THE ANALYSIS OF GENES EXPRESSED IN THE SPERM STORAGE ORGANS OF *DROSOPHILA MELANOGASTER:* PATTERNS OF EVOLUTION AND EXPRESSION

by

Adrianne Marie Prokupek-Pickett

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Biological Sciences

Under the Supervision of Professor Lawrence G. Harshman

Lincoln, NE

May, 2008

Prokupek 2008

UMI Number: 3297753

Copyright 2008 by
Prokupek-Pickett, Adrianne Marie

All rights reserved.

UMI Microform 3297753
Copyright 2008 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

THE ANALYSIS OF GENES EXPRESSED IN THE SPERM STORAGE ORGANS OF *DROSOPHILA MELANOGASTER:* PATTERNS OF EVOLUTION AND EXPRESSION

Adrianne Marie Prokupek-Pickett, Ph.D

University of Nebraska, 2008

Advisor: Lawrence G. Harshman

The molecular evolutionary rate of genes associated with sexual traits and reproduction is (on average) faster than that of genes coding for non-sex-related traits. In Drosophila, this trend has been observed in genes expressing male accessory gland products, as well as in genes expressed in the entire reproductive tract of the females. Though a general trend of rapid evolution has been observed in the *Drosophila* female reproductive tract, relatively little is known about the evolutionary patterns of the genes expressed in specific organs within the female reproductive system. Sperm storage organs are important to the reproductive success of both males and females, as sperm in storage must remain viable and be properly released for successful fertilization. Further more, sperm storage organs provide an arena for important evolutionary processes such as sperm competition and female sperm choice. This study investigated the evolutionary and expression patterns of genes found in the sperm storage organs of *Drosophila*. A high percentage of genes found in sperm storage organs were found to be evolving at rapid rates. Polymorphism data confirmed that for five of the genes discovered, the rapid evolution can be attributed to positive selection acting on a subset of the codons of these genes. The function of

many of the genes identified suggests that these genes could interact with seminal products, and this interaction could be a factor in the elevation of evolutionary rates. Regions upstream of genes typically contain a number of non-coding regulatory regions; changes in these regions could result in changes in gene expression. Regions upstream of five of the sperm storage genes were found to be over twice as polymorphic as neutral regions of non-coding DNA. Using microarrays to investigate expression patterns, it was found that the expression of genes associated with sperm storage showed time and organ dependent patterns, presumably associated with different sperm storage events. The identification of genes involved in sperm storage and the characterization of their expression and evolutionary patterns is a new and important development in the understanding of the mechanics and timing of fertilization. This understanding has broader impacts such as insight into important processes (entrance, maintenance and release) of sperm storage for a wide range of species. This study is one of the first to describe important female proteins that may be interacting (possibly co-evolving), either in concert or antagonistically with male seminal products.

TABLE OF CONTENTS

CONTENTS PAGE

TITLE PAGE …..……i

ABSTRACT ..……...……...…..ii

TABLE OF CONTENTS …….…….…….iv

LIST OF FIGURES ...…….……..vi

LIST OF TABLES ……..…….…….vii

LIST OF ABBREVIATIONS ……viii

ACKNOWLEDGEMENTS ……..……..……...…....x

CHAPTER I: INTRODUCTION ……..…….……1

CHAPTER II: An evolutionary expressed sequence tag analysis of *Drosophila* spermatheca genes

Abstract…….…….23

Introduction……24

Materials and Methods ……..……28

Results ……...…….34

Discussion …….39

Acknowledgements……48

References……..49

CHAPTER III: A resequencing study of sperm storage organ genes: a cluster of proteases and a gene of unknown classification.

Abstract…….…….73

Introduction……74

Materials and Methods ……..……77

Results ……..…….80

Discussion …….84

Acknowledgements……87

References……..88

CHAPTER IV: Patterns of gene expression in the sperm storage organs of *Drosophila melanogaster*

Abstract…….……...123

Introduction……..124

Materials and Methods ……..……..127

Results/Discussion ……...……...131

Acknowledgements……..…140

References……141

LIST OF FIGURES

FIGURE Page

1.1……20

2.1……59

2.2……61

3.1……96

3.2……98

3.3……100

3.4……102

3.5……104

3.6……106

3.7……108

4.1……165

4.1 (Supplementary) ……167

4.2 (Supplementary)……169

LIST OF TABLES

TABLES Page

2.1……63

2.2……65

2.1 (Supplementary)……67

3.1…… 110

3.2…… 112

3.3……..114

3.4……..116

3.5……...118

3.6……120

4.1……..163

4.1 (Supplementary) …….....182

4.2 (Supplementary)…….....183

4.3 (Supplementary) …….....184

4.4 (Supplementary)…….....185

4.5 (Supplementary) …….....186

4.6 (Supplementary)…….....187

4.7 (Supplementary) …….....188

4.8 (Supplementary)…….....189

List of Abbreviations

Acp …….….. male accessory gland protein

bp …….…….…….. base pair

cDNA ……...…… complementary Deoxyribonucleic acid

cRNA……... complementary Ribonucleic acid

D. ana ……..…… *Drosophila ananassae*

D. ere …… *Drosophila erecta*

D. grim ……..……. *Drosophila grimshawi*

D. mel……. *Drosophila melanogaster*

D. moj…… *Drosophila mojavensis*

D. per…….……. *Drosophila persimilis*

D. pse…….. *Drosophila pseudoobscura*

D. sec ……..……. *Drosophila sechellia*

D. sim …….. *Drosophila simulans*

D. vir…..…… *Drosophila virilis*

D. wil……..…..…… *Drosophila willistoni*

D. yak…. …… *Drosophila yakuba*

dN ……..……nonsynonymous substitutions per nonsynonymous site

dS ……. synonymous substitutions per synonymous site

hLRT …… hierarchial likehood ratio test

Ka ……..……nonsynonymous substitutions per nonsynonymous site

Ks ……. synonymous substitutions per synonymous site

Kb …… Kilobases

mRNA …… messenger Ribonucleic Acid

PAML ……..…phylogenetic analysis by maximum likelihood

PCR …….……. polymerase chain reaction

RNA ……..……..……... ribonucleic acid

SP…….……....sex peptide

SP……...…….signal peptide

SR ……..……..……..…. seminal receptacle

SSO ……..…….sperm storage organ

ST …….. spermathecae

TM ……..…… transmembrane

Acknowledgments

I am deeply grateful and eternally indebted to all of my family, friends and collegues who helped me both academically and emotionally, through all stages of this dissertation.

To my advisor, Dr. Lawrence Harshman, thank you for all of the time you spent helping to guide me through the rough terrains of graduate school. Thank you for providing me with the beginnings of a project, and for you support and patience in allowing me to take that project make it my own. During the six years I spent in your lab I was allowed to develop my research, critical thinking, writing and teaching skills to levels I never imagined I could reach. You truly are a great teacher, scientist and friend.

I am grateful to my committee members for all of your comments and suggestions. I have benefited greatly from all of your advice and insight. Thank you for taking time out of your lives to make sure that I stayed focused and continued on in the right direction.

I do not posses the words to fully express the depths of my graditute to my entire family. To come into and grow up in a world filled with so much love and support is more than I could wish for, and I am grateful everyday for every opportunity I was provided. To my parents, your unwavering belief in me allowed me to fully believe in myself. My big brother, you gave me the courage to take chances, whether I wanted to or not. My little sister, you helped teach me patience, and kept my life silly and young. And to my husband, thank you for always being my

“biggest fan”. You gave me a safe place to run to when I felt like things were getting too big for me, but also gave me the strength to go back out and keep trying.

To all of the faculty and staff members in School of Biological Sciences who were always kind and willing to help. I came in with requests that were sometimes outrageous, and almost always last minute, but you always came through for me. I am especially thankful to all of the members of the Manter lunch table, no matter how stressed out I was you always made me laugh and have fun.

I could not have succeeded without the friendship and understanding of all of my lab mates, both past and present. You made this more than a place to come and work, you made it like a home. You opened up my small Nebraska world to a wide diversity of cultures and experiences, and forever changed me. I hope while we all pursue our many differing interests we never lose track of each other.

To my friends who have stuck with me through all of this. As I look back on all of the adventures we have had together, I am sometimes so surprised we have all made it so far. Thank you for filling my life with surprises, laughter and love.

A final thank you to all of my gracious sources of for providing teaching and research assistantships, travel and research funding throughout my graduate career: School of Biological Sciences, Initiative for Ecology and Evolutionary Analysis (IEEA), Nebraska Idea Networks of Biomedical Research Excellence (Ne-INBRE), U.S. Department of Education (GAANN), and NSF grant DEB-ESP0346476.

Chapter I.

Introduction

An emerging theme in evolutionary biology is that reproductive proteins evolve rapidly in animals and plants (Clark et al., 2006). One example is genes in the accessory glands of male *Drosophila melanogaster,* which produce proteins transferred to females at the time of mating (Swanson et al., 2001a, Wolfner, 2002). Some of these genes exhibit evidence for positive selection based elevated rates of non-synonymous substitutions inferred from comparative DNA sequence analysis (Swanson et al., 2001a). Genes expressed in the female reproductive tract of *D. melanogaster* also show evidence for positive selection (Swanson et al., 2004b). However, in *D. melanogaster* almost nothing is known about what genes are expressed specifically in sperm storage organs (SSOs), how these genes contribute to SSO function, and how these genes evolve. The absence of data about the evolution of these genes is a major gap in knowledge because genes that are expressed specifically in SSOs potentially play important roles in evolutionary processes such as sexual antagonistic coevolution, sperm competition, and speciation. The second chapter of this dissertation addresses this problem using *Drosophila* as a model organism and a unique and powerful technique, hybrid selection, to create a cDNA library enriched for genes that are specifically expressed in the spermathecae, the long term SSO of *Drosophila*. This procedure allowed for the focus to be on a relatively small set of genes that were hypothesized to evolve rapidly and to play an important role in sperm storage. The ***hypothesis*** for this study was that a relatively high proportion of genes identified by enriched expression in the spermathecae evolve rapidly and exhibit the molecular signatures of positive Darwinian selection. The third chapter of this dissertation describes a polymorphism study conducted to help define the evolutionary patterns of genes expressed in the spermathecae. Polymorphism studies are helpful in

determining the rate and pattern of evolutionary change at the population level, versus the long-term evolution inferred from divergence studies alone. The ***hypothesis*** was that there would be various indicators of selection based on the pattern of variation within and between populations, and between related species. The fourth chapter of this dissertation uses another powerful technique, microarrays, to conduct a transcriptome analysis of the two types of SSOs of *D. melanogaster* (spermathecae and seminal receptacle). This analysis provided a means to investigate gene expression and identify proteins in these organs. The ***hypothesis*** was that expression patterns of the genes are distinctive between the two SSOs based on their differing roles after sperm acquisition.

A. Sperm Storage

Sperm storage is an important reproductive strategy utilized by females in species with internal fertilization. Sperm storage allows for the female acquisition of sperm to take place days or even months before it is needed to fertilize a mature oocyte. The functional reproductive benefits of this process are great as contact with a suitable male does not necessary correspond with egg availability. For multiply mated females, sperm storage can offer an additional selection step via sperm competition and/or cryptic female sperm choice, as well as give the female a mechanism to compensate for infertile/genetically incompatible males. The duration of time sperm spend in storage is highly variable, extending from a few hours (e.g., mouse), up to decades (e.g., honey bee). This complex and dynamic process includes components such as female and male anatomical structures, muscular contraction of the female genital tract, site of ejaculate deposition, molecules in the seminal fluid, female genital tract secretions, sperm

mobility, and proteins associated with the sperm or other seminal products (Wolfner, 1997, Bloch-Qazi et al., 1998, Neubaum & Wolfner, 1999c, Simmons & Kotiaho, 2002). Specialized SSOs have evolved in a number of species including sperm storage tubules in birds, spermathecae in amphibians and arachnids and in insects SSOs typically appear as sac-like structures (spermathecae) or long tubules (seminal receptacles). These organs are responsible for maintaining sperm viability, organizing the sperm in storage, and facilitating the proper release of sperm from storage (reviewed in (Neubaum & Wolfner, 1999c, Bloch-Qazi et al., 2003).

Drosophila melanogaster females have evolved two types of specialized SSOs, the seminal receptacle and spermathecae (Figure 1.1). The seminal receptacle (SR) is a coiled structure which is innervated and partially surrounded by muscle (Fowler, 1973). In addition, spermathecae (ST), a pair of sperm storage pouches, are the site of longer-term sperm storage (upwards of two weeks) (Figure 1.1). Aging of sperm in storage is an issue (Snook & Hosken, 2004) implying that one of the important functions sperm storage, especially in the long term sperm (the ST) is sperm maintenance. In the course of a single mating, 4,000-6,000 sperm are transferred to a female but only approximately 1,000 of the transferred sperm are stored (Gilbert, 1981). More sperm are stored in the SR (65-80%) than in the ST (Gilbert, 1981, Tram & Wolfner, 1999, Bloch-Qazi et al., 2003). The number of sperm in storage is correlated with net fecundity and the timing of female remating (Gromko et al., 1984, Newport & Gromko, 1984, Harshman & Clark, 1998).

When a female mates with two males in succession, most of the offspring are fertilized by sperm from the second male (Lefevre & Jonsson, 1962), this process is referred to as sperm precedence. Sperm precedence could be attributed to a variety of

processes which include sperm displacement (Lefevre & Jonsson, 1962), sperm stratification (layering of stored sperm based on time of entrance into storage) (Waage, 1986, Schlager, 1960), or female sperm selection (Eberhard, 1996, Birkhead & Moller, 1998). The relative number of sperm transferred is a determinant of the proportion of progeny sired by the second male (Gilchrist & Partridge, 2000), possibly due to the displacement of the previous males sperm. Direct observation of fluorescently labeled sperm indicates that the first male's sperm can be physically displaced from storage by sperm from the second male to mate (Civetta, 1999). Male accessory gland fluid can also act to reduce the amount of progeny sired by previously stored sperm (Scott & Richmond, 1990, Scott & Williams, 1993, Harshman & Prout, 1994) . Male accessory gland products could impact sperm in storage by incapacitation of the first male's sperm by the second male's seminal products (Price et al., 1999). Sperm storage and sperm competition are related issues and many studies using *D. melanogaster* have played an important role in investigating sperm competition (Prout & Bundgaard, 1977, Gilchrist & Partridge, 2000, Harshman & Prout, 1994, Clark et al., 1995, Clark & Begun, 1998).

The nature of the relationship between sperm competition and sperm storage remain undefined for any species. The major area of deficit lies in the understanding of the processes involved in sperm storage. How do female SSOs function and what selective forces act on genes specifically expressed in these structures? SSOs are black boxes and this lack of knowledge is surprising given that the SSOs could play a major role in evolutionary phenomena such as sperm competition, sexually antagonistic coevolution, and speciation.

B. Sexual Conflict and Male - Female Coevolution

Sexual conflict is defined by Parker (1979) as "a conflict between the evolutionary interests of individuals of the two sexes" (Parker, 1979). This conflict stems from anisogamy, the large difference in gamete size in the two sexes. Anisogamy leads to different parental investment, and therefore different optimal values of reproductive traits in the two sexes (Trivers, 1972, Parker, 1979). The reciprocal selection pressures on reproductive traits associated with sexual conflict can lead to sexually antagonistic coevolution in which males and females are locked in an evolutionary race (Rice & Holland, 1997, Rice, 1996, Gavrilets, 2000, Rice, 2000, Rowe & Day, 2006). Sexually antagonistic coevolution can drive rapid evolution, even over a short time frame, and have important ramifications for a wide range of traits and phenomena including gamete interactions, sperm use, mate choice and speciation (e.g. (Rowe et al., 1994, Chapman et al., 1995, Pitnick et al., 1999, Civetta & Clark, 2000, Czesak & Fox, 2003, Lessells, 2006). In a key study using *D. melanogaster*, male traits were allowed to evolve while the evolution of female traits was arrested. In as little as 40 generations, dramatic changes were seen in the ability of adapted males to both secure mates and induce female mating refractoriness. A marked increase in the toxic effects of products in the seminal fluid was observed in terms of shortened female lifespan (Rice, 1996). A subsequent study showed that females can evolve resistance to the toxic effects of males (Holland & Rice, 1999). It is important to identify the molecules and understand the mechanisms that underlie this remarkably strong selection as it is likely to provide insight into a range of evolutionary processes. Male accessory gland protein gene products (Acps) could be a proximate cause of sexual antagonism. Acps induce post-mating responses in females which may

have a negative impact on female fitness. For example, experiments using males from a transgenic line, in which the main cells (the source of accessory gland proteins) of the male accessory gland had been ablated by transgenic expression of a toxin, showed that Acps cause the cost of mating (Chapman et al., 1995). In general, sexual conflict could be found widely as a source of strong selection (Rowe & Arnqvist, 2002, Arnqvist & Rowe, 1995). Antagonistic interactions between male seminal products and female proteins in the SSOs could be responsible for the rapid evolution of SSO genes.

C. Male Accessory Gland Proteins in *D. melanogaster*

Male accessory gland proteins (Acps) are transferred to the female at the time of mating in *D. melanogaster*. Thereafter they exert a range of effects on female reproduction and physiology including the induction of sperm storage, ovulation and egg-laying, and decreasing female receptivity to future mates and female lifespan (Wolfner, 1997, Swanson et al., 2001a). Interaction of Acps with proteins of the SSOs could be a strong driver of selection. Genes that encode these proteins show evidence for positive selection measured by the analysis of d_N (nonsynonymous substitutions per nonsynonymous site) to d_S (synonymous substitutions per synonymous site) ratios (Swanson et al., 2001a). A lower proportion of the Acps show evidence for positive selection compared to the proportion of genes identified by enriched expression in spermathecae (Chapter II).

Three *D. melanogaster* Acps will be briefly discussed based on the fact that after mating all three are found in the female SSOs. For decades it has been known that mating increases female egg production and reduces female receptivity to additional mates.

Acp70A (sex peptide, (SP)) is largely responsible for these effects (Kubli, 1996) as demonstrated by a mutation in the SP gene and RNAi suppression of the expression of this gene in males (Liu & Kubli, 2003, Chapman et al., 2003b). SP enters SSOs attached to sperm and, when cleaved, stimulates the biosynthesis of juvenile hormone which induces ovulation (Moshitzky et al., 1996). Two of the genes identified in the spermathecal library (Chapter II) are associated with juvenile hormone function. Another male accessory gland protein, Acp36DE, is required for normal levels of sperm storage (Neubaum & Wolfner, 1999b, Chapman et al., 2000, Bertram et al., 1996). It moves in loose association with the sperm and accumulates in both SSOs (Tram & Wolfner, 1999, Bloch-Qazi & Wolfner, 2003). Acp62F is a protease inhibitor shown to be toxic to females when ectopically expressed (Lung et al., 2002a). Importantly, Acp62F is one of a substantial number of protease inhibitors among the protein products encoded by Acps. This is important because a high proportion of the rapidly evolving genes in the spermathecae are serine proteases (Chapter II). Approximately 20 percent of the predicted Acps are proteases or protease inhibitors (Swanson et al., 2001a, Wolfner, 2002).

D. Evolution of Reproductive Proteins

Elevated rates of evolution for reproductive proteins have been observed in various taxa (Clark et al., 2006). Rapid evolutionary divergence has been documented in protistans, fungi, plants, and animals in gamete surface proteins (Swanson & Vacquier, 2002). This divergence is most often attributed to positive selection, indicating that changes in the structure of the protein are adaptive (Yang & Bielawski, 2000).

High rates of evolution have been observed in both egg proteins and sperm receptors. In abalone, egg vitelline envelope receptor (VERL) and sperm lysin are a cognate pair of gamete recognition proteins that regulate fertilization. These proteins mediate species-specific sperm binding to eggs. The amino-terminal end of the egg VERL, as well as sperm lysin, has undergone very rapid positive Darwinian selection (Galindo et al., 2003b). In mammals, egg coat (zona pellucida) glycoproteins and several sperm proteins evolve rapidly due to positive selection (Swanson et al., 2001b, Swanson et al., 2003). Positive selection has also been shown for *Drosophila* Acps (Begun et al., 2000b, Swanson et al., 2001a) as well as proteins within the *Drosophila* female reproductive tract, including proteases (Swanson et al., 2004b). Protease inhibitors have been found to be rapidly evolving among the male accessory gland proteins (Swanson et al., 2001a) perhaps due to interactions with rapidly evolving female proteases. The selective pressures driving the high rate of evolution are unknown, but they could be due to a conflict in the reproductive interests. It is possible that the two sexes are locked in a "coevolutionary chase" driven by sperm competition, sexual conflict, or sexual selection (Galindo et al., 2003b, Swanson et al., 2003).

E. The Effect of Mating on Gene Expression in Female *D. melanogaster*

The identification of changes triggered in females by mating is important for the understanding of the female role in post-copulatory processes such as sperm storage. McGraw *et al.* (2004) used microarrays to compare gene expression of *D. melanogaster* females which where virgin, mated to normal males, mated to spermless males or mated to males lacking Acps. It was determined that 1,783 genes showed a change in expression

level due to mating, though only 46 showed a two fold or greater change. A genome-wide comparative microarray analysis done on virgin, courted, and two-hour post-mated females found that 23 genes were differentially expressed in virgin courted versus uncourted females, and 38 differentially expressed genes were observed in the comparison of virgin and mated females (Lawniczak & Begun, 2004). In both studies, immune-related genes and serine proteases were found to be effected by mating. The number of serine proteases influenced by mating was determined to be higher than expected by chance (Lawniczak & Begun, 2004). The role of such proteases within the female is unknown. It is interesting to note that serine proteases, as well as serine protease inhibitors (serpins), are also present in seminal fluid (Coleman et al., 1995, Swanson et al., 2001a). Proteases observed to be differentially expressed in microarray studies (Arbeitman et al., 2004, McGraw et al., 2004, Lawniczak & Begun, 2004) were found to partially overlap with protease genes identified in the spermathecal evolutionary EST study (Chapter II). Mack *et al*. (2006) investigated the effect of mating on gene expression in a region of the female reproductive tract that included ST and the SR using microarrays and two-dimensional protein electrophoresis (Mack et al., 2006b). The study by Mack *et al.* showed a peak in gene expression at 6 hours post-mating which helped to define one of the time points for the microarray study in this dissertation (Chapter III).

A study investigating a nuclear hormone receptor HR39 of *Drosophila* found that female mutants for Hr39 have defective or absent spermathecae. Hr39 was found to function similarly to the mammalian splicing factor 1 (SF1), which is necessary to produce both androgens and Mullerian-inhibiting substances to prevent oviduct degeneration (Allen & Spradling, 2008). The spermathecae and parovaria (Figure 1.1)

were found to secrete proteins which function in sperm maturation and in storage, similar to those found in the male epididymis in mammals. The Allen and Spradling study helped to reveal close connections between Dipteran and mammalian reproductive biology, suggesting that there is conservation in specific steps of reproduction including sperm storage. The results of this dissertation could provide relevant ties to mammalian (human) reproduction biology.

LITERATURE

Allen, A. K., and A. C. Spradling. 2008. The Sf1-related nuclear hormone receptor Hr39 regulates *Drosophila* female reproductive tract development and function. Development 135:311-321.

Arbeitman, M. N., A. A. Fleming, M. L. Siegal, B. H. Null, and B. S. Baker. 2004. A genomic analysis of *Drosophila* somatic sexual differentiation and its regulation. Development 131:2007-2021.

Arnqvist, G., and L. Rowe. 1995. Sexual conflict and arms races between the sexes: a morphological adaptation for control of mating in a female insect. Proc. R. Soc. Lond B 261:123-127.

Begun, D. J., P. Whitley, B. L. Todd, H. M. Waldrip-Dail, and A. G. Clark. 2000. Molecular population genetics of male accessory gland proteins in *Drosophila*. Genetics 156:1879-1888.

Bertram, M. J., D. M. Neubaum, and M. F. Wolfner. 1996. Localization of the *Drosophila* male accessory gland protein Acp36DE in the mated female suggests a role in sperm storage. Insect Biochem Mol Biol 26:971-980.

Birkhead, T. R., and A. P. Moller. 1998. Sperm competition and sexual selection. Academic Press, London.

Bloch-Qazi, M. C., J. R. Aprile, and S. M. Lewis. 1998. Female role in sperm storage in the red flour beetle, *Tribolium castaneum*. Comparative Biochemistry and Physiology 120:641-648.

Bloch-Qazi, M. C., Y. Heifetz, and M. F. Wolfner. 2003. The developments between gametogenesis and fertilization: ovulation and female sperm storage in *Drosophila melanogaster*. Developmental Biology 256:195-211.

Bloch-Qazi, M. C., and M. F. Wolfner. 2003. An early role for the *Drosophila melanogaster* male seminal protein Acp36DE in female sperm storage. Journal of Experimental Biology 206:3521-3528.

Chapman, T., J. Bangham, G. Vinti, B. Seifried, O. Lung, M. F. Wolfner, H. K. Smith, and L. Partridge. 2003. The sex peptide of *Drosophila melanogaster*: female post-mating responses analyzed by using RNA interference. Proc Natl Acad Sci U S A 100:9923-9928.

Chapman, T., L. F. Liddle, J. M. Kalb, M. F. Wolfner, and L. Partridge. 1995. Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. Nature 373:241-244.

Chapman, T., D. M. Neubaum, M. F. Wolfner, and L. Partridge. 2000. The role of male accessory gland protein Acp36DE in sperm competition in *Drosophila melanogaster*. Proc Biol Sci 267:1097-1105.

Civetta, A. 1999. Direct visualization of sperm competition and sperm storage in *Drosophila*. Current Biology 9:841-844.

Civetta, A., and A. G. Clark. 2000. Correlated effects of sperm competition and postmating female mortality. Proc Natl Acad Sci U S A 97:13162-13165.

Clark, A. G., M. Aguade, T. Prout, L. G. Harshman, and C. H. Langley. 1995. Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. Genetics 139:189-201.

Clark, A. G., and D. J. Begun. 1998. Female genotypes affect sperm displacement in *Drosophila*. Genetics 149:1487-1493.

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

Coleman, S., B. Drahn, G. Peterson, J. Stolorv, and K. Kraus. 1995. A *Drosophila* male accessory gland proteins that is a member of the serpin superfamily of proteinase inhibitors is transferred to females during mating. Insect Biochem Mol Biol 25:203-207.

Czesak, M. E., and C. W. Fox. 2003. Genetic variation in male effects on female reproduction and the genetic covariance between the sexes. Evolution Int J Org Evolution 57:1359-1366.

Eberhard, W. G. 1996. Female control: sexual selection by cryptic female choice. Princeton University Press, New Jersey.

Fowler, G. 1973. Some aspects of the reproductive biology of *Drosophila*: sperm transfer, sperm storage and sperm utilization. Adv in Genet 17:293-360.

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 U S A 100:4639-4643.

Gavrilets, S. 2000. Rapid evolution of reproductive barriers driven by sexual conflict. Nature 403:886-889.

Gilbert, D. G. 1981. Ejaculate esterase 6 and initial sperm use by female *Drosophila melanogaster*. J. Insect Physiol. 27:641 - 650.

Gilchrist, A. S., and L. Partridge. 2000. Why it is difficult to model sperm displacement in *Drosophila melanogaster*: the relation between sperm transfer and copulation duration. Evolution Int J Org Evolution 54:534-542.

Gromko, M. H., M. E. A. Newport, and M. G. Kortier. 1984. Sperm dependence of female receptivity to remating in *Drosophila melanogaster*. Evolution 38:1273-1282.

Harshman, L. G., and A. G. Clark. 1998. Inference of sperm competition from broods of field-caught *Drosophila*. Evolution 52:1334-1341.

Harshman, L. G., and T. Prout. 1994. Sperm displacement without sperm transfer in *Drosophila melanogaster*. Evolution 48:758-766.

Holland, B., and W. R. Rice. 1999. Experimental removal of sexual selection reverses intersexual antagonistic coevolution and removes a reproductive load. Proc Natl Acad Sci U S A 96:5083-5088.

Kubli, E. 1996. The *Drosophila* sex-peptide: A peptide pheromone involved in reproduction. Adv. in Dev. Biol. Chem 4:99-128.

Lawniczak, M. K., and D. J. Begun. 2004. A genome-wide analysis of courting and mating responses in *Drosophila melanogaster* females. Genome 47:900-910.

Lefevre, G., and U. B. Jonsson. 1962. Sperm transfer, storage, displacement, and utilization in *Drosophila melanogaster*. Genetics 47:1719-1736.

Lessells, C. M. 2006. The evolutionary outcome of sexual conflict. Philos Trans R Soc Lond B Biol Sci 361:301-317.

Liu, H., and E. Kubli. 2003. Sex-peptide is the molecular basis of the sperm effect in *Drosophila melanogaster*. Proc Natl Acad Sci U S A 100:9929-9933.

Lung, O., U. Tram, C. M. Finnerty, M. A. Eipper-Mains, J. M. Kalb, and M. F. Wolfner. 2002. The *Drosophila melanogaster* seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression. Genetics 160:211-224.

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

McGraw, L. A., G. Gibson, A. G. Clark, and M. F. Wolfner. 2004. Genes regulated by mating, sperm, or seminal proteins in mated female *Drosophila melanogaster.* Curr Biol 14:1509-1514.

Moshitzky, P., I. Fleischmann, N. Chaimov, P. Saudan, S. Klauser, E. Kubli, and S. W. Applebaum. 1996. Sex-peptide activates juvenile hormone biosynthesis in the *Drosophila melanogaster* corpus allatum. Arch. Insect Biochem. Physiol. 32:363-374.

Neubaum, D. M., and M. F. Wolfner. 1999a. Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. Genetics 153:845-857.

Neubaum, D. M., and M. F. Wolfner. 1999b. Wise, winsome, or weird? Mechanisms of sperm storage in female animals. Curr Top Dev Biol 41:67-97.

Newport, K. M., and M. H. Gromko. 1984. The effect of experimental design on female receptivity to remating and its impact on reproductive success in *Drosophila melanogaster*. Evolution 38:1261-1272.

Parker, G. A. 1979. Sexual selection and reproductive competition in insects. Pp. 123-166 *in* M. S. Blum, and N. A. Blum, eds. Academic Press, New York.

Pitnick, S., T. A. Markow, and G. S. Spicer. 1999. Evolution of multiple kinds of female sperm-storage organs in *Drosophila*. Evolution 53:1804-1822.

Price, C. S. C., K. A. Dyer, and J. a. Coyne. 1999. Sperm competition between *Drosophila* males involves both displacement and incapacitation. Nature 400:449-452.

Prout, T., and J. Bundgaard. 1977. The population genetics of sperm displacement. Genetics 85:95-124.

Rice, W. R. 1996. Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. Nature 381:232-234.

Rice, W. R. 2000. Dangerous liaisons. Proc Natl Acad Sci U S A 97:12953-12955.

Rice, W. R., and B. Holland. 1997. The enemies within: intergenomic conflict, interlocus contest evolution (ICE), and the intraspecific Red Queen. Behav. Ecol. Sociobiol. 41:1-10.

Rowe, L., and G. Arnqvist. 2002. Sexually antagonistic coevolution in a mating system: combining experimental and comparative approaches to address evolutionary processes. Evolution Int J Org Evolution 56:754-767.

Rowe, L., G. Arnqvist, A. Sih, and J. J. Krupa. 1994. Sexual conflict and the evolutionary ecology of mating patterns: water striders as a model system. TREE 9.

Rowe, L., and T. Day. 2006. Detecting sexual conflict and sexually antagonistic coevolution. Philos Trans R Soc Lond B Biol Sci 361:277-285.

Schlager, G. 1960. Sperm precedence in the fertilization of eggs in *Tribolium castaneum*. Ann. Entomol. Soc. Am. 53:557-560.

Scott, D., and R. C. Richmond. 1990. Sperm loss by remating in *Drosophila melanogaster* females. J. Insect Physiology 36:451-456.

Scott, D., and E. Williams. 1993. Sperm displacement after remating in *Drosophila melanogaster*. J. Insect Physiology 39:201-206.

Simmons, L. W., and J. S. Kotiaho. 2002. Evolution of ejaculates: patterns of phenotypic and genotypic variation and condition dependence in sperm competition traits. Evolution Int J Org Evolution 56:1622-1631.

Snook, R. R., and D. J. Hosken. 2004. Sperm death and dumping in *Drosophila*. Nature 428:939-941.

Swanson, W. J., A. G. Clark, H. M. Waldrip-Dail, M. F. Wolfner, and C. F. Aquadro. 2001a. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc Natl Acad Sci U S A 98:7375-7379.

Swanson, W. J., R. Nielsen, and Q. Yang. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. Mol. Biol. Evol. 20:18-20.

Swanson, W. J., and V. D. Vacquier. 2002. The rapid evolution of reproductive proteins. Nat Rev Genet 3:137-144.

Swanson, W. J., A. Wong, M. F. Wolfner, and C. F. Aquadro. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. Genetics 168:1457-1465.

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

Tram, U., and M. F. Wolfner. 1999. Male seminal fluid proteins are essential for sperm storage in *Drosophila melanogaster*. Genetics 153:837-844.

Trivers, R. L. 1972. Parental investment and sexual selection. Pp. 1871-1971 *in* B. Cambell, ed. Sexual selection and the decent of man. Aldine-Atherton, Chicago.

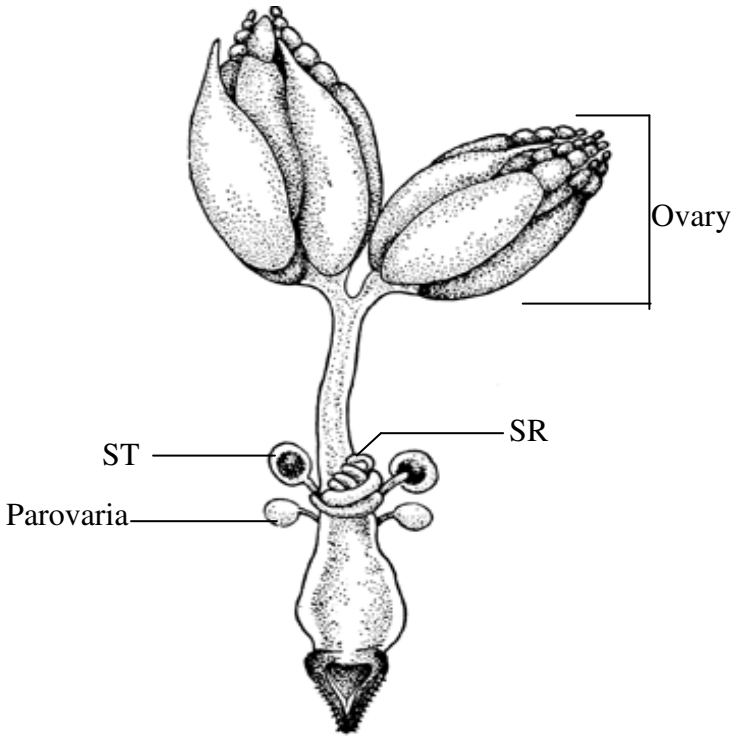
Waage, J. K. 1986. Evidence for widespread sperm displacement ability among Zygoptera (Ordonata) and the means for predicting its presence. Biol. J. Linn. Soc. 28:285-300.

Wolfner, M. F. 1997. Tokens of love: functions and regulation of *Drosophila* male accessory gland products. Insect Biochem Mol Biol 27:179-192.

Wolfner, M. F. 2002. The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in *Drosophila* Heredity 88:85-93.

Yang, Z., and J. Bielawski. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431-449.

Figure 1.1: Diagram of the female reproductive system in *Drosophila melanogaster*. SR, seminal receptacle; Sp, spermatheca; Ovary; Parovaria (female accessory gland). (Reprinted from Patterson and Stone 1952)


Ovary
SR
ST
Parovaria

Chapter II.

Expressed Sequence Tag Analysis of *Drosophila* Spermatheca Genes: Protein Identity and Evolution

[Based on the manuscript submitted to *Evolution* by Adrianne M. Prokupek, Federico Hoffmann, Seong-il Eyun, Etsuko N. Moriyama, Min Zhou, Lawrence G. Harshman]

ABSTRACT

Sperm competition, cryptic female choice of sperm, and sexually antagonistic co-evolution are forms of sexual selection that can drive strong selection and play an important role in speciation. Although there is keen interest in molecular mechanisms of sperm competition and female choice of sperm, very little is known in any species about female genes which affect sexual selection. Sperm storage organ genes are likely to play an important role in sexual selection and speciation. In this study, genes enriched for expression in the long term sperm storage organ of *Drosophila*, the spermatheca, were investigated. A high proportion of genes enriched for expression in the spermatheca have evolved rapidly. Especially notable was the high incidence of genes that exhibit the molecular signature of positive selection even when compared to male gland protein genes, which are a paradigm for this pattern of evolution. Genes that encode serine proteases were a prominent category of genes associated with the spermatheca. They tend to evolve rapidly and all have secretion signals. Genes with secretion signals and transmembrane domains encode proteins that might directly interact with ejaculate proteins and coevolve with them. This study identified spermatheca genes which are candidates to affect sexual selection and speciation.

INTRODUCTION

Darwin (1871) defined sexual selection as the advantage some individuals have over others of the same sex in relation to reproduction. This type of selection can result in the evolution of conspicuous traits such as extravagant secondary sexual characteristics in some species or, more subtly, may be working on a molecular level through the evolution of proteins involved in reproductive processes such as sperm competition, female sperm storage or female sperm choice. Rapidly evolving reproductive proteins are likely involved in sexual selection, playing specific roles in inter (between), intra (within) sex competitions, or a combination of both. One major area where there is a deficit in information is in the identification and classification of female reproductive proteins, especially those proteins interacting directly with male seminal products (e.g. sperm storage proteins).

Reproductive proteins associated with female sperm storage organs (SSOs) are candidates to play important roles in evolutionary phenomena such as sperm competition, female sperm choice, sexual selection and consequently in speciation. However, SSOs are poorly understood in terms of the evolution of proteins associated with these organs, how the organs function at the molecular level, and the relationship between evolution and function. Identification of female proteins expressed within SSOs is a vital step in the development of a comprehensive understanding of the role of SSOs in evolution.

Specific genes and proteins known to play a role in sperm competition have, thus far, only been identified in males. In *D. melanogaster,* second male sperm precedence (P2) is due, in part, to a non-sperm component of the ejaculate (Harshman & Prout, 1994). It is now established that male accessory gland proteins play a role in sperm

competition in *Drosophila* (Ravi-Ram & Wolfner, 2007). Allelic variation in male accessory gland protein genes has been associated with the fertilization success of both the first male to mate with a female (first male sperm precedence (P1)), and the second male to mate with the female (second male sperm precedence (P2)) in this species (Clark et al., 1995, Fiumera et al., 2005, Fiumera et al., 2007). Genetic studies reveal that female processes have a major effect of the outcome of sperm competition in *D. melanogaster* (Clark & Begun, 1998, Clark et al., 1999), but the specific female genes that have these effects have not been identified. Female SSOs provide an arena for sperm competition, making female SSOs a likely place to find female reproductive molecules that affect sperm competition.

Studies on conspecific sperm precedence have provided clues about the importance of male and female reproductive proteins in sperm competition. Conspecific sperm precedence occurs when a female mated to both conspecific and heterospecific males, regardless of the mating order, preferentially produces conspecific rather than hybrid offspring (Howard, 1999). Conspecific sperm precedence is likely to play a major role in speciation by the reproductive isolation of closely related species (Coyne & Orr, 2004). This phenomenon has been observed in a diverse range of taxa including flour beetles, sea urchins, *Drosophila,* rabbits, and several plant species (reviewed in Howard et al., in prep, Howard, 1999). In *D. melanogaster*, conspecific sperm precedence can involve the incapacitation of sperm of the first male to mate by the seminal fluid of the second male to mate (Price, 1997, Price et al., 1999). In this species, heterospecific sperm are not typically displaced by the second male, therefore, the heterospecific sperm in the female SSOs are exposed to the seminal fluid of the second male to mate. Thus, sperm

precedence may be due to interactions between the seminal fluid and reproductive proteins in SSOs. In *D. mauritiana*, stored heterospecific sperm are rapidly lost from SSOs (Price et al., 2001), presumably due to improper storage. Female SSOs and the genes they express could play a major role in conspecific sperm precedence, or precluding fertilization by heterospecific sperm. Direct interactions between female proteins and sperm, or male seminal products may be necessary for the identification of conspecific versus heterospecific sperm, providing another scenario for important male-female interaction.

Rapid evolution of reproductive proteins has been documented in protistans, fungi, plants, and animals (Clark et al., 2006) in both male and female gametic proteins (Swanson & Vacquier, 2002a, Galindo et al., 2003a). For example, in sea urchins male sperm bindin evolves rapidly (Palumbi, 1999) as does the corresponding receptor for bindin on the egg (Palumbi, 1999, Kamei et al., 2000). In mammals, egg coat (zona pellucida) glycoproteins and several sperm proteins evolve rapidly and exhibit the molecular signature of positive (adaptive) selection (Swanson et al., 2001b, Swanson et al., 2003b).

Rapidly evolving *Drosophila* male accessory gland proteins (Acps) have been foci of molecular population genetic and molecular evolution studies. The average rate of sequence divergence of *D. melanogaster* Acps is approximately twice that of non-reproductive proteins (Begun et al., 2000a, Swanson et al., 2001a-b, Mueller et al., 2005a, Wagstaff & Begun, 2004a). At least seven Acps are transferred to the SSOs after mating (Ravi-Ram & Wolfner, 2005), including a rapidly evolving protease inhibitor. By contrast, female reproductive genes are understudied in *Drosophila*, but the signature of

positive selection has been revealed by evolutionary expressed sequence tag (EST) studies using the lower reproductive tract of both *D. simulans* (Swanson et al., 2004a) and *D. arizonae* (Kelleher et al., 2007). These studies did not investigate the rate of evolution of a broad sample of genes from a specific organ as was done in the present study; only a small proportion of the genes identified in the present study overlap with the previously identified proteins in the lower reproductive tract of the same species (Swanson et al., 2004a). The study presented here is the first molecular evolutionary study of genes sampled from a female SSO in any species.

Drosophila species typically have two types of organs dedicated to sperm storage (Fowler, 1973, Pitnick et al., 1999). The seminal receptacle contains the majority (65 - 80%) of the sperm (Lefevre & Jonsson, 1962, Neubaum & Wolfner, 1999a), while a pair of spermathecae are the site of long term storage. Sperm are stored in the lumen of the spermatheca which receives proteins of unknown function from surrounding secretory epithelial cells (Filosi & Perotti, 1975). Evolutionary interactions have been identified between sperm and SSOs. For example, evolutionary changes in sperm length resulted in corresponding changes in the length of the seminal receptacle (Miller & Pitnick, 2002, Miller & Pitnick, 2003). One rationale for investigating genes in the spermatheca was that rapidly evolving genes in this SSO might coevolve with rapidly evolving *Drosophila* Acps.

The present study is an evolutionary investigation into genes that are enriched for expression in the long term sperm storage organ of *Drosophila*, the spermathecae. This study uses RNA isolated from the spermathecae of *Drosophila simulans* to identify expressed sequence tagged sites (ESTs), allowing for the identification of gene expressed

in the organ at the time of collection. The focus is on spermathecae because, not only are these organs in extended contact with male seminal products during storage, they also secrete proteins into the sperm storage lumen which could interact with male proteins in the female (Acps and sperm proteins). Rapidly evolving proteins in the spermatheca are prime candidates to play an important role in female-ejaculate interactions. The results suggest that a high proportion of spermathecal proteins evolve rapidly. Such proteins include those with secretion signals and thus are capable of directly interacting with male reproductive proteins in this SSO. Female-ejaculate interactions are thought to mediate key features of sperm storage and important evolutionary phenomena.

MATERIALS and METHODS

cDNA library preparation and DNA sequence generation

RNA was isolated from both of the spermathecae, including the spermathecal ducts, dissected from 250 *D. simulans* females. The females were held as virgins until the fourth day of adult life when each was paired with a single male. Dissection occurred 3 hours after mating was observed. Total RNA was isolated using the TRIzol reagent (Invitrogen). Total RNA was also purified from female whole bodies minus spermathecae to be used as the driver in subtractive hybridization. cDNA was generated from total RNA using the SMART approach (Zhu et al., 2001). A cDNA library was generated (Evrogen, Moscow) using the suppressive subtraction hybridization (SSH) method in both directions (tester vs. driver and driver vs. tester) (Diatchenko et al., 1996, Diatchenko et al., 1999). An aliquot of the library was plated and 384 colonies were used for DNA template generation by rolling circle amplification using TempliPhi (Amersham

Biosciences). 383 DNA sequences were generated using the MegaBACE 400 automated DNA sequencer (Amersham Biosciences). Vector sequences were masked and the sequences assembled into 383 contigs (expressed sequence tags, ESTs) using the CAP3 program (Huang & Madan, 1999).

Identification of genes expressed in spermathecae

To ensure that all of the genes represented by the *D. simulans* spermathecae EST sequences and been found, we queried both the *D. simulans* and *D. melanogaster* genomes. The *D. melanogaster* genome provides a more complete genome, making it less likely to contain annotation errors. Each of the 383 *D. simulans* ESTs was used as a query in a blastn DNA similarity search (Altschul et al., 1990) conducted against the entire CDS sets of *D. simulans* and *D. melanogaster*. Sequences were excluded if the calculated similarity values were 80% or lower or if their expected (E) values were greater than 0.01.

Ortholog identification

Using each of the *D. simulans* CDSs obtained above as a query, a blastp protein similarity search (Altschul et al., 1990) was performed to identify ortholog candidates from five additional *Drosophila* genomes (*D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*). This similarity search was also performed with *D.meloanogster* CDSs as the query to ensure that all identifiable ortholog candidates were found. The Comparative Analysis Freeze 1 (CAF1) genomic sequences of the *Drosophila* species were downloaded from the AAAWiki web site

(http://rana.lbl.gov/drosophila) for the 12 *Drosophila* genome project. The entire set of coding sequences of *D. melanogaster* was obtained from FlyBase (Release 5.1; http://flybase.org). The top hit from each species was then used as a query and reciprocal blastp search was performed against the entire *D. simulans* CDS set to confirm the orthologous relationships. When multiple sequences were identified with almost identical lowest E-values, all were used as the queries for the reciprocal blastp search. After examining the results of the reciprocal search, ortholog candidates from each species were identified for each of the *D. simulans* genes. To determine the presence or absence of possible distant orthologs in other species, reciprocal blast was performed against an additional five (more-distantly related) *Drosophila* genomes (*D. persimilis, D. willistoni, D. mojavensis, D. virilis,* and *D. grimshawi*). In addition to blastp, tblastn against these DNA scaffolds was also used.

Species-specific duplications

Our orthologous gene set was compared to the list of homologs provided by the 12 *Drosophila* genome project (http://rana.lbl.gov/~venky/AAA/freeze_20061030/protein_coding_gene). As the list provided by the genome project identified only homolog candidates regardless of whether a gene was an ortholog or paralog, we confirmed that all of our ortholog candidates were included among their homolog candidates. If two or more genes were identified as the top hits with almost identical E-values, then these genes were analyzed as possible duplicates by further investigation including DNA and protein phylogenetic analysis to identify paralog/ortholog relationships.

Reconstruction of multiple alignments from orthologous gene sets

We first reconstructed protein alignments using MUSCLE (Edgar, 2004), and each alignment was adjusted manually. Protein alignments were reverse translated to nucleotide alignments based on their nucleotide sequences using the protal2dna web server (http://bioweb.pasteur.fr/seqanal/interfaces/protal2dna.html). Each nucleotide alignment was again adjusted manually. Finally, the nucleotide alignments were translated to protein alignments for confirmation. The final nucleotide alignments were used for molecular evolutionary analyses.

Evolutionary analyses

The relative contribution of nonsynonymous (d_N) and synonymous (d_S) changes to the patterns of nucleotide variation was compared using the codon-based maximum-likelihood framework described by Goldman and Yang (Goldman & Yang, 1994) which is implemented in the program PAML (phylogenetic analysis of maximum likelihood) (version 3.15) (Yang et al., 2000). A pairwise comparison was performed between the 42 orthologs of *D. simulans* and *D. melanogaster*. The likelihood of d_N being higher than d_S was evaluated by comparing a model where d_N and d_S were estimated as free parameters (L_1) to a model where d_N equals d_S (L_0). The two models were compared in a likelihood ratio test with one degree of freedom. Historically the signature of adaptive evolution is defined as the gene having a d_N/d_S ratio > 1. Ratios equal to one are indicative of netural evolution; equal numbers of relative synonymous and nonsynonymous substitutions are presumably due to a lack of functional constraints. Ratios less than one indicate fewer relative nonsynonymous substitutions due to the processes of purifying (negative

selection). In a study by Swanson et al. 2004, it was found that when looking along the entire length of a gene finding $d_N/d_S > 1$ is rare since, presumably only a subset of the codons are subjected to positive selection. Using a survey of literature, Swanson (2004) found that when lowering the d_N/d_S threshold to 0.5, 95 % of genes had statistical evidence for positive selection in a subset of their codons. Given that the purpose of the current study is to identify candidate genes which are subject to positive selection, possibly in only a subset of their codons, we adapted the lowered d_N/d_S threshold of 0.5 to be indicative of positive selection.

In addition to the pairwise comparison, we explored variation in the d_N/d_S ratio among sites using the tree based models described by Yang *et al.* (Yang et al., 2000) based on the alignment of *D. simulans*, *D. melanogaster*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae* and *D. pseudoobscura* orthologs. The accuracy and power of PAML models increase with more sequences and longer length (Anisimova et al. 2001). PAML was run using the maximum number of orthologs possible; a minimum of four orthologs were used to circumvent problems caused by model convergence. The assumptions of the models and test statistics are briefly described, for a full description see Yang *et al.* (Yang et al., 2000). PAML was used to determine estimates of d_N/d_S for models of varying complexity; six models were used in this study (M0, M1a, M2a, M7 and M8). M0 and M3 are discrete models in that they assign codons to populations of distinct d_N/d_S values; d_N/d_S can be any value greater than 0. M0 assigns all sites to a single d_N/d_S ratio whereas M3 assigns codons to three categories of sites with d_N/d_S free to vary for each site. M1 assigns codons to one of two site classes, $d_N/d_S >1$ and $d_N/d_S = 1$. M2a is similar to M1a, but adds an additional class of sites with $d_N/d_S >1$. M7 assigns

codons using 10 categories of sites with 10 d_N/d_S ratios in the range of 0 – 1 from a discrete approximation of the β distribution. M8 is similar to M7 with the addition of a category of sites with a d_N/d_S ratio that is free to vary from 0 to >1. Each PAML model generates a log likelihood indicating how well the models fit the input data. Since PAML models are nested within each other (M0 with M3; M1 within M2; M7 within M8) twice the log likelihood difference between the two models is compared with a χ^2 distribution with degrees of freedom equal to the difference in degrees of freedom between the two models. *P* values for sites potentially under positive selection are obtained using a Baysian approach in PAML.

Transmembrane and signal peptide prediction and functional domain detection

Protein sequences from *D. melanogaster* orthologs were used for motif prediction. Transmembrane (TM) region prediction was conducted using two programs: HMMTOP (version 2.0; (Tusnady & Simon, 2001) and Phobius (Käll et al., 2004). Both methods use hidden Markov models for predicting the transmembrane topology. Phobius combines TM prediction and signal peptide prediction in order to identify signal peptides from N-terminal regions, often misidentified as a TM region by TM prediction methods. We list a protein as having a transmembrane domain if both HMMTOP and Phobius predicted TM regions, or if one program predicted more than one TM region. For signal peptide prediction, we used TargetP version 1.1 (Emanuelsson et al., 2007) in addition to Phobius. The TargetP program ranks support for the signal peptides. Only the genes in the highest class of support, which were also identified as having signal peptides by Phobius, were listed as having signal peptides.

Functional categories

The function of each gene was inferred from a combination of information gained from methods including: conserved domain searches, FlyBase classification, Gene Ontology database classification, and searches of relevant literature. Conserved domain searches by CD-Search at National Center for Biotechnology Information (Marchler-Bauer & Bryant, 2004) (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml), compared the genes discovered in the EST library against a database of protein modules and domains in order to identify regions of conservation. Conserved domain searches are useful for identifying regions of the genes which may be functionally important, as well as identifying protein domains specific to classes of proteins in order to identify punitive functions of unclassified genes. Gene Ontology (GO) database (http://www.geneontology.org) is a database of genes described by a controlled vocabulary in terms relating to cellular components, biological processes and molecular functions. GO allows for a way to provide consistent descriptions of gene products. GO terminology is utilized by FlyBase (http://flybase.bio.indiana.edu/), a website dedicated to gene and genomic information for Drosophila.

RESULTS

Coding sequences in the cDNA library

Of the 383 EST library clone sequences, 244 matched coding domain sequences (CDSs) in the *D. simulans* genome representing 44 unique CDSs. This number is consistent with to the number of genes typically discovered in EST studies of male accessory gland genes (see Discussion for details). The remaining 139 EST sequences

had no significant matches in either the *D. simulans* or *D. melanogaster* genome and were excluded from further analysis because they did not meet the criteria described in Materials and Methods. The matched sequences found against excluded ESTs were very short, typically less than 20 base pairs, and had E values greater than 0.01 (accepted sequences had an average E value of 4.5×10^{-6}).

Orthologs and functional categories

Sequence similarity was used to identify orthologs in seven *Drosophila* genomes: *D. simulans, D. melanogaster, D. sechellia, D. yakuba, D. erecta, D. ananassae, and D. pseudoobscura*. Of the 44 genes, 30 had identifiable orthologous genes in all seven species. For the remaining 14 genes, orthologs were found in some, but not all of the seven species. For two of 44 genes (dsim_GLEANR_6594 and dsim_GLEANR_15604) there were no orthologous genes shared by *D. simulans* and *D. melanogaster*. These genes were not used for molecular evolutionary analysis because of the lack of identifiable orthologs. The method used to identify orthologs was comprehensive and it sometimes revealed annotation errors. For example, two *D. simulans* genes (dsim_GLEANR_16297 and 17130) appeared to be an incorrect annotation of two exons from a larger gene based on the gene structure of the *D. melanogaster* ortholog (CG32702). Therefore, these two exons were combined into one gene for *D. simulans* as well as in *D. erecta* and *D. ananassae* (dere_GLEANR_3759 and 3760, and dana_20818 and 20819).

Gene function was investigated for all genes having orthologs present in both *D. simulans* and *D. melanogaster* (42 genes in total; Table 2.1; Supplementary Table 2.1).

The most likely function was determined by the Gene Ontology database, conserved domains, and relevant literature. Of the 42 genes, 11 are putative serine proteases each of which has the catalytic triad of amino acids typical of active serine proteases. Expression Analysis Systematic Explorer (EASE) integrated into the DAVID bioinformatics database (http://david.abcc.ncifcrf.gov/home.jsp) indicated that proteases were significantly (P-value 4.5 x 10^{-9}) overrepresented in the spermathecae (26%) when compared to the percentage of such genes in the entire *D. melanogaster* genome (5%) (Ross et al., 2003).

Evolutionary analysis

Non-synonymous and synonymous substitution rates were determined for each of the 42 genes. One analysis was a pairwise comparison between *D. simulans* and *D. melanogaster.* The rates of synonymous and nonsynonymous substitutions were calculated by a maximum-likelihood method using PAML (Yang, 2007, Yang, 1997). The average d_N/d_S ratio from the pairwise comparisons between the *D. simulans* and *D. melanogaster* sequences is 0.269 ±0.2932, with an average d_N of 0.032 and an average d_S of 0.119. In this comparison, 10 out of the 42 genes (Figure 1; Supplementary Table 2.1) have d_N/d_S higher than the 0.5 threshold adopted by Swanson *et al* (Swanson et al., 2004a).

Another analysis compared the fit of the data to different models of codon evolution (Yang & Nielsen, 2000). PAML models were used to explore heterogeneity in d_N/d_S along the gene, and to test for positive selection (Table 2.2; Supplementary Table 2.1). These comparisons were restricted to the 37 genes for which sequences were

available from at least four species. The first comparison examined heterogeneity along the length of the gene by comparing the fit of data to the one-ratio model (M0) against a model that classifies sites into 3 classes (M3). For 27 of the 38 genes, the fit of data to M3 was significantly better than M0, indicating heterogeneity of evolutionary rates along the gene for a high percentage of the genes analyzed. Direct tests of positive selection were also performed. Briefly, two likelihood ratio tests (LRTs) were used to compare null models that do not allow $d_N/d_S > 1$, M1a and M7 , with alternative models that allow a class of sites to have $d_N/d_S > 1$, M2a and M8 (Yang & Nielsen, 2002, Yang & Swanson, 2002). In the first test the null model (M1a) assumes two site classes, the first with $d_N/d_S < 1$, and the second with $d_N/d_S = 1$; this is compared with the alternative model (M2a) which adds a class of sites with $d_N/d_S > 1$. The second test uses M7 as the null model, where d_N/d_S estimates are drawn from a beta distribution with $0 \leq d_N/d_S \leq 1$, with the alternative model M8, which adds a class of sites with $d_N/d_S > 1$. If the LRTs were significant, positive selection was inferred (Yang & Nielsen, 2002, Yang & Swanson, 2002). Comparisons of M7 to M8 and M1a to M2a provided evidence for positive selection in 14 of 37 genes. Genes having elevated d_N/d_S from the pairwise comparison, and/or support for positive selection from PAML models, are listed in Table 2.2 (for complete data on all genes see Supplementary Table 2.1).

In order to obtain extended taxonomic insight into the evolution of spermathecal genes, the presence or absence of homologous genes was examined by comparing *D. melanogaster* to 11 sequenced genomes (*D. simulans, D. sechellia, D. yakuba, D. erecta, D. ananassae, D. pseudoobscura, D. persimilis, D. willistoni, D. mojavensis, D. virilis,* and *D. grimshawi)* (Supplementary Table 2.1). A number of genes, including predicted

proteases, a peptidase, an actin biosynthesis gene, Drosomycin, and a gene of unknown function, were not detectable in taxa distant from the *melanogaster* subgroup (*D. mojavensis, D. virilis,* and *D. grimshawi*). A high percentage of the genes with undetectable orthologs in more distantly related species encode serine proteases (see Discussion). One of the two *D.simulans* genes (dsim_GLEANR_6594) discovered in the library, but having no *D. melanogaster* ortholog, is predicted to be a serine protease by the use of conserved domain searches. Orthologs to this gene were also found in *D. sechellia*, and *D. yakuba*.

Possible species-specific duplications were found in *D. erecta*, *D. ananassae*, and *D. pseudoobscura.* Duplicated copies in *D. erecta* (dere_GLEANR_9251, dere_GLEANR_13114, dere_GLEANR_16834) and in *D. ananassae* (dana_GLEANR_20091 and dana_GLEANR_20093) share 80 and 90% similarities, respectively, against their corresponding *D. simulans* genes. A pair of *D. ananassae* genes (dana_GLEANR_9014 and dana_GLEANR_10165) was almost identical (only one nucleotide difference) to a pair of *D. pseudoobscura* genes (dpse_GLEANR_6308 and dpse_GLEANR_6306). Further investigation would be needed to determine if these apparent duplications are due to artifacts such as assembly mistakes.

Secretion signal sequence and transmembrane region prediction

Of the 42 genes examined, 20 (47%) are predicted to have signal peptides and 12 (29%) predicted to have transmembrane regions (Table 2.1; Supplementary Table 2.1). All of the proteases have predicted secretion signal sequences.

DISCUSSION

Rapidly evolving reproductive proteins are candidates to play an important role in sexual selection and speciation. In order to identify candidate genes that could play a role in these evolutionary processes, the molecular evolution of genes expressed in the spermatheca was analyzed and likely gene function characterized. A high proportion of spermatheca genes are predicted to encode serine proteases. Many of the serine proteases discovered in this study evolve rapidly, and all have secretion signals. Serine proteases expressed in the spermatheca are prime candidates to participate in evolutionarily dynamic interactions with male seminal products. Overall, a very high percentage of the genes in the spermatheca are rapidly evolving and a very high proportion exhibit the molecular signal of positive selection based on d_N/d_S ratios and PAML analysis. Insight into the function of the spermatheca was obtained from the identity of genes expressed in this SSO.

Although spermathecae are found in a wide range of invertebrate and vertebrate taxa (Eberhard, 1996), the function of proteins and other macromolecules associated with this organ are understudied. An exception is social insects - bees and ants - in which the spermatheca are known to be important for long term sperm storage. Reproductively capable females (queens) mate with several males early in life and rely on stored sperm to fertilize eggs over many years. Since the reproductive success of the queen is directly dependent on the availability and viability of stored sperm, there is strong selection for both sperm viability as well as efficient utilization of sperm stores (Baer et al., 2006). In a social ant (*Crematogaster opuntiae*), as well as in honey bees, the secretory cells surrounding the spermatheca have ample glycogen, serving as a possible energy reserve

for sperm in storage (Wheeler & Krutzsch, 1994). In honeybees the spermathecal fluid contains sugars including glucose, trehalose, and fructose, as well as a high level of trehalase activity (Alumot et al., 1969a). Enzymes with antioxidant activity have been found in the spermathecae of honeybees, presumably acting to protect stored sperm from oxidative damage (Weirich et al., 2002, Collins et al., 2004c).

In the present study, 44 unique genes were identified in the hybrid-selected cDNA library. The small number of genes found in this study is quite similar to the number of genes found in comparable studies in *Drosophila* and other species. Multiple studies, using cDNA hybridization or EST sequences, were needed to raise the level of Acp genes *Drosophila melanogaster* to a total of 57 genes (DiBenedetto et al., 1987, Monsma & Wolfner, 1988, Wolfner et al., 1997, Swanson et al., 2001a-b). EST studies of the accessory glands of mosquito *Anopheles gambiae* yielded 46 genes, and similar studies in crickets (*Gryllus firmis* and *Gryllus pennsylvanicus*) yielded 30 genes (Andres et al., 2006). Using Mediterranean fruit flies (*Ceratitis capitata*), DNA sequence analysis of a subtractive hybridization cDNA library from RNA isolated from the male accessory gland resulted in identification of 13 unique genes (Davies & Chapman, 2006).

It is informative to compare the evolution of the 42 spermatheca genes identified in this study with relevant previous studies. Based on the pairwise comparison between *D. simulans* and *D. melanogaster*, 24% of the spermatheca genes have an overall $d_N/d_S > 0.5$ (Figure 1). Two especially relevant previous studies (Swanson et al. 2001a; Swanson et al. 2004) have made pairwise comparisons of *D. simulans* and *D. melanogaster*. The incidence of genes with $d_N/d_S > 0.5$ in the spermatheca (present study) is appreciably higher than that observed in the female reproductive tract of *D. melanogaster*, minus

ovaries, in which 6% of the genes had a $d_N/d_S > 0.5$ (Swanson et al., 2004a) and at least as high as that observed for male accessory gland genes among which 19% of genes had a $d_N/d_S > 0.5$ (Swanson et al., 2001a-b). The number of genes in the present study that overlapped with the most similar study (Swanson et al. 2004) was only 5, even though the Swanson study included spermathecae in the mix of tissue investigated. This suggests the possibility that a large number of genes (especially those of lower expression) were missed when using the entire lower reproductive tract. Of the 10 spermathecal genes with elevated d_N/d_S (> 0.5), the most rapidly evolving was a gene of unknown function ($d_N/d_S = 0.95$), followed by 5 serine proteases ($d_N/d_S = 0.72 – 0.89$), and two more genes of unknown function ($d_N/d_S = 0.64 – 0.66$). The average rate of sequence divergence of *D. melanogaster* Acps is approximately twice that of non-reproductive proteins (Begun et al., 2000a, Swanson et al., 2001a-b, Mueller et al., 2005a, Wagstaff & Begun, 2004a). Acp male genes are notable for their rapid rates of evolution and it appears that female spermatheca genes are similarly evolutionarily dynamic.

The analysis of the pattern of molecular evolution among a larger set of related species is also informative. Molecular evolutionary analyses showed that 14 of 37 (37.8%) spermatheca genes contain at least one region that conforms to a model of positive selection (Table 2.2; Supplementary Table 2.1). Categories of genes showing evidence for positive selection include serine proteases, cell communication, translation, sugar metabolism, peptidase activity, protein-protein interaction and genes of unknown function (Table 2.2; Supplementary Table 2.1). The proportion of positively selected spermathecae genes can be compared to the molecular evolution rates of *Drosophila melanogaster* seminal fluid proteins (Haerty et al., 2007). Twenty five seminal fluid

genes had orthologs in all of the *melanogaster* subgroup species and 4 of these genes (16%) exhibited positive selection by the criteria of acceptance of M8 over M7. Acceptance of M8 over M7 indicates that the data better fit a models which includes a class of sites that allow $d_N/d_S > 1$ (positively selected). Using the same criteria in female Drosophila, of 679 genes in the reproductive tract, and of 9921 genes not related to sex or reproduction, 6.2% and 6.0% respectively were consistent with the hypothesis of positive selection by acceptance of model 8 (Haerty et al., 2007). Overall, a much higher proportion of genes enriched for expression in the spermatheca showed evidence for positive selection (37%) compared to seminal fluid proteins (16%) or a collection of genes in the female reproductive tract (6.2%). The incidence of directional selection among spermathecal genes is striking.

In the present study, serine proteases are the predominant category of genes which are lack detectable orthologs as a function of evolutionary distance. Among spermatheca proteases, four have lost orthologs in species belonging to the *melanogaster* subgroup and five protease genes have no detectable ortholog in the obscura subgroup. At the level of differentiation between *D. melanogaster/D. simulans* and the repleta group, seven protease genes have no orthologs. Between the *melanogaster* subgroup and a Hawaiian *Drosophila*, 10 protease genes have no orthologs. In other studies using *Drosophila* species, reproductive system proteases show evidence of accelerated and positive evolution (Kelleher et al., 2007, Swanson et al., 2004a, Kern et al., 2004, Panhuis & Swanson, 2006, Lawniczak & Begun, 2007, Wong et al., 2007), indicating that the evolutionary patters for this group of proteins is dynamic.

All of the spermatheca serine proteases have secretion signals (Table 2.1; Supplementary Table 2.1) and are possibly secreted into the lumen of this SSO. Potential roles for male and female proteases are discussed in Ravi-Ram and Wolfner (Ravi-Ram & Wolfner, 2007). Spermathecal proteases may be involved in interactions with male reproductive proteins, or play roles functionally analogous to male reproductive proteins. Previous studies have described at least two *Drosophila* male proteins that are transferred to females and undergo cleavage within the female reproductive tract, perhaps as a mechanism to control activity levels of the proteins (Monsma et al., 1990, Bertram et al., 1996, Ravi-Ram & Wolfner, 2007). Female proteases might act to control the viscosity of the internal milieu of the lumen of the spermatheca analogous to the semen coagulation role played by the primate prostate specific antigen (PSA) in males (Malm et al., 2000). PSA is a serine protease and its role in humans suggests an analogous function for spermathecal proteases in *Drosophila*.

An Acp protein that plays a key role in sperm storage (Acp36DE) is found in the spermatheca after mating and it is rapidly evolving. Moreover a protease Acp associated with regulation of sperm use is also evolving rapidly (Wong et al., 2007) and it also is a candidate for a coevolutionary interaction with spermatheca proteases based on direct interaction. As an exciting possibility, male derived protease inhibitors might inhibit female proteases secreted into the lumen spermatheca in a specific male-female (ejaculate-female) molecular interaction. Seven protease inhibitors have been reported among Acps of *D. melanogaster*. Acp 62F, which is able to transverse the female reproductive tract and enter the hemolymph, is toxic upon ectopic expression; this Acp is present in the spermatheca after mating (Lung et al., 2002b). The evolutionary

importance of the interactions which occur between females proteins and male ejaculates are being increasingly recognized (Pitnick et al., in prep.). Spermathecal proteases are prime candidates to be involved in evolutionarily dynamic interactions with male reproductive proteins such as protease inhibitors and proteases.

Four of the protease genes identified in the present study are found in a cluster on chromosome arm 2L. These genes exhibit approximately 30% sequence similarity to each other in *D. melanogaster* and each gene is approximately 90% similar to its ortholog in *D. simulans*. They have no introns and they encode proteins with the canonical serine protease catalytic triad of amino acids. The cluster of proteases has been found to be transcriptionally activated by mating (Lawniczak & Begun, 2007). These proteases, and several others, are rapidly evolving between populations of *D. melanogaster* and diverging between *D. melanogaster* and *D. simulans.* (Lawniczak & Begun, 2007). Five of the proteases found in the present study (CG18125, and the cluster on chromosome II) have been foci for previous molecular population genetic and molecular evolution studies. These studies showed that the sites of molecular changes in these proteases were associated with the active site, suggesting the evolution of functional changes related to catalysis (Panhuis & Swanson, 2006, Lawniczak & Begun, 2007). This rapid divergence can further be exemplified by the two spermatheca genes found in *simulans* without a *melanogaster* ortholog (see Results). One of these genes, dsim_GLEANR_6594, is found in the middle of the four clustered proteases in the *D. simulans* genome, and is predicted to be a serine protease based on conserved domains. A large corresponding portion of this region is missing from the *D. melanogaster* genome, which provides an explanation for

the lack of ortholog found in *D. melanogaster* and yields a picture of rapid change between the two genomes.

One class of spermathecal proteins identified in this study contains at least one protein-protein interaction motif called a CUB domain. CUB domains, which consist of approximately 110 amino acids with four positionally conserved cysteines, (Bork & Beckman, 1993) play a variety of roles including interaction with sperm in both vertebrate and invertebrate taxa (Haley & Wessel, 2004, Kamei & Glabe, 2003). CUB domains bind other proteins with high specificity (Song et al., 2006) and tend to exist as a cluster of multiple repeats along the length of a single gene. A gene identified in the present study, CG32702, contains approximately 20 CUB domains in one region of the protein, along with a repeat of five EGF-CA like domains at the C-terminal end (Figure 2.2). A second gene (CG30371) encodes a trypsin-like serine protease domain and a motif that is 67% similar to a CUB domain. CG32702 (the gene with many CUB domains) exhibits evolutionary stasis in much of the gene, but relatively rapid evolution in some regions of the gene. Having multiple CUB domains potentially allows for a relaxation of selective constraints. Changes could be tolerated in a subset of the domains because the original specificity may be retained by the remaining (unchanged) domains. A general argument about redundancy and relaxation of selective constraints when repeated motifs are present in a protein has been made by Metz and Palumbi (Metz & Palumbi, 1996) and is used to interpret the evolution of VERL domains in reproductive proteins (Swanson & Vacquier, 1998). There is evidence for positive selection in regions of the CUB protein even though the protein is sufficiently conserved to be found in all 12 sequenced *Drosophila* genomes.

Other genes encoding proteins with potentially important roles associated with sperm storage and maintenance were identified. Trehalase activity (sugar metabolism gene in Table 2.1) could play a role in sperm nutrition. The *Drosophila* trehalase RNA encodes a predicted secretion signal and thus its protein could be active in the lumen of the spermatheca. A gene encoding an antifungal defense peptide (*Drs*) also was identified in the present study. This gene is not spermatheca-specific, it is constitutively expressed in both types of SSOs of *D. melanogaster* (Ferrandon et al., 1998). The SSOs are apparently the only site of constitutive expression whereas the gene is expressed in many locations after induction with a pathogen (Ferrandon et al., 1998). Juvenile hormone epoxide hydrolase 3 (JHEH3) is an example of a gene that could play an interesting role in evolution. This enzyme catabolizes juvenile hormone (JH) to an inactive metabolite. Its protein product is predicted to have a secretion signal and six transmembrane domains suggesting it could be a receptor. It is possible that JHEH3 could be acting to control JH levels in the spermatheca and as a systemic hormone regulator if it is secreted into the hemolymph. In *D. melanogaster*, Acp70 is transferred to females at the time of mating stimulates juvenile hormone synthesis (Peng et al., 2005a). An enzyme that produces a precursor of juvenile hormone is elevated in the lower female reproductive tract after mating (Mack et al., 2006a). The presence of enzyme activity that degrades juvenile hormone (JHEH3) in the long-term SSO is intriguing because it might oppose the male effect of stimulating the synthesis of JH. Genes with unknown function in *Drosophila* or other species (Table 2.1) might be quite interesting in terms of having spermatheca - specific roles because functions for such genes have not been identified in other tissues or taxa. Four of these genes have transmembrane domains which could be receptors having

spermatheca-specific function. These receptors could potentially interact with male accessory gland proteins or other proteins found on sperm. Identification of such receptors would be important for understanding the evolution of Acps and how they function in females.

CONCLUSIONS

This study has produced insight into the evolution and function of genes enriched for expression in *Drosophila* spermathecae. We find that genes expressed in the spermatheca evolve as rapidly as genes in the male accessory gland. Importantly, the proportion of genes with the overall signature of positive selection is higher than Acp genes which are a paradigm for rapid evolution. Rapidly evolving spermatheca proteins of established and novel function could participate in female reproductive molecule-ejaculate interactions which are increasingly recognized as evolutionarily important (Pitnick et al., in prep., Ravi-Ram & Wolfner, 2007).

ACKNOWLEDGMENTS

We thank Mariana Wolfner, Alex Wong, Tony Zera and Andy Clark for comments on the manuscript and other contributions to the research. This research was supported by a National Science Foundation grant (DEB-ESP0346476).

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic local alignment search tool. *J Mol Biol* **215**: 403-10.

Alumot, E., Lensky, Y. & Holstein, P. 1969. Sugars and trehalase in the reproductive organs and hemolymph of the queen and drone honey bees (Apis mellifera L.Var. Ligustica Spi.),. *Comp Biochem Physiol B* **28**: 1419-25.

Andres, J. A., Maroja, L. S., Bogdanowicz, S. M., Swanson, W. J. & Harrison, R. G. 2006. Molecular evolution of seminal proteins in field crickets. *Mol. Biol. Evol.* **23**: 1574-84.

Baer, B., Armitage, S. A. O. & Boomsma, J. J. 2006. Sperm storage induces an immunity cost in ants. *Nature* **441**: 872 - 76.

Begun, D. J., Whitley, P., Todd, B. L., H.M., W.-D. & Clark, A. G. 2000. Molecular population genetics of male accessory gland proteins in *Drosophila*. *Genetics* **156**: 1879 - 88.

Bertram, M. J., Neubaum, D. M. & Wolfner, M. F. 1996. Localization of the *Drosophila* male accessory gland protein Acp36DE in the mated female suggests a role in sperm storage. *Insect Biochem Mol Biol* **26**: 971-80.

Bork, P. & Beckman, G. 1993. The CUB domain. A widespread module in developmentally regulated proteins. *J. Mol. Biol*: 539-45.

Clark, A. G., Aguade, M., Prout, T., Harshman, L. G. & Langley, C. H. 1995. Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. *Genetics* **139**: 189-201.

Clark, A. G. & Begun, D. J. 1998. Female genotypes affect sperm displacement in *Drosophila*. *Genetics* **149**: 1487-93.

Clark, A. G., Begun, D. J. & Prout, T. 1999. Female x male interactions in *Drosophila* sperm competition *Science* **283**: 217-220.

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

Collins, A. M., Williams, V. & Evans, J. D. 2004. Sperm storage and antioxidative enzyme expression in the honey bee, Apis mellifera. *Insect Mol Biol* **13**: 141-146.

Coyne, J. A. & Orr, H. A. 2004. *Speciation*. Sinauer Associates, Sunderland, Massachusetts.

Darwin, C. 1871. *The Descent of Man and Selection in Relation to Sex*. London.

Davies, S. J. & Chapman, T. 2006. Identification of genes expressed in the accessory glands of male Mediterranean Fruit Flies (Ceratitis capitata). *Insect Biochem Mol Biol* **36**: 846-56.

Diatchenko, L., Lau, Y., Campbell, A., Chenchik, A., Mogadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. & Siebert, P. 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci USA* **93**: 6025-30.

Diatchenko, L., Lukyanov, K., Lau, Y. & Siebert, P. 1999. Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes. *Method Enzym* **303**: 349-80.

DiBenedetto, A. J., Lakich, D. M., Kruger, W. D., Belote, J. M., Baker, B. S. & Wolfner, M. F. 1987. Sequences expressed sex-specifically in *Drosophila melanogaster* adults. *Dev Biol* **119**: 242-51.

Eberhard, W. G. 1996. *Female control: sexual selection by cryptic female choice*. Princeton University Press, New Jersey.

Edgar, R. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-97.

Emanuelsson, O., Ren-Brunak, S., vonHeijne, G. & Nielsen, H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**: 953-71.

Ferrandon, D., Jung, A., Criqui, M., Lemaitre, B., Uttenweiler-Joseph, S., Michaut, L., Reichhart, J. & Hoffman, J. 1998. A drosomycin-GFP reporter transgene reveals a local immune response in *Drosophila* that is not dependent on the Toll pathway. *EMBO* **17**: 1217-27.

Filosi, M. & Perotti, M. 1975. Fine structure of spermatheca of *Drosophila melanogaster* Meig. *J. Submicr. Cytol* **7**: 259-70.

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**: 243-57.

Fiumera, A. C., Dumont, B. L. & Clark, A. G. 2007. Associations between sperm competition and natural variation in male reproductive genes on the third chromosome of *Drosophila melanogaster*. *Genetics* **176**: 1245-60.

Fowler, G. 1973. Some aspects of the reproductive biology of *Drosophila*: sperm transfer, sperm storage and sperm utilization. *Adv in Genet* **17**: 293-360.

Galindo, B., Vacquier, V. & Swanson, W. 2003. Positive selection in the egg receptor for abalone sperm lysine. *Proc Natl Acad Sci USA* **100**: 4639-43.

Goldman, N. & Yang, Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* **11**: 725-36.

Haerty, W., Jagadeeshan, S., Kulathinal, R. J., Wong, A., Ravi Ram, K., Sirot, L. K., Levesque, L., Artieri, C. G., Wolfner, M. F., Civetta, A. & Singh, R. S. 2007. Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics* **177**: 1321-35.

Haley, S. & Wessel, G. 2004. Proteolytic cleavage of the cell surface protein p160 is required for detachment of the fertilization envelope in the sea urchin. *Dev. Biol* **272**: 191-202.

Harshman, L. G. & Prout, T. 1994. Sperm displacement without sperm transfer in *Drosophila melanogaster*. *Evolution* **48**: 758-766.

Howard, D. J. 1999. Conspecific sperm and pollen precedence and speciation. *Ann. Rev. Ecol. Syst.* **30**: 109-132.

Howard, D. J., Palumbi, S. R., Birge, L. & Manier, M. K. (in prep) Sperm and speciation. In: *Sperm Biology: an Evolutionary Approach*, (Birkhead, T. R. & Hosken, D. J., eds.). pp. Elsevier Press.

Käll, L., Krogh, A. & Sonnhammer, E. L. L. 2004. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol* **338**: 1027-36.

Kamei, N. & Glabe, C. 2003. The species-specific egg receptor for sea urchin sperm adhesion is EBR1, a novel ADAMTS protein. *Genes and Devel.* **17**: 2501-07.

Kamei, N., Swanson, W. & Glabe, C. 2000. A rapidly diverging EGF protein regulates species-specific signal transduction in early sea urchin development. *Dev. Biol* **225**: 267-76.

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 Genet* **3**: e148.

Kern, A. D., Jones, C. D. & Begun, D. J. 2004. Molecular population genetics of male accessory gland proteins in the *Drosophila simulans* complex. *Genetics* **167**: 725-35.

Lawniczak, M. K. & Begun, D. J. 2007. Molecular population genetics of female-expressed mating-induced serine proteases in *Drosophila melanogaster. Mol Biol Evol* **24**: 1944-51.

Lefevre, G. & Jonsson, U. B. 1962. Sperm transfer, storage, displacement, and utilization in *Drosophila melanogaster*. *Genetics* **47**: 1719-1736.

Lung, O., Tram, U., Finnerty, M., Eipper-Mains, M., Kalb, J. & Wolfner, M. 2002. The *Drosophila melanogaster* seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression. *Genetics* **160**: 211-14.

Mack, P., Kapelnikov, A., Heifetz, Y. & Bender, M. 2006. Mating-responsive genes in reproductive tissues of female *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **103**: 10358-63.

Malm, J., Hellman, J., Hog, P. & Lilja, H. 2000. Enzymatic action of prostate-specific antigen (PSA or hK3): substrate specificity and regulation by Zn (2+). *Prostate* **45**: 132-39.

Marchler-Bauer, A. & Bryant, S. 2004. CD-Search:protein domain annotations on the fly. *Nucleic Acids Research* **32**: 327-31.

Metz, E. & Palumbi, S. 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Mol Biol Evol* **13**: 397-406.

Miller, G. & Pitnick, S. 2002. Sperm-female co-evolution in *Drosophila*. *Science* **298**: 1230-33.

Miller, G. & Pitnick, S. 2003. Functional significance of seminal receptacle length in *Drosophila melanogaster*. *J. Evol Biol* **16**: 114-16.

Monsma, S. A., Harada, H. A. & Wolfner, M. F. 1990. Synthesis of two *Drosophila* male accessory gland proteins and their fate after transfer to the female during mating. *Dev Biol* **142**: 465-75.

Monsma, S. A. & Wolfner, M. F. 1988. Structure and expression of a *Drosophila* male accessory gland gene whose product resembles a peptide pheromone precursor. *Genes Dev* **2**: 1063-73.

Mueller, J., Ravi-Ram, K., McGraw, M., Bloch-Qazi, M., Siggia, E., Clark, A., Aquadro, C. & Wolfner, M. 2005. Cross-species comparison of *Drosophila* male accessory gland protein genes. *Genetics* **171**: 131-43.

Neubaum, D. M. & Wolfner, M. F. 1999. Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. *Genetics* **153**: 857-869.

Palumbi, S. 1999. All males are not created equal: Fertility difference depend on gamete recognition polymorophisms in sea urchins. *Proc Natl Acad Sci U S A* **99**: 12632-37.

Panhuis, T. M. & Swanson, W. J. 2006. Molecular evolution and population genetic analysis of candidate female reproductive genes in *Drosophila*. *Genetics* **173**: 2039-47.

Peng, J., Chen, S., Busser, S., Liu, H., Honegger, T. & Kubli, E. 2005. Gradual release of sperm bound sex-peptide controls female postmating behavior in *Drosophila*. *Current Biology* **15**: 207-13.

Pitnick, S., Markow, T. A. & Spicer, G. S. 1999. Evolution of multiple kinds of female sperm-storage organs in *Drosophila*. *Evolution* **53**: 1804-22.

Pitnick, S., Wolfner, M. & Suarez, S. (in prep.) Sperm-female interactions. In: *Sperm Biology: An Evolutionary Perspective*, (Birkhead, T., Hosken, D. & Pitnick, S., eds.). pp. Elsevier Press.

Price, C. S. 1997. Conspecific sperm precedence in *Drosophila*. *Nature* **388**: 663-6.

Price, C. S., Kim, C. H., Gronlund, C. J. & Coyne, J. A. 2001. Cryptic reproductive isolation in the *Drosophila simulans* species complex. *Evolution Int J Org Evolution* **55**: 81-92.

Price, C. S. C., Dyer, K. A. & Coyne, J. a. 1999. Sperm competition between *Drosophila* males involves both displacement and incapacition. *Nature* **400**: 449-52.

Ravi-Ram, K. & Wolfner, M. 2005. Fates and targets of male accessory gland proteins in mated female *Drosophila melanogaste*r. *Insect Biochem Mol Biol* **35**: 1059-71.

Ravi-Ram, K. & Wolfner, M. F. 2007. Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduction. *Int. Comp. Biol.*: 19.

Ross, J., Jiang, H., Kanost, M. & Wang, Y. 2003. Serine proteases and their homologs in *Drosophila melanogaster* genome: an initital analysis of sequence conservation and phylogenetic relationships. *Gene* **304**: 117-131.

Song, J., Wong, J. & Wessel, G. 2006. Oogenesis: Single cell development and differentiation. *Dev. Biol* **300**: 385-405.

Swanson, W., Nielsen, R. & Yang, Z. 2003b. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.* **20**: 18-20.

Swanson, W. & Vacquier, V. 1998. Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. *Science* **281**: 710-12.

Swanson, W. & Vacquier, V. 2002a. The rapid evolution of reproductive proteins. *Nat Rev Genet* **3**: 137-144.

Swanson, W., Wong, A., Wolfner, M. & Aquadro, C. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies several genes subjected to positive selection. *Genetics* **168**: 1457-65.

Swanson, W. J., Clark, A. G., Waldrip-Dail, H., Wolfner, M. F. & Aquadro, C. F. 2001a. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc. Natl. Acad. Sci. U S A* **98**: 7375-7379.

Swanson, W. J., Yang, Z., Wolfner, M. F. & Aquadro, C. F. 2001b. Positive selection droves the evolution of several female reproductive proteins in mammals. *Proc. Natl. Acad. Sci. U S A* **98**: 2509-5214.

Tusnady, G. E. & Simon, I. 2001. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**: 849-50.

Wagstaff, B. & Begun, D. J. 2004. Comparative genomics of acessory gland protein genes in *Drosophila melanogaster* and *D. pseudoobscura*. *Mol. Biol. Evol.* **22**: 818-32.

Weirich, G. G., Collins, A. M. & Williams, V. P. 2002. Antioxidant enzymes in the honey bee, *Apis mellifera*. *Apidologie* **33**: 3-14.

Wheeler, D. E. & Krutzsch, P. H. 1994. Ultrastructure of the spermathecae and its associated gland in the and *Crematogaster opuntiae* (Hymenoptera: Formicidae). *Zoomorphology* **114**: 203-214.

Wolfner, M. F., Harada, H. A., Bertram, M. J., Stelick, T. J., Kraus, K. W., Kalb, J. M., Lung, Y. O., Neubaum, D. M., Park, M. & Tram, U. 1997. New genes for male accessory gland proteins in *Drosophila melanogaster*. *Insect Biochem Mol Biol* **27**: 825-34.

Wong, A., Turchin, M. C., Wolfner, M. F. & Aquadro, C. F. 2007. Evidence for Positive Selection on *Drosophila melanogaster* Seminal Fluid Protease Homologs. *Mol Biol Evol*.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**: 555-6.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**: 1586-91.

Yang, Z. & Nielsen, R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *J.Mol.Evol.* **46**: 409-18.

Yang, Z. & Nielsen, R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**: 908-17.

Yang, Z. & Swanson, W. J. 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol Biol Evol* **19**: 49-57.

Yang, Z., Swanson, W. J. & Vacquier, V. D. 2000. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Mol Biol Evol* **17**: 1446-1454.

Zhu, Y., Machleder, E., Chenchik, A. & Siebert, P. 2001. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* **30**: 892-97.

Figure 2.1. The number of nonsynonymous substitutions per site (d_N) plotted against the number of synonymous substitutions per site (d_S) for the *D. simulans* spermathecae EST library. The solid line represents d_N/d_S of 1, and the dashed line represents d_N/d_S of 0.5. All points on the graph represent genes found in the spermathecae EST library and open squares correspond to the eleven serine proteases.

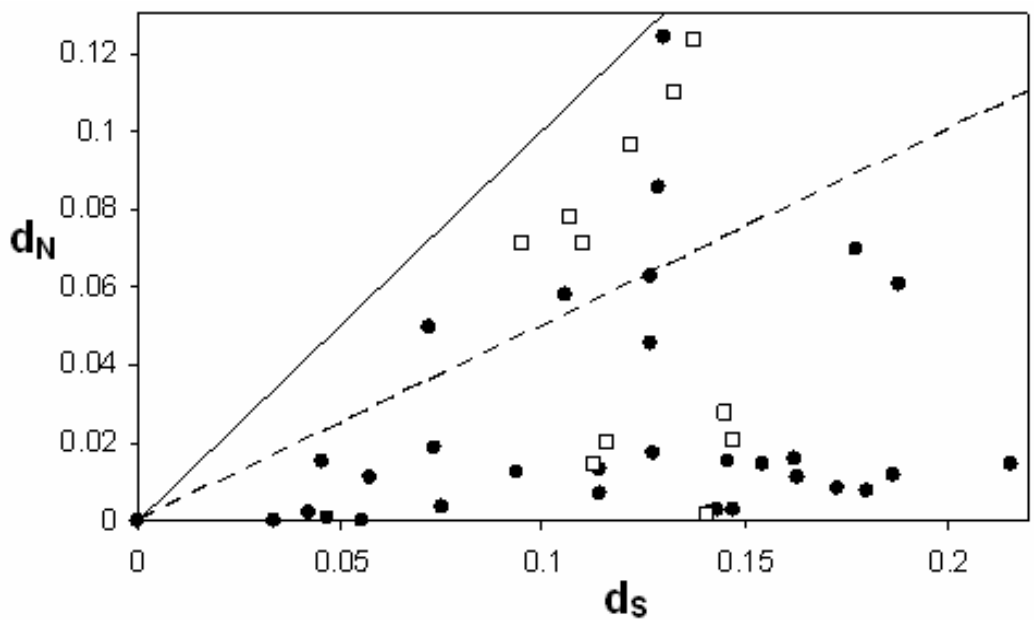

0.12
0.1
0.08
0.06
0.04
0.02
0
d_N
0
0.05
0.1
0.15
0.2
d_S

Figure 2.2. Schematic representation of the CUB domains (putative sperm binding domains) of gene CG32702. This gene is located on the X chromosome. CG32702 is 12,484 bp in length, has twelve introns and a transcript size of 11,275 nucleotides (http://flybase.org). The gene is relatively conserved throughout the twelve *Drosophila* genomes although there are sites that have evolved rapidly. Image obtained from the National Center for Biotechnology Information (Marchler-Bauer & Bryant, 2004).

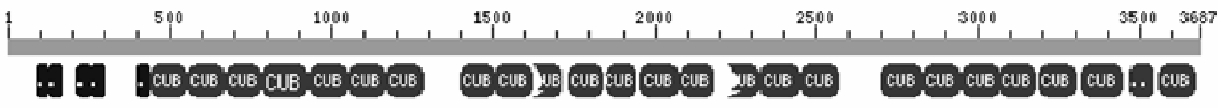

1
500
1000
1500
2000
2500
3000
3500
3687
CUB

Table 2.1: Functional annotation of 42 genes enriched for expression in the spermatheca

Function[1]	Number[2]	SP[3]	TM[4]
Serine protease	11	11	
Cell communication	3		2
Peptidase	3	2	2
Translation	2		
Actin formation/biosynthesis	2		
Amino acid transport	1	1	1
Antimicrobial	1	1	
Apoptosis	1		1
Cation transport	1		1
Dehydrogenase	1	1	
Helicase	1		
Juvenile Hormone Catabolism	1	1	1
Nerve signaling	1		1
Phospholipid metabolism	1	1	1
Secondary metabolism	1		
Protein-Protein interaction	1		
Sugar metabolism	1	1	
Unknown	9	4	3

[1]Predicted function of encoded proteins; [2]Number of genes; [3]Number of genes predicted to encode proteins which have secretion signal peptides; [4]Number of genes predicted to encode proteins with transmembrane regions

Table 2.2: Genes with elevated pairwise d_N/d_S (>0.5) and those identified as having regions which provide evidence for evolution by positive selection (PAML analysis).

				M0 vs. M3		M1 vs. M2		M7 vs. M8	
Genes	**Function[1]**	**d_N/d_S [2]**	**Species [3]**	**p_s[4]**	**d_N/d_S[5]**	**p_s[6]**	**d_N/d_S[7]**	**p_s[8]**	**d_N/d_S[9]**
CG8331	cell comm.	0.19	M,S,Sc,Y,E,A,P	0.21***	0.59	0.01	35.52	0.01**	30.11
me31B	helicase	0.05	M,S,Sc,Y,E,A	0.02***	0.82	0	1.00	0.02*	1.00
CG10650	peptidase	0.55	M,S,Sc,Y,A,P	0.09***	2.97	0.03	5.36	0.05**	3.81
ηTry	ser. protease	0.14	M,S,Sc,Y,E	0.03***	2.69	0.07	1.00	0.03*	2.65
Ser12	ser. protease	0.64	M,S						
CG17012	ser. protease	0.89	M,S,Sc,A	0.07***	6.08	0.05***	8.21	0.06***	6.96
CG17234	ser. protease	0.74	M,S,Sc,Y,E,P	0.06***	5.94	0.10***	4.20	0.20***	2.82
CG17239	ser. protease	0.73	M,S,Sc,Y,E,A	0.16***	2.70	0.09**	3.41	0.13***	2.94
CG18125	ser. protease	0.79	M,S						
CG31681	ser. protease	0.82	M,S						
CG32834	ser. protease	0.19	M,S,Sc,Y,E,A,P	0.09***	2.74	0.02	5.15	0.06*	3.25
CG32702	protein int.	0.14	M,S,Sc,Y,E,A,P	0.06***	1.41	0.05	1.00	0.03***	2.11
Treh	sugar metab.	0.12	M,S,Sc,Y,E	0.05***	2.95	0.04**	3.11	0.05***	2.93
Ef1α48D	translation	0.02	M,S,Sc,Y,E,A,P	0.01**	3.44	0.01*	3.27	0.01***	3.42
Qm	translation	0.02	M,S,Sc,Y,E,A,P	0.01**	1.76	0.0003	1.69	0.01*	1.75
CG2233	unknown	0.95	M,S,Sc,Y,E,A	0.23***	2.64	0.10***	4.18	0.17***	3.10
CG11137	unknown	0.1	M,S,Sc,Y,E,A,P	0.02***	0.84	0.01	1.00	0.02*	1.00
CG15098	unknown	0.69	M,S,Sc,Y,E,A,P	0.09***	2.72	0.05	3.46	0.08***	2.89
CG30197	unknown	0.05	M,S,Sc,Y,E	0.02**	4.37	0.02	4.37	0.02*	4.37
CG31686	unknown	0.66	M,S						

[1]Predicted protein function; [2]d_N/d_S based on pairwise comparison of *D. melanogaster* and *D. simulans* sequences, estimated assuming no rate heterogeneity; [3]Species: refers to the species of *Drosophila* from which sequences were obtained for PAML analysis M= *melanogaster*, S = *simulans*, Sc= *sechellia*, Y = *yakuba*, E = *erecta*; A= *ananassae*; P = *pseudoobscura*. The following statistics are all derived from PAML analysis: [4]p_S: the proportion of sites estimated to belong to the class that has the highest d_N/d_S in M3; [5]d_N/d_S: for the highest class in M3;[6]p_S: the proportion of sites estimated to belong to the class that has $d_N/d_S > 1$ in M2; [7]d_N/d_S: the estimate for the class with the ratio > 1 in M2; [8]p_S: the proportion of sites estimated to belong to the class that has $d_N/d_S > 1$ in M8; [9]d_N/d_S: the estimate for the class with the ratio > 1 in M8; Statistical significance for the LRT: *P <.05; **P<.01; ***P<.001.

Supplementary Table 2.1 : Gene identification numbers for *D. simulans* and identification numbers for *D. melanogaster* corresponding to the 44 genes identified by the spermathecae EST library; Function of predicted proteins; Secretion signal prediction; Transmembrane region prediction; Pairwise d_N/d_S; PAML data for all genes with sufficient number of orthologs;

							M0 vs M3		M1 vs M2		M7 vs. M8		
D.sim ID	**D.mel CG**	**Function[1]**	**SP[2]**	**TM[3]**	**d_N/d_S[4]**	**PAML Species[5]**	**p_s[6]**	**d_N/d_S[7]**	**p_s[8]**	**d_N/d_S[9]**	**p_s[10]**	**d_N/d_S[11]**	**Genome Species[12]**
dsim_GLEANR_16371	CG2233	Unknown	X		0.95	M,S,Sc,Y,E	0.02***	33.31	0.18***	4.16	.17***	4.37	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,P	0.23***	2.64	0.10***	4.18	0.17***	3.10	
dsim_GLEANR_3734	CG3066	Serine Protease	X		0.17	M,S,Sc,Y,E	0.09***	1.45	0.09	1.45	0.09	1.46	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.03***	2.63	0.13	1.00	0.02	2.91	
dsim_GLEANR_9093	CG3831	Unknown		X	0.32	M,S,Sc,Y,E	0.14***	1.54	0.14	1.54	0.14	1.54	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.34***	0.19	0.00	33.13	0.00	1.00	
dsim_GLEANR_7778	CG4214	Nerve signaling		X	0.36	M,S,Sc,Y,E	0.31**	0.54	0.07	1.00	0.00	1.00	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.19***	0.46	0.00	17.20	0.00	1.00	
dsim_GLEANR_8979	CG4254	Actin metabolism			0.00	M,S,Sc,Y,E	0.25	0.00	0.00	1.00	0.00	1.52	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.02**	0.90	0.00	1.00	0.02	1.00	
dsim_GLEANR_2206	CG4370	Cation transport		X	0.01	M,S,Sc,Y,E	0.09***	1.56	0.09	1.56	0.09	1.56	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.04***	2.23	0.00	5.26	0.03	2.85	
dsim_GLEANR_7402	CG4916	Helicase			0.05	M,S,Sc,Y,E	0.01	2.92	0.01	2.92	0.006	2.92	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A	0.02***	0.82	0.00	1.00	0.02*	1.00	
dsim_GLEANR_2360	CG7415	Peptidase		X	0.01	M,S,Sc,Y,E	0.002	2.92	0.00	2.92	0.002	2.93	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.1***	0.44	0.00	1.00	0.00	1.95	
dsim_GLEANR_10797	CG8280	Translation			0.02	M,S,Sc,Y,E	0.003**	11.54	0.00	11.54	0.002*	11.54	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.01***	3.44	0.01*	3.27	0.01***	3.42	
dsim_GLEANR_11000	CG8331	Cell communication		X	0.19	M,S,Sc,Y,E	0.01***	33.67	0.01	34.35	0.01	33.74	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.21***	0.59	0.01	35.52	0.01***	30.11	
dsim_GLEANR_11526	CG9364	Sugar metabolism	X		0.12	M,S,Sc,Y,E	0.05***	2.95	0.04**	3.11	.05***	2.93	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.05***	2.47	0.01	5.44	.03***	2.93	

dsim_GLEANR_11514	CG10067	Actin formation			0.00	M,S,Y,E,A,P	0.37	0.00	0.00	1.00	0.00	1.00	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
dsim_GLEANR_13207	CG10469	Serine Protease	X		0.07	M,S,Sc,Y,E	0.11	0.97	0.04	1.00	0.10	1.00	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.21***	0.64	0.04	1.00	0.11	1.00	
dsim_GLEANR_5555	CG10650	Peptidase	X		0.55	M,S,Sc,Y							M,S,Sc,Y,A,P,P r,W,Mj,V,G
						M,S,Sc,YA,P	0.09***	2.97	0.03	5.36	.05**	3.81	
dsim_GLEANR_13789	CG10810	Antimicrobial defense	X		0.00	M,S,Sc,Y,E	0.46	0.03	0.00	1.00	0.00	1.00	M,S,Sc,Y,E,A,P ,Pr,Mj
						M,S,Sc,Y,E,A,P	0.32	0.10	0.00	23.07	0.00	1.00	
dsim_GLEANR_12120	CG11137	Unknown		X	0.10	M,S,Sc,Y,E	0.15	0.24	0.01	1.00	0.00	1.00	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.02***	0.84	0.01	1.00	.02*	1.00	
dsim_GLEANR_9282	CG11200	Dehydrogenase	X		0.02	M,S,Sc,Y,E	0.18	0.41	0.04	1.00	0.00	1.00	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.06***	0.36	0.00	7.37	0.00	1.00	
dsim_GLEANR_12517	CG12232	Cell communication			0.05	M,S,Sc,Y,E	0.09	0.47	0.02	1.00	0.00	2.32	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.03	0.79	0.00	10.79	0.00	10.24	
dsim_GLEANR_9897	CG12386	Serine Protease	X		0.14	M,S,Sc,Y,E	0.05***	2.35	0.06	2.09	0.05	2.31	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.03***	2.69	0.07	1.00	0.03*	2.64	
dsim_GLEANR_10438	CG12840	Cell communication		X	0.25	M,S,Sc,Y,E	0.005*	14.33	0.00	14.19	0.00	14.22	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.15***	0.68	0.03	1.00	0.08	1.00	
dsim_GLEANR_9414	CG14495	Unknown			0.09	M,S,Sc,Y,E	0.09***	1.56	0.09	1.56	0.09	1.56	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.01***	2.66	0.00	16.63	0.00	3.18	
dsim_GLEANR_9351	CG15098	Unknown		X	0.69	M,S,Sc,Y,E	0.02***	4.63	0.04	3.84	0.05	3.23	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.09***	2.72	0.05	3.46	0.08***	2.89	
dsim_GLEANR_9344	CG15106	Juv. Horm. Catabolism	X	X	0.06	M,S,Sc,Y,E	0.15***	0.48	0.03	1.00	0.00	1.00	M,S,Sc,Y,E,A,P ,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.15***	0.22	0.03	1.00	0.00	1.47	

dsim_GLEANR_5780	CG15293	Unknown	X		0.49	M,S,Sc,Y,E	0.20***	1.42	0.07	1.69	0.49	1.15	M,S,Sc,Y,E,A,P,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.07***	2.32	0.15	1.00	0.05	2.41	
dsim_GLEANR_5271	CG15533	Phospholipid metabolism	X	X	0.13	M,S,Sc,Y,E	0.19***	0.47	0.05	1.00	0.00	1.00	M,S,Sc,Y,E,A,P,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.01***	1.08	0.01	1.00	0.00	1.11	
dsim_GLEANR_6593	CG17012	Serine Protease	X		0.89	M,S,Sc,A	0.07***	6.08	0.05***	8.21	0.06***	6.96	M,S,Sc,A,V
dsim_GLEANR_2235	CG17119	Amino acid transport	X	X	0.07	M,S,Sc,Y,E	0.04***	1.54	0.04	1.54	0.04	1.55	M,S,Sc,Y,E,A,P,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.08***	0.98	0.00	2.44	0.06	1.06	
dsim_GLEANR_6882	CG17234	Serine Protease	X		0.74	M,S,Sc,Y,E	0.05***	7.23	0.09***	5.25	0.10***	4.86	M,S,Sc,Y,E,P,Pr
						M,S,Sc,Y,E,P	0.06***	5.94	0.10***	4.20	0.20***	2.82	
dsim_GLEANR_6595	CG17239	Serine Protease	X		0.73	M,S,Sc,Y,E	0.22***	2.40	0.22*	2.41	0.22*	2.40	M,S,Sc,Y,E,A,Pr,V
						M,S,Sc,Y,E,A	0.16***	2.70	0.09**	3.41	.13***	2.94	
dsim_GLEANR_6596	CG17240	Serine Protease	X		0.64								M,S,Y,E
dsim_GLEANR_12116	CG17521	Translation			0.02	M,S,Sc,Y,E	0.005**	2.77	0.01	2.77	.001*	2.77	M,S,Sc,Y,E,A,P,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	.01**	1.75	0.00	1.69	.005*	1.75	
dsim_GLEANR_5775	CG18125	Serine Protease	X		0.79								M,S,Sc
dsim_GLEANR_11034	CG30197	Unknown	X		0.05	M,S,Sc,Y,E	0.02**	4.37	0.02	4.37	.002*	4.37	M,S,Sc,Y,E,A,P,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.33	0.16	0.00	1.00	0.00	4.26	
dsim_GLEANR_15401	CG30371	Serine Protease	X		0.13	M,S,Sc,Y,E	0.36*	0.48	0.05	1.00	1.05	1.00	M,S,Sc,Y,E,A,P,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.02***	1.40	0.00	27.59	0.01	1.75	
dsim_GLEANR_4658	CG31343	Peptidase		X	0.14	M,S,Sc,Y,E	0.13***	1.11	0.13	1.11	0.12	1.12	M,S,Sc,Y,E,A,P,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.04***	1.37	0.04	1.00	0.01	2.02	
dsim_GLEANR_18113	CG31681	Serine Protease	X		0.82								M,S,Sc,Y,E

dsim_GLEANR_6881	CG31686	Unknown			0.66								M,S,Sc,Y,E
dsim_GLEANR_14267	CG32068	Secondary metabolism			0.39	M,S,Sc,Y,E	0.08***	1.18	0.08	1.18	0.08	1.18	M,S,Sc,Y,E,A,P,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.33***	0.20	0.00	5.32	0.00	1.23	
dsim_GLEANR_17130	CG32072	Protein-protein interact.			0.10	M,S,Sc,Y,E	0.02***	2.55	0.00	6.09	0.01**	3.58	M,S,Sc,Y,E,A,P,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.06***	1.41	0.05	1.00	0.03***	2.11	
dsim_GLEANR_15288	CG32834	Serine Protease	X		0.19	M,S,Sc,Y,E	0.21***	1.76	0.21	1.76	0.20	1.77	M,S,Sc,Y,E,A,P,Pr
						M,S,Sc,Y,E,A,P	0.09***	2.74	0.02	5.15	0.06*	3.25	
dsim_GLEANR_10398	CG33134	Apoptosis	X	X	0.06	M,S,Sc,Y,E	0.06	0.97	0.03	1.00	0.06	1.00	M,S,Sc,Y,E,A,P,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.06***	0.71	0.00	2.11	0.04	1.00	
dsim_GLEANR_5795	CG33306	Unknown	X		0.04	M,S,Sc,Y,E	0.27	0.47	0.03	1.00	0.00	1.00	M,S,Sc,Y,E,A,P,Pr,W,Mj,V,G
						M,S,Sc,Y,E,A,P	0.34***	0.4	0.05	1.00	0.02	1.75	
dsim_GLEANR_15604													
dsim_GLEANR_6594													

[1]Predicted protein function; [2]"X" indiates that the proteins has a predicted secretion signal;[3]"X" indicates the proteins has a predicted transmembrane region;[4]d_N/d_S: pairwise comparison of *D. melanogaster* and *D. simulans* sequences, estimated assuming no rate heterogeneity; [5]Species of *Drosophila* from which sequences were obtained for PAML analysis M= *melanogaster*, S = *simulans*, Sc= *sechellia*, Y = *yakuba*, E = *erecta*; A= *ananassae*; P = *pseudoobscura*. The following statistics are all derived from PAML analysis: [6]p_S: the proportion of sites estimated to belong to the class that has the highest d_N/d_S in M3; [7]d_N/d_S: for the highest class in M3;[8]p_S: the proportion of sites estimated to belong to the class that has $d_N/d_S > 1$ in M2; [9]d_N/d_S: the estimate for the class with the ratio > 1 in M2; [10]p_S: the proportion of sites estimated to belong to the class that has $d_N/d_S > 1$ in M8;[11]d_N/d_S: the estimate for the class with the ratio > 1 in M8; Statistical significance for LRT: *P <.05; **P<.01; ***P<.001.[12]The genomes in which orthologs were found using the 12 Drosophila genomes. Pr = *persimilis*, W = *willistoni*, Mj = *mojavensis*, V = *virilis*, and G =*grimshawi.*

Chapter III.

Resequencing study of spermathecal genes: a cluster of proteases and a gene of unknown function.

ABSTRACT

Genes expressed within the *Drosophila* spermathecae (a female sperm storage organ) are diverging at a rate exceeding that of *Drosophila* male accessory gland proteins, which are a paradigm for rapid evolution. Extensive studies of male reproductive genes and their products, such as accessory gland proteins in male *Drosophila*, have alluded to the importance of male-female interactions in the rapid evolution of reproductive proteins. Some of the most rapidly evolving genes within the spermathecae are serine-type proteases. Female proteases are candidate genes for such interactions; they have secretion signals and could be functioning in cascades, processing of male proteins or interactions with male serine protease inhibitors. Genes of unknown classification identified in the spermathecae are candidates because they may be playing a role specific to spermathecae function. In this study, we sequenced 5 genes, (4 proteases and 1 gene of unknown classification) located on the second chromosome in *Drosophila melanogaster* previously identified as expressed in the spermathecae. Polymorphism data reveals significant polymorphism-to-divergence heterogeneity in specific regions of the genes and a significant departure from equilibrium-neutral expectations in 3 genes. This data supports the hypothesis of positive selection occurring in these genes within *Drosophila* populations.

INTRODUCTION

Female sperm storage is an essential reproductive process utilized by most animals with internal fertilization (reviewed in (Neubaum & Wolfner, 1999c, Bloch-Qazi et al., 2003). Sperm storage is thought to protect sperm from displacement, extend the amount of time sperm are available for fertilization, and provide an area for competition between the ejaculates of males and for female sperm choice (Parker, 1970, Thornhill, 1983, Birkhead & Moller, 1998, Eberhard, 1996). Despite the importance of sperm storage, the identity and function of reproductive proteins associated with female sperm storage organs (SSOs) are poorly understood. These proteins, which are likely to mediate SSO processes such as storing, maintaining and releasing sperm, could also play key roles in female interactions with male reproductive proteins such as male accessory gland proteins (Acps), possibly mediating sperm competition or facilitating female cryptic choice of sperm or counteracting the detrimental effects of these proteins. An understanding of sperm storage is essential to defining important co-evolutionary processes such as sexual selection and sexual conflict.

Insect male accessory gland proteins (Acps) have been foci for molecular population genetic and molecular evolutionary studies. The average rate of sequence divergence of *D. melanogaster* Acps is approximately twice that of non-reproductive proteins (Begun et al., 2000a, Swanson et al., 2001a-a, Mueller et al., 2005a, Wagstaff & Begun, 2004b). At least seven Acps are transferred to the SSOs after mating (Ravi-Ram & Wolfner, 2005), including rapidly evolving Acps such as the protease inhibitor Acp62F (Lung et al., 2002b) and a protein that plays a key role in sperm storage (Acp36DE). Comparatively, the female reproductive genes are understudied, but the signature of positive selection has been revealed by a *D. simulans* evolutionary expressed sequence tag (EST) study (Swanson et al., 2004a), *D. melanogaster* molecular population genetic studies of female reproductive tract genes (Lawniczak & Begun, 2007), and an evolutionary EST study using *D. arizonae* reproductive tracts (Kelleher et al., 2007). A study focusing directly on Drosophila spermathecae, the long-term sperm storage organ in female, found that genes enhanced for spermathecae-specific expression were evolving under positive selection (Prokupek et al.). The proportion of positively selected genes in

the spermathecae (38%) (Prokupek et al.) detected by maximum likelihood analysis is greater than that seen in male seminal fluid genes (16%) (Haerty et al., 2007).

In *D. melanogaster,* second male sperm precedence (P2) can be partially attributed to a non-sperm component of the ejaculate (Harshman & Prout, 1994). Male Acps play a major role in sperm competition in *Drosophila* (Ravi-Ram & Wolfner, 2007). Allelic variation in male Acp genes has been associated with first male sperm precedence (P1) and P2 in this species (Clark et al., 1995, Fiumera et al., 2005, Fiumera et al., 2007). Although the specific genes remain unknown, genetic studies reveal that females play a major role in sperm competition in *D. melanogaster* (Clark & Begun, 1998, Clark et al., 1999). SSOs are a likely place to find specific female reproductive molecules that affect sperm competition.

Drosophila species typically have two types of organs dedicated to sperm storage (Fowler, 1973, Pitnick et al., 1999). The seminal receptacle contains the majority (65 - 80%) of the sperm (Lefevre & Jonsson, 1962, Neubaum & Wolfner, 1999a), while a pair of spermathecae are the site of long term storage. *Drosophila* spermatheca are cuticle-lined organs derived from ectodermal tissue partially surrounded by secretory cells. Sperm are stored in the spermathecal lumen, which receives proteins of unknown function from surrounding secretory epithelial cells (Filosi & Perotti, 1975). The interaction between sperm and SSOs are known to be evolutionarily important. For example, changes in sperm length result in the evolution of changes in the length of the seminal receptacle (Miller & Pitnick, 2002, Miller & Pitnick, 2003). Dynamic evolutionary interactions could exist between female reproductive proteins and male reproductive proteins, particularly in the spermathecae due to the extended period of time sperm spend in this organ.

A high proportion of genes expressed in the spermathecae show rapid rates of evolution, and an especially high level of positive selection, characteristic of reproductive proteins (i.e., sperm-egg interaction proteins, Acps) (Prokupek et al.). Serine-type proteases were overrepresented in the spermathecae, and were among the most rapidly evolving genes identified. Serine proteases could interact with male seminal products in a number of ways, such as the digestion of male proteins, proteolytic cleavage of male proteins for activation or in cascades (immune response, semen coagulation). Genes of

unknown classification were also among the most rapidly evolving genes expressed in the spermathecae. Genes with unknown classification might be quite interesting in terms of having spermatheca - specific roles because functions for such genes have not been identified in other tissues or taxa.

A proper investigation into the genetic evolution within a species requires an accurate description of variation both within and between populations, as well as the ability to distinguish between the potential causes of the observed variation distribution (Begun & Aquadro, 1993). There are two types of measurement for genetic variation at the DNA level: the average number of pairwise nucleotide differences and the number of segregating (polymorphic) sites among a sample of DNA sequences (Takahata & Clark, 1993). Comparative approaches (i.e., between species analysis by a maximum likelihood such as that conducted by the PAML program) use a comparison of the ratio of nonsynonymous changes per nonsynonymous site to synonymous changes per synonymous site in order to test for past selection. These tests, while informative, can not sufficiently differentiate between the forces of selection, nor do they allow inference into present day selection. Population genetic approaches aim at detecting ongoing selection in a population by analyzing current levels of polymorphisms. In addition, polymorphic approaches are much more sensitive than interspecific approaches for the identification and classification of selective pressures. Measures of the level of polymorphism can be revealing in terms of discovering forces acting on particular genomic loci. Low levels of polymorphism indicate either directional selection or genetic drift while high levels of polymorphism are reflective of high mutation rates, large historic population size, neutral evolution or balancing selection. The combination of multiple test statistics can be used to reveal more about the patterns of selection (Otto, 2000).

Relatively little research has been done to detect positive selection in noncoding regions based on comparative data. Methods similar to those used to detect elevated K_a/K_s have been devised as a way of estimating selection in these regions (Wong & Nielsen, 2004, Haygood et al., 2007). The presence of highly variable sites in noncoding regions may be interesting in terms of the evolutionary changes which occur in non-coding regulatory regions (typically upstream regions) which result in phenotypic differences due to changes in gene regulation. A large portion of the non-translated

genome of *Drosophila* has been suggested as functionally important, and therefore subject to both purifying and adaptive evolution (Andolfatto, 2005).

Studies on genetic variation provide a powerful means for elucidating the genetic, evolutionary and demographic factors shaping the *Drosophila* genome. In the present study we examined genetic variation for 5 genes (and regions associated with these genes) on the chromosome 2L in 35 lines from 2 populations of *Drosophila melanogaster* (19 North American, and 16 African). Four of the genes investigated were serine-proteases, and one was a gene of unknown classification. These genes all have enhanced expression in the spermathecae as well as signal peptides (proteases) or transmembrane domains (gene of unknown classification) and thus may be involved in interactions with male seminal products.

This study addressed two major questions: (1) Is the pattern of variation in these genes incompatible with neutral evolution? (2) What is the pattern of variation in non-coding regions immediately upstream and downstream of coding regions (likely containing promoter and regulatory regions) compared to a non-coding region not associated with a gene?

MATERIALS and METHODS

Experimental line origins and DNA sequencing

D. melanogaster flies used in this study came from second chromosome extracted lines from two populations (Uganda and Pennsylvania). Nineteen lines were used from the Pennsylvania (PA) population, and 16 lines from the Uganda (UG) population. DNA was extracted from single males from each line.

DNA sequences of the complete coding and non-coding (upstream, downstream, and intronic) regions for each of the proteases genes, and the gene of unknown function were produced. Eleven primer pairs were designed along the entire regional cluster of four protease genes (CG17012, CG17234, CG17239, CG17240 (Ser12)). The resulting sequence included the coding regions, all intergenic sequences between the protease genes and stretches of DNA (~500bp) immediately upstream and downstream of the cluster (genomic sequence location 2L: 2250003 – 2255500). The genes of the cluster contain no introns. Five primer pairs were designed for a gene of unknown function

(CG15098) located on the second chromosome which included upstream and downstream regions (~500bp) (genomic sequence location 2R: 14720524 - 147722760). One primer pair was designed for a randomly selected region (~600bp) of non-coding DNA from chromosome 2 (genomic sequence location 2L: 2257882 – 2258882). The closest gene to this region is 2832bp upstream. DNA samples were sent to High Throughput Sequencing Solutions (Seattle, WA) for primer design, PCR amplification and 2-way sequencing.

Clean-up and Alignment

The Contig Express program (Invitrogen) was used to clean-up the DNA sequences, as well as to create contiguous sequences out of the individual sequence reads. Both forward and reverse strand sequences were used to create the consensus sequence for each sample set. Regional sequences from Flybase (Release 5.1; http://flybase.org) were used as a reference for proper alignment of the contigs. Regions where sequences were ambiguous between the two strands, or missing in sequence data, were replaced with dummy (N) sequences. Alignments of each region were generated with the AlignX program of Vector NTI Advance. Sequences from Flybase were used as a reference for the coding frame; all sequences were adjusted by hand when necessary to keep the proper coding frame.

DNA sequence analysis

Polymorphism

Polymorphism analysis tests were done using the DnaSP version 4.20 software package (Rozas et al., 2003). A sliding window analysis of the ratio of polymorphism to divergence using a window of 50bp with a step size of 10 was also conducted using DnaSP. To test for departures from neutrality we calculated Tajima's D (Tajima, 1989), Fu and Li's D (Fu & Li, 1993), and Fay and Wu's H (Fay & Wu, 2000). Significance for Fay and Wu's H statistic was determined by coalescent simulations using a recombination parameter, R, estimated by DnaSP for each region. Tests based on the frequency spectrum (Fay and Wu's H, Tajima's D, and Fu and Li's D) detect levels of variation that are inconsistent with the expectation of neutral equilibrium model, but are influenced by demographic effects. Fu and Li's D compares the number of mutations

which occur in external versus internal branches of the genealogy (Fu & Li, 1993). Tajima's D test compares a standardized measure of the total number of polymorphic sites in the sample with the average number of mutations between pairs in the sample (Tajima, 1989). If D (either Tajima's or Fu and Li's) is significantly different from 0, we reject the neutral mutation hypothesis. A D significantly smaller than 0 is suggestive of directional selection, greater than 0 is suggests balancing selection. One problem in using the statistics from these tests alone is that selective sweeps, bottlenecks and population expansion/reduction can result in similar effects at particular loci. Fay and Wu's H compares the proportion of alleles at intermediate versus high frequency. Positive selection is indicated by a negative value of H (Fay & Wu, 2000). This test is useful in the detection of past hitchhiking events because H is sensitive to an excess of new, high frequency alleles.

Divergence

The McDonald – Kreitman (MK) (McDonald & Kreitman, 1991) test was conducted to compare levels of synonymous and nonsynonymous variation within and between species. The neutral prediction is that the ratio of nonsynonymous to synonymous substitutions is equal to the ratio of nonsynonymous to synonymous polymorphisms.

The Hudson-Kreitman-Aguade (HKA) (Hudson et al., 1987) test, used in the present study, predicts a correlation between the levels of polymorphism within a species, and divergence between two closely related species. HKA is a goodness of fit test, comparing estimates of parameters describing within-species diversity and between-species divergence using data from two or more loci (Hudson et al., 1987). If functional constraints limit evolution, then levels of polymorphism and divergence are equally constrained. This test is typically applied to sites of silent variation (untranslated or synonymous sites), testing if the pattern of silent variation within and between species is different for two regions. The HKA test uses a neutral locus for comparison; a random, non-coding region of the 2nd chromosome was used in the present study as a neutral loci. This locus was relatively far from any up or downstream coding regions, and thus is unlikely to contain promoter or regulatory regions. Tests of neutrality were also

performed on this region to ensure that it conforms to the hypothesis of neutral evolution. Homogeneity between the loci of interest and the assigned neutral loci would support a hypothesis of neutrality. Failure to reject this model implies that the differences among locus polymorphisms arise from differences in neutral mutation rates.

In tests requiring between species comparisons (HKA and MK), the corresponding chromosome region from *D. simulans* (a sister species of *D. melanogaster*) was used. The orthologous region of *D. simulans* was determined from NCBI BLAST (Altschul et al., 1990), and aligned using Clustal W. All sequences were manually adjusted when necessary in order to keep the coding frame. The use of a sister species prevents the outgroup species from being too distantly related. If the species are too distant, the tests will be biased in the direction of excess changes within a species by violating the assumption of a single mutation per site. If the species are too close, the test could be biased with a finding of excess changes between species, because shared polymorphisms will be confounded with fixed differences (reviewed in(Wayne & Simonsen, 1998).

RESULTS

The evolutionary patterns of genes expressed within the spermathecae of *Drosophila melanogaster* were examined. The intention was to expand knowledge of the forces driving the evolution of female reproductive traits, in this case, sperm storage. Six loci on chromosome 2L, five gene regions including upstream, downstream, and intronic regions, as well as a single non-coding region in 35 lines from 2 populations of *Drosophila melanogaster* (19 North American, and 16 African) and one outgroup (*D. simulans*) were used for analyses. Analyses were performed when it was determined that data would not be skewed by limited sequence information.

Numerous indels were detected in the intron and non-coding region alignments and were ignored in the analysis. A large indel (~1000bp) was present in the *D. simulans* genome between the ortholog to CG17239 and CG17234; this region was missing in the *D.melanogaster* genome. This region in the *D. simulans* genome was ignored for analysis. This region encodes the *D. simulans* gene dsim_GLEANR_6594. Using a conserved domain search, this gene is predicted to be a serine protease (E-value 6^{-46}).

Polymorphism analyses

Polymorphism studies allow for more comprehensive inferences about the nature of selective forces driving the evolution of genes compared to divergence studies alone. The data from polymorphism studies are used for testing deviations for neutrality. A summary of polymorphisms is presented in Table 3.1. The genome average of heterozygosity calculated from a random sample of autosomal genes is .001 (Andolfatto, 2001). The average replacement heterozygosity calculated for the genes in this study is .007. This number is much greater than the genome average and similar to that seen in the Acps of *D. melanogaster* (Begun et al., 2000b).

Polymorphism tests show that three of the genes significantly deviate from the neutral expectations (CG17012, CG17234, and CG17240) (Table 3.2). In all three cases Tajima's D values were negative. Negative values can be due to the presence of an excess of low-frequency variants usually caused by a sudden population expansion or by the selection of one specific allele over alternative alleles. The same three genes had significantly negative Fu and Li's D which is consistent with positive selection.

Divergence

K_a (nonsynonymous substitutions per nonsynonymous site) and K_s (synonymous substitutions per synonymous site) estimates for the comparisons between *D. melanogaster* and *D. simulans* are presented in Table 3.3. The genome average synonymous site divergence (K_s) between these two species is estimated to be ~0.12 (Begun & Whitley, 2000) and the average amino acid divergence (K_a) average is estimated to be ~.011 (Begun, 2002). The average K_a for the protease genes of this study is 0.07, over six times the calculated genome wide average, whereas the K_s average for the protease genes of this study is similar to the genome wide average (0.11). We also compared our values to the average K_a, and K_s values for 166 randomly chosen serine proteases (K_s =0.1279, K_a = 0.0258, K_a/K_s = 0.203) (Lawniczak & Begun, 2007). On average, the K_a (0.07) and K_a/K_s (0.66) of the spermathecae serine proteases was 2 to 3-fold higher than that of a random selection of serine proteases.

When comparing the gene of unknown function to the genomic average, we see that although the K_a value is markedly higher than average, the K_s value is lower (Table 3.4). This indicates that the elevated K_a/K_s ratios are affected by changes in both the nonsynonymous and synonymous substitution rates. Decreases in synonymous substitution rates may be due to selection for optimal codon usage (Moriyama & Powell, 1997), or an overall decrease in the mutation rate at this locus. The K_a/K_s ratio of the gene of unknown function is as high, or higher than the protease ratios due to the low K_s value.

Hudson, Kreitman, Aguade Test

The HKA test was used to asses the neutral prediction that the ratio of polymorphism to divergence is the same for different loci for a series of pairwise comparisons with *D. simulans* as the outgroup. This test is used for detecting a deficit or excess of polymorphism either due to directional selection or balancing selection, respectively. A single non-coding region of ~600 base pairs (over 2kb away from the closest coding region) was used as a 'neutral' locus for comparison. We also used the upstream region of each gene as an additional locus for each comparison. Additionally for CG15098 we concatenated the introns and used them as an additional neutral locus. All five genes conformed to the neutral prediction in all comparisons.

McDonald-Kreitman Test.

The MK test is analogous to the HKA test, but by tracking synonymous versus nonsynonymous substitution it is more sensitive to the detection of positive selection. The MK test was not significant for any of the genes tested. The ratio of nonsynonymous to synonymous fixed substitutions was not statistically different from the ratio of nonsynonymous to synonymous polymorphic sites.

For all of the protease genes, there are a number of replacement polymorphisms, but no amino acid fixations between the PA and UG populations. The excess of replacement polymorphisms is consistent with both neutral processes (relaxation of functional constraints) and selective processes (segregation of deleterious mutation or balancing selection).

Non-coding DNA

Evolutionary changes which occur in non-coding regulatory regions (typically upstream regions) may result in phenotypic differences due to changes in gene regulation. A large portion of the non-translated genome of *Drosophila,* including regulatory regions, is potentially functionally important, and therefore subject to both purifying and adaptive evolution (Andolfatto, 2005). Promoter regions were predicted using the Neural Network Promoter Prediction from the Berkley Drosophila Genome Project (http://www.fruitfly.org/seq_tools/promoter.html). Significant negative values for Tajima's D and Fu and Li's D were seen in upstream regions of CG17012, CG17234, CG17239 and CG17240, consistent with positive selection.

The average divergence between *D.melanogaster* and *D.simulans* in untranslated regions (UTRs) is estimated to be 0.045 (Andolfatto, 2005). We chose to use the randomly selected genomic region of the second chromosome for comparison with the upstream and intronic regions associated with the genes in our study (divergence (K_g) = ~0.064). To get an idea of the evolutionary rates of the upstream and intronic regions of our study genes we calculated the ratio of substitution rates in the upstream/intronic regions compared to those in the random genomic regions (Tables 3.2 and 3.3). This ratio is analogous to the ratio of substitution rates at nonsynonymous sites to synonymous sites in coding regions, and allows a way to estimate the forces of positive selection acting on non-coding regions (Wong & Nielsen, 2004, Haygood et al., 2007). The average value for the ratio of upstream regions to genomic region substitutions was 2.5 ± 0.75. The average value for the intronic regions of CG15098 was 1.56, and for the downstream region of CG15098 was 1.25. The upstream regions are on average twice as diverse as the random genomic, intronic and downstream regions.

Sliding Window Analysis

Oftentimes different regions of a single gene can be exposed to different selective pressures. In these cases, calculating K_a/K_s over the entire length of the gene does not provide an accurate picture of how selection is acting on the gene. Both the MK and the HKA tests use the average polymorphism and divergence rate for each gene. Failure to

reject the neutral predictions for these tests may be explained by the over-all gene averages masking selection on specific regions. Sliding window analysis of both polymorphism and divergence can help determine if specific regions and codons in a gene are under selection, or if they are highly conserved. The sliding widow analysis (Figures 3.1-5) shows inflated K_a/K_s (divergence) values relative to Pi_a/Pi_s (polymorphism) values for all five of the genes analyzed. This result is strongly indicative of these regions are evolving under positive selection. Moreover, the sliding window analysis was helpful in finding non-coding regions which showed higher than average levels of polymorphism (Figures 3.6 and 3.7).

DISCUSSION

Female reproductive proteins are evolving rapidly in *Drosophila* (Civetta & Singh, 1995, Panhuis & Swanson, 2006). Sexual selection has been implicated to explain the evolution of reproductive proteins (e.g. gamete recognition proteins) (Swanson & Vacquier, 2002). The exact nature of selection driving the evolution of spermatheca genes is not known, but it is likely to be a form of sexual selection possibly driven by interactions with male seminal products.

Sperm storage is a dynamic process in which both males and females play active roles (Adams & Wolfner, 2007). In *D. melanogaster*, a feminized nervous system is required for accumulation of sperm in storage (Arthur et al., 1998). Female secretions, as seen in the honeybee *Apris mellifera*, are necessary to support sperm storage (Weirich et al., 2002, Collins et al., 2004a). Male seminal proteins are also required for proper sperm storage and accumulation (Kalb et al., 1993, Tram & Wolfner, 1999). Roles of both the male and female have been identified in sperm competition (Clark et al., 1995, Clark et al., 1999, Clark & Begun, 1998, Hosken, 1999, Hosken & Ward, 2000, Lawniczak & Begun, 2005). The interactions between male and female reproductive proteins have been recognized as important in evolutionary processes such as male-female co-evolution, sexual selection and speciation (Pitnick et al., in prep.).

Very little is known about reproductive proteins present in SSOs. A previous evolutionary expressed sequence tag (EST) study on genes expressed within the spermathecae identified this organ as a region of high rates of evolution (Prokupek et al.).

PAML analysis conducted in the EST study predicted that a high proportion of genes in the spermathecae are rapidly evolving. The five genes in the present study were previously analyzed using overall K_a/K_s ratios as well as in a PAML analysis using multiple species. It was determined that all five of these genes showed signs of positive selection on a subset of their codons, but without polymorphism data the previous study lacked information about evolution within species and it lacked the comprehensive battery of tests used in the present study.

Examination of both divergence and polymorphism provided insight into the evolutionary history of a set of proteases expressed in the spermathecae. The failure to reject Fay and Wu's H indicates that the reduced level of variation observed is not due to hitchhiking (Otto, 2000). A non-significant H, combined with the results of other neutrality tests (Tajima's D, Fu and Li's D), is suggestive that three of the proteases are undergoing positive selection. Positive Fu and Li's D test values were found in the upstream region of CG17239 as well as the downstream region of CG16098, indicating that balancing selection may be operating in these regions. Positive strong peaks (greater than 1) of K_a/K_s, along with low levels of relative polymorphism at the same point, were observed in the sliding window analysis. These peaks suggest that positive selection is operating on a subset of the codons in all of the protease genes.

Serine –type endopeptidases are recognized as important reproductive proteins both in the female reproductive tract, and in seminal fluids. Female reproductive tract secretion of digestive enzymes alludes to a form of ejaculate-female interaction in which females actively degrade or activate protein components of the male ejaculate. Serine proteases, found in the spermathecae may be interacting with Acps which enter the sperm storage organs. Four of the Acps found in the sperm storage organs are serine protease inhibitors. Serine proteases inhibitors (serpins) bind irreversibly to serine proteases, functionally inactivating them (Potempa et al., 1994). While no direct interaction between the serine proteases of the female reproductive tract and serpin Acps has been found thus far, it does present an interesting scenario for these two classes of proteins. Serine proteases could also be acting in a non-antagonistic manner with male proteins in coordinating immune response or semen coagulation cascades.

The four protease genes investigated in the present study are found in a cluster on chromosome arm 2L. They have no introns and they encode proteins with the canonical serine protease catalytic triad of amino acids. The cluster is transcriptionally activated by mating (Lawniczak & Begun, 2007). In *Drosophila arizonae*, reproductive tract-specific digestive proteases show evidence of directional selection, with certain amino acids experiencing strong positive selection (Kelleher et al., 2007).

Tests of CG15098 (gene of unknown classification) failed to reject the hypothesis of neutrality. However, this gene showed higher than average levels of heterogeneity in polymorphism along the length of the gene. Sliding window analysis revealed two peaks which could indicate that positive selection is operating in these specific regions of the gene. Genes of unknown function could play roles that are spermathecae-specific. The gene analyzed in this study has a predicted transmembrane region, and could be interacting with male proteins in the spermathecae. The two regions identified by sliding window analysis could possibly be the regions of male-female interaction.

Polymorphism tests on the non-coding regions of the genes revealed that regions upstream of the genes are two to three times more polymorphic than the designated neutral regions. Analysis of specific predicted promoter regions showed that these regions, present in the upstream regions, had relatively low levels of polymorphism. Upstream regions, though not translated, could be undergoing positive selection which affects expression levels of the genes.

Patterns of nucleotide variation can reveal rapid evolution by positive selection as seen in *Drosophila* male reproductive proteins (Acps) (Swanson et al., 2001a). Acps, which are transferred to the female at the time of mating, play important roles in sperm storage, sperm competition/defense, female mate receptivity, egg laying rate, and female lifespan (reviewed in (Wolfner, 1997). Some Acps have been found to localize in specific regions of the female reproductive tract including the genital opening, the distal portion of the uterus and in sperm storage structures (Bertram et al., 1996, Heifetz et al., 2000, Ravi-Ram & Wolfner, 2005). Due to the important role Acps play in reproduction it is not surprising that these proteins have been found to evolve at twice the rate of non-reproductive proteins. What is perhaps more intriguing, is that female counter-parts to

these proteins remains undiscovered, although secreted proteins encoded by spermathecae genes are promising candidates to play this role.

This study provided an in depth look into the evolutionary patterns shaping five genes expressed within the spermatheca, the long term sperm storage organ of *Drosophila*. The results of this study indicate that the evolutionary patterns of female reproductive proteins present in the spermathecae is consistent with positive selection operating in a subset of the codons. The evolutionary pressures shaping these genes could be generated by male-female interactions. Further investigation is needed to identify specific proteins interacting with these candidate genes.

ACKNOWLEDGEMENTS

This research was supported by a National Science Foundation grant (DEB-ESP0346476).

REFERENCES

Adams, E. M., and M. F. Wolfner. 2007. Seminal proteins but not sperm induce morphological changes in the *Drosophila melanogaster* female reproductive tract during sperm storage. J. Insect Physiol. 53:319-331.

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.

Andolfatto, P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. Nature 437:1149-1152.

Arthur, B. I., E. Hauschteck-Jungen, R. Nothiger, and P. I. Ward. 1998. A female nervous system is necessary for normal sperm storage in *Drosophila melanogaster*: a masculinized nervous system is as good as none. Proc. R. Soc. Lond B 265:1749-1753.

Begun, D. J. 2002. Protein variation in *Drosophila simulans,* and comparison of genes from centromeric versus noncentromeric regions of chromosome 3. Mol Biol Evol 19:201-203.

Begun, D. J., and C. F. Aquadro. 1993. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature 365:548-550.

Begun, D. J., and P. Whitley. 2000. Reduced X-linked nucleotide polymorphism in *Drosophila simulans.* Proc Natl Acad Sci U S A 97:5960-5965.

Begun, D. J., P. Whitley, B. L. Todd, W.-D. H.M., and A. G. Clark. 2000a. Molecular population genetics of male accessory gland proteins in *Drosophila*. Genetics 156:1879 - 1888.

Begun, D. J., P. Whitley, B. L. Todd, H. M. Waldrip-Dail, and A. G. Clark. 2000b. Molecular population genetics of male accessory gland proteins in *Drosophila*. Genetics 156:1879-1888.

Bertram, M. J., D. M. Neubaum, and M. F. Wolfner. 1996. Localization of the *Drosophila* male accessory gland protein Acp36DE in the mated female suggests a role in sperm storage. Insect Biochem Mol Biol 26:971-980.

Birkhead, T. R., and A. P. Moller. 1998. Sperm competition and sexual selection. Academic Press, London.

Bloch-Qazi, M. C., Y. Heifetz, and M. F. Wolfner. 2003. The developments between gametogenesis and fertilization: ovulation and female sperm storage in *Drosophila melanogaster*. Developmental Biology 256:195-211.

Civetta, A., and R. S. Singh. 1995. High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *Drosophila virilis* group species. J Mol Evol 41:1085-1095.

Clark, A. G., M. Aguade, T. Prout, L. G. Harshman, and C. H. Langley. 1995. Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. Genetics 139:189-201.

Clark, A. G., and D. J. Begun. 1998. Female genotypes affect sperm displacement in *Drosophila*. Genetics 149:1487-1493.

Clark, A. G., D. J. Begun, and T. Prout. 1999. Female x male interactions in *Drosophila* sperm competition Science 283:217-220.

Collins, A. M., V. Williams, and J. D. Evans. 2004. Sperm storage and antioxidative enzyme expression in the honey bee, Apis mellifera. Insect Mol Biol 13:141-146.

Eberhard, W. G. 1996. Female control: sexual selection by cryptic female choice. Princeton University Press, New Jersey.

Fay, J. C., and C. I. Wu. 2000. Hitchhiking under positive Darwinian selection. Genetics 155:1405-1413.

Filosi, M., and M. Perotti. 1975. Fine structure of spermatheca of *Drosophila melanogaster* Meig. J. Submicr. Cytol 7:259-270.

Fiumera, A. C., B. L. Dumont, and A. G. Clark. 2005. Sperm competitive ability in *Drosophila melanogaster* associated with variation in male reproductive proteins. Genetics 169:243-257.

Fiumera, A. C., B. L. Dumont, and A. G. Clark. 2007. Associations between sperm competition and natural variation in male reproductive genes on the third chromosome of *Drosophila melanogaster*. Genetics 176:1245-1260.

Fowler, G. 1973. Some aspects of the reproductive biology of *Drosophila*: sperm transfer, sperm storage and sperm utilization. Adv in Genet 17:293-360.

Fu, Y. X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. Genetics 133:693-709.

Harshman, L. G., and T. Prout. 1994. Sperm displacement without sperm transfer in *Drosophila melanogaster*. Evolution 48:758-766.

Haygood, R., O. Fedrigo, B. Hanson, K. D. Yokoyama, and G. A. Wray. 2007. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. Nat Genet 39:1140-1144.

Heifetz, Y., O. Lung, E. J. Frongillo, and M. F. Wolfner. 2000. The *Drosophila* seminal fluid protein Acp26Aa stimulates release of oocytes by the ovary. Current Biology 10:99-103.

Hosken, D. J. 1999. Sperm displacement in yellow dung flies: a role for females. Trends Ecol Evol 14:251-252.

Hosken, D. J., and P. I. Ward. 2000. Copula in yellow dung flies (*Scathophaga stercoraria*): investigating sperm competition models by histological observation. J. Insect Physiol. 46:1355-1363.

Hudson, R. R., M. Kreitman, and M. Aguade. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116:153-159.

Kalb, J. M., A. J. DiBenedetto, and M. F. Wolfner. 1993. Probing the function of *Drosophila melanogaster* accessory glands by directed cell ablation. Proc Natl Acad Sci U S A 90:8093-8097.

Kelleher, E. S., W. J. Swanson, and T. A. Markow. 2007. Gene duplication and adaptive evolution of digestive proteases in *Drosophila arizonae* female reproductive tracts. PLoS Genet 3:e148.

Lawniczak, M. K., and D. J. Begun. 2005. A QTL analysis of female variation contributing to refractoriness and sperm competition in *Drosophila melanogaster*. Genet Res 86:107-114.

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

Lefevre, G., and U. B. Jonsson. 1962. Sperm transfer, storage, displacement, and utilization in *Drosophila melanogaster*. Genetics 47:1719-1736.

Lung, O., U. Tram, M. Finnerty, M. Eipper-Mains, J. Kalb, and M. Wolfner. 2002. The *Drosophila melanogaster* seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression. Genetics 160:211-214.

McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351:652-654.

Miller, G., and S. Pitnick. 2002. Sperm-female co-evolution in *Drosophila*. Science 298:1230-1233.

Miller, G., and S. Pitnick. 2003. Functional significance of seminal receptacle length in *Drosophila melanogaster*. J. Evol Biol 16:114-116.

Moriyama, E. N., and J. R. Powell. 1997. Synonymous substitution rates in *Drosophila*: mitochondrial versus nuclear genes. J Mol Evol 45:378-391.

Mueller, J., K. Ravi-Ram, M. McGraw, M. Bloch-Qazi, E. Siggia, A. Clark, C. Aquadro, and M. Wolfner. 2005. Cross-species comparison of *Drosophila* male accessory gland protein genes. Genetics 171:131-143.

Neubaum, D. M., and M. F. Wolfner. 1999a. Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. Genetics 153:857-869.

Neubaum, D. M., and M. F. Wolfner. 1999b. Wise, winsome, or weird? Mechanisms of sperm storage in female animals. Curr Top Dev Biol 41:67-97.

Otto, S. P. 2000. Detecting the form of selection from DNA sequence data. Trends Genet 16:526-529.

Panhuis, T. M., and W. J. Swanson. 2006. Molecular evolution and population genetic analysis of candidate female reproductive genes in *Drosophila*. Genetics 173:2039-2047.

Parker, G. A. 1970. Sperm competition and its evolutionary consequences in the insects. Biol. Rev. 45:525-567.

Pitnick, S., T. A. Markow, and G. S. Spicer. 1999. Evolution of multiple kinds of female sperm-storage organs in *Drosophila*. Evolution 53:1804-1822.

Pitnick, S., M. Wolfner, and S. Suarez. in prep. Sperm-female interactions *in* T. Birkhead, D. Hosken, and S. Pitnick, eds. Sperm Biology: An Evolutionary Perspective. Elsevier Press.

Potempa, J., E. Korzus, and J. Travis. 1994. The serpin superfamily of proteinase inhibitors: structure, function, and regulation. J Biol Chem 269:15957-15960.

Prokupek, A. P., F. Hoffman, S. Eyun, E. N. Moriyama, and L. G. Harshman. An evolutionary expressed sequence tag analysis of *Drosophila* spermatheca genes. In Review.

Ravi-Ram, K., and M. Wolfner. 2005. Fates and targets of male accessory gland proteins in mated female *Drosophila melanogaster*. Insect Biochem Mol Biol 35:1059-1071.

Ravi-Ram, K., and M. F. Wolfner. 2007. Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduction. Int. Comp. Biol.:19.

Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analysis by the coalescent and other methods. Bioinformatics 19:2496-2497.

Swanson, W., A. Wong, M. Wolfner, and C. Aquadro. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies several genes subjected to positive selection. Genetics 168:1457-1465.

Swanson, W. J., A. G. Clark, H. Waldrip-Dail, M. F. Wolfner, and C. F. Aquadro. 2001a. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc Natl Acad Sci USA 98:7375-7379.

Swanson, W. J., A. G. Clark, H. M. Waldrip-Dail, M. F. Wolfner, and C. F. Aquadro. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc Natl Acad Sci U S A 98:7375-7379.

Swanson, W. J., and V. D. Vacquier. 2002. The rapid evolution of reproductive proteins. Nat Rev Genet 3:137-144.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

Takahata, N., and A. G. Clark. 1993. Mechanisms of molecular evolution : introduction to molecular paleopopulation biology. Japan Scientific Societies Press ; Tokyo

Thornhill, R. 1983. Cryptic female choice and its implications in the scorpionfly *Harpobittacus nigriceps*. Am Nat 122:1548-1556.

Tram, U., and M. F. Wolfner. 1999. Male seminal fluid proteins are essential for sperm storage in *Drosophila melanogaster.* Genetics 153:837-844.

Wagstaff, B., and D. J. Begun. 2004. Comparative genomics of acessory gland protein genes in *Drosophila melanogaster* and *D. pseudoobscura*. Mol Biol Evol 22:818-832.

Wayne, M. L., and K. L. Simonsen. 1998. Statistical tests of neutrality in the age of weak selection. TREE 13:236-240.

Weirich, G. G., A. M. Collins, and V. P. Williams. 2002. Antioxidant enzymes in the honey bee, *Apis mellifera*. Apidologie 33:3-14.

Wolfner, M. F. 1997. Tokens of love: functions and regulation of *Drosophila* male accessory gland products. Insect Biochem Mol Biol 27:179-192.

Wong, W. S., and R. Nielsen. 2004. Detecting selection in noncoding regions of nucleotide sequences. Genetics 167:949-958.

Figure 3.1 Sliding window analysis of divergence K_a/K_s ratios and polymorphism π_a/π_s in the gene CG15098. The solid line indicates a K_a/K_s ratio of one, ratios larger than one are indicative of positive selection.

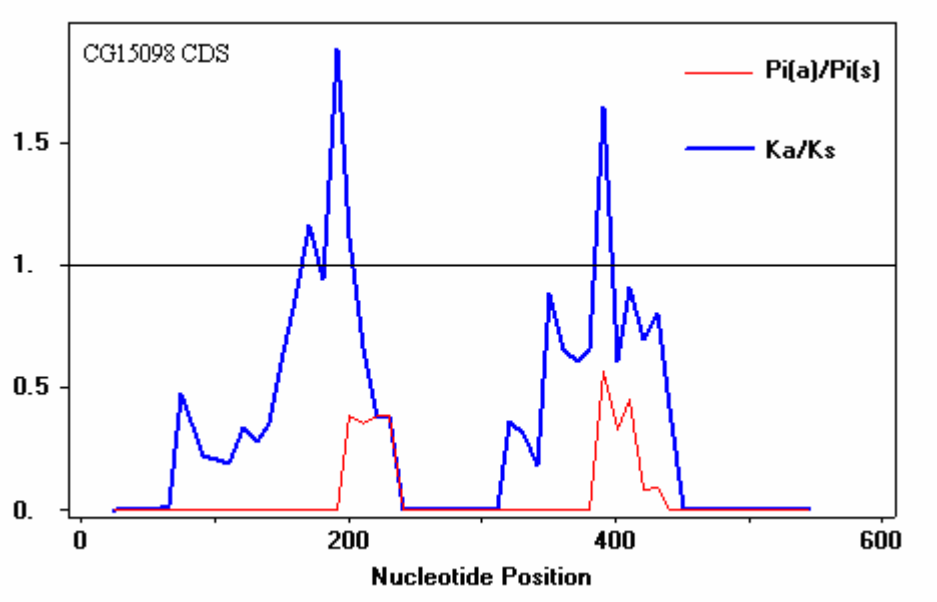

CG15098 CDS
Pi(a)/Pi(s)
Ka/Ks
1.5
1.
0.5
0.
0
200
400
600
Nucleotide Position

Figure 3.2 Sliding window analysis of divergence K_a/K_s ratios and polymorphism π_a/π_s in the gene CG17012. The solid line indicates a K_a/K_s ratio of one, ratios larger than one are indicative of positive selection.

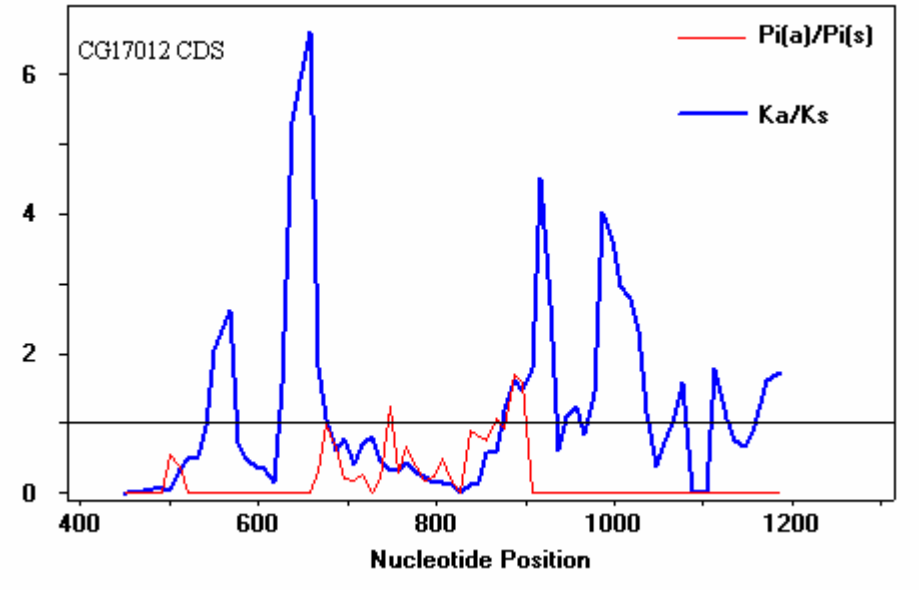

CG17012 CDS
Pi(a)/Pi(s)
Ka/Ks
6
4
2
0
400
600
800
1000
1200
Nucleotide Position

Figure 3.3 Sliding window analysis of divergence K_a/K_s ratios and polymorphism π_a/π_s in the gene CG17234. The solid line indicates a K_a/K_s ratio of one, ratios larger than one are indicative of positive selection.

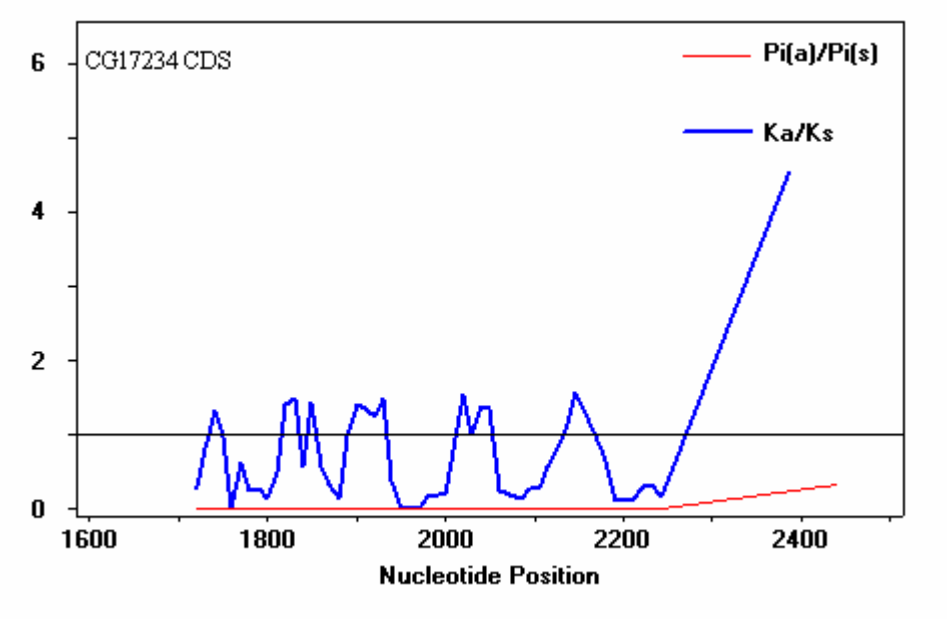

CG17234 CDS
Pi(a)/Pi(s)
Ka/Ks
6
4
2
0
1600
1800
2000
2200
2400
Nucleotide Position

Figure 3.4 Sliding window analysis of divergence K_a/K_s ratios and polymorphism π_a/π_s in the gene CG17239. The solid line indicates a K_a/K_s ratio of one, ratios larger than one are indicative of positive selection.

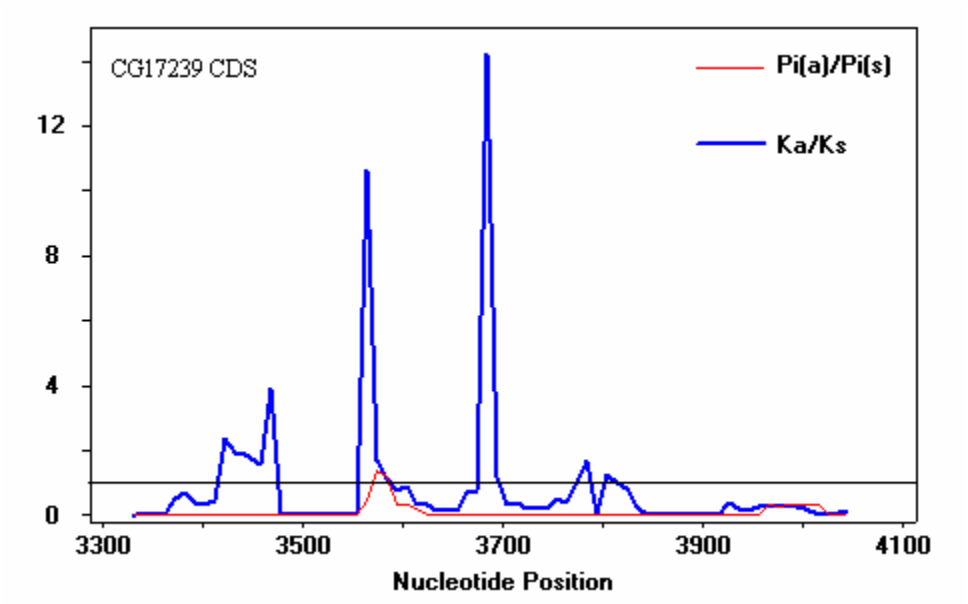

CG17239 CDS
Pi(a)/Pi(s)
Ka/Ks
12
8
4
0
3300
3500
3700
3900
4100
Nucleotide Position

Figure 3.5 Sliding window analysis of divergence K_a/K_s ratios and polymorphism π_a/π_s in the gene CG17240 (Ser12). The solid line indicates a K_a/K_s ratio of one, ratios larger than one are indicative of positive selection.

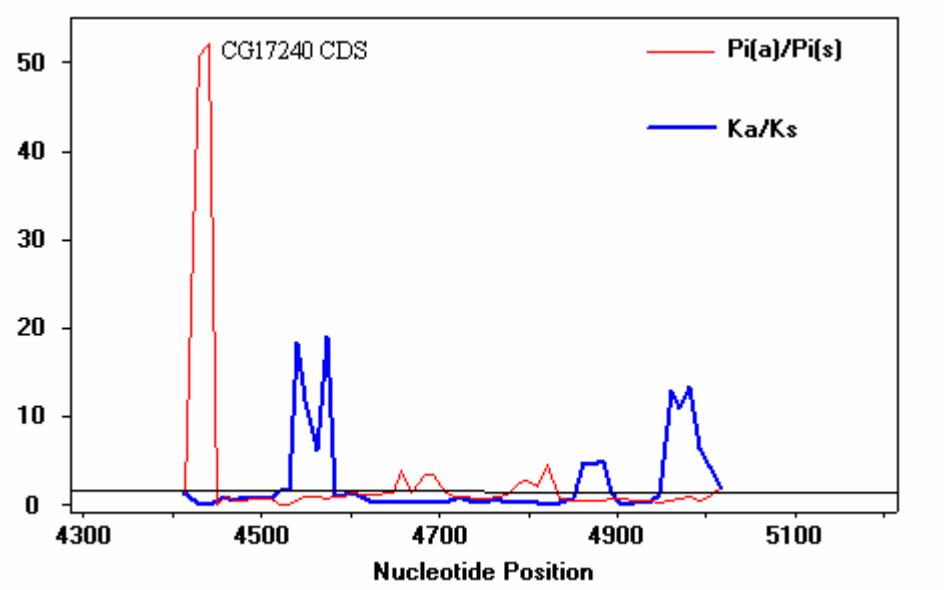

CG17240 CDS
Pi(a)/Pi(s)
Ka/Ks
50
40
30
20
10
0
4300
4500
4700
4900
5100
Nucleotide Position

Figure 3.6 Sliding window analysis of substitutions (K_{nc}) and polymorphism (π_{nc}) in the non-coding regions (upstream, downstream and introns) of gene CG15098. Coding regions were excluded from analysis.

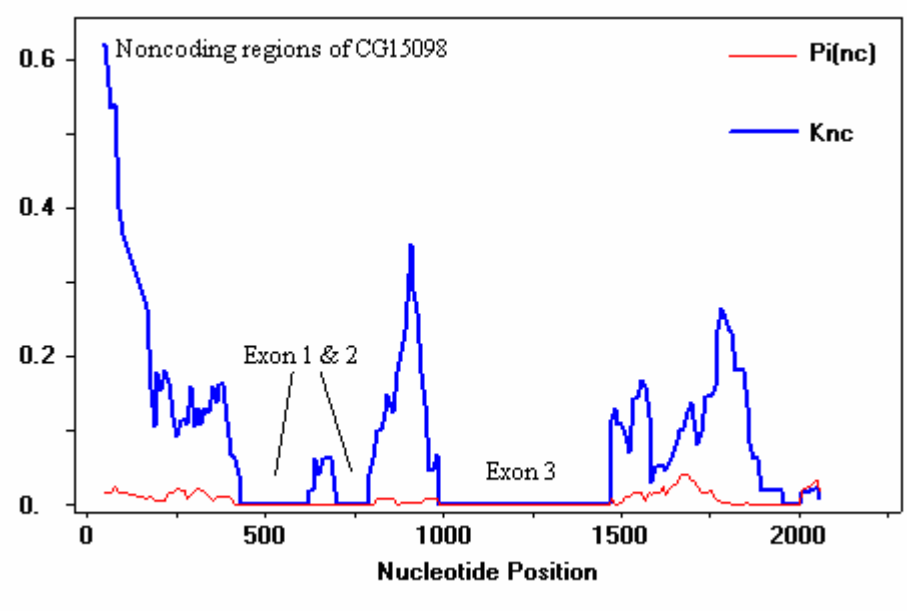

Noncoding regions of CG15098
Pi(nc)
Knc
0.6
0.4
0.2
0.
Exon 1 & 2
Exon 3
0
500
1000
1500
2000
Nucleotide Position

Figure 3.7 Sliding window analysis of substitutions (K_{nc}) and polymorphism (π_{nc}) in the non-coding regions (upstream) of the genes in the protease cluster (CG17012, CG17234, CG17239 and CG17240 (Ser12)). Coding regions were excluded from analysis.

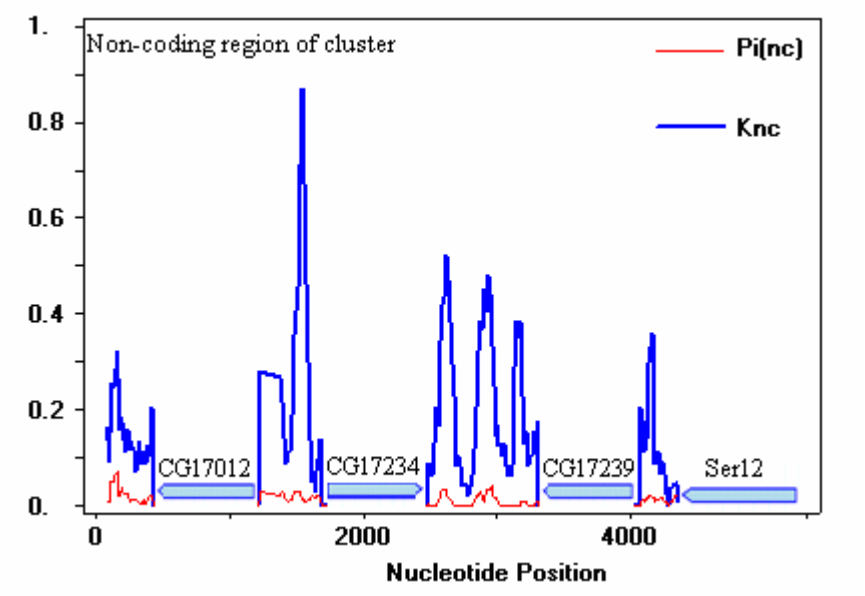

Non-coding region of cluster
Pi(nc)
Knc
1.
0.8
0.6
0.4
0.2
0.
CG17012
CG17234
CG17239
Ser12
0
2000
4000
Nucleotide Position

Table 3.1 Polymorphism summary statistics for the coding regions in the cluster of proteases and the gene of unknown classification.

Gene	Pop	size	n	S	π_A	π_S	π_A/π_S
CG17012	all	240	35	21	0.006	0.017	0.368
	PA	766	19	27	0.008	0.016	0.480
	UG	240	16	19	0.008	0.017	0.445
CG17234	all	179	35	26	0.007	0.020	0.367
	PA	572	19	6	0.0003	0.002	0.162
	UG	202	16	29	0.014	0.032	0.424
CG17239	all	724	35	22	0.004	0.011	0.377
	PA	747	19	16	0.004	0.008	0.454
	UG	724	16	18	0.005	0.013	0.387
CG17240	all	n/a					
	PA	161	19	23	0.019	0.017	1.112
	UG	n/a					
CG15098	all	980	35	25	0.006	0.007	0.808
	PA	980	19	16	0.005	0.005	1.086
	UG	981	16	22	0.005	0.009	0.533
Gene average	all				0.006	0.009	1.306
	PA				0.007	0.010	0.659
	UG				0.008	0.018	0.447
Genomic average from literature					0.001	0.009	0.111

all: combined data analysis from both the Pennsylvania and Uganda lines; PA: Pennsylvania lines; UG: Uganda lines; Size: the length of sequence data used for each gene; n: the number of lines used for analysis; S; the number of segregating sites; π_A:a measure of nonsynonymous nucleotide diversity; π_S: a measure of synonymous nucleotide diversity.

Table 3.2 Neutrality test summary statistics for the coding and non-coding regions associated with the cluster of genes and gene of unknown function, and the non-coding random genomic region.

		Fu and Li's D	p	Tajima's D	p	Fay and Wu's H	p
Genomic	PA	0.48	ns	0.09	ns	-2.86	ns
	UG	-1.70	ns	-1.27	ns	-8.38	0.02
Upstream 17012	PA	0.38	ns	0.81	ns	-0.02	ns
	UG	-2.18	0.05	-1.58	0.05	-0.29	ns
CG17012 CDS	PA	-0.59	ns	-0.42	ns	-2.65	ns
	UG	-2.44	0.05	-1.94	0.05	0.94	ns
Upstream 17234	PA	-3.05	0.02	-2.03	0.05	-0.88	ns
	UG	-2.59	0.02	-2.03	0.05	-0.05	ns
CG17234 CDS	PA	-2.76	0.02	-2.02	0.05	0.66	ns
	UG	-3.31	0.02	-2.21	0.01	2.80	ns
Upstream 17239	PA	1.51	0.02	1.56	ns	-1.14	ns
	UG	-2.81	0.02	-1.01	ns	0.73	ns
CG17239 CDS	PA	-0.60	ns	-0.90	ns	-5.86	0.01
	UG	-0.69	ns	-0.06	ns	-0.16	ns
Upstream 17240	PA	-2.66	0.05	-2.05	0.05	-3.90	0.05
CG17240 CDS	PA	-3.25	0.02	-2.35	0.01	0.74	ns
Upstream 15098	PA	-0.47	ns	-0.63	ns	-1.31	ns
	UG	0.03	ns	0.51	ns	-2.90	ns
15098 CDS	PA	-0.12	ns	-0.34	ns	-2.83	ns
	UG	-0.99	ns	-0.65	ns	-3.40	ns
15098 introns	PA	0.40	ns	0.55	ns	-0.97	ns
	UG	0.43	ns	-0.20	ns	-1.92	ns
Downstream 15098	PA	2.17	0.05	-1.50	0.01	1.46	ns
	UG	-0.88	ns	-0.35	ns	2.83	ns

Population summary statistics for neutrality tests. PA: Pennsylvania lines; UG: Uganda lines; P-values for Fu and Li's D and Tajima's D were obtained using a 2-tailed Fisher's exact test. P-values for Fay and Wu's H were calculated using coalescent simulations; ns: non-signficant.

Table 3.3 Analysis of the divergence in the cluster of proteases between D.melanogaster and D. simulans, including non-coding, promoter and coding regions.

		Coding regions			Non-coding regions			
Gene		**K_a**	**K_s**	**K_a/K_s**	**K_u**	**K_u/K_g**	**K_p**	**K_p/K_g**
CG17012	all	0.09	0.11	0.83	0.10	1.55	0.08	1.27
	PA	0.11	0.13	0.84	0.13	2.12	0.01	0.09
	UG	0.09	0.11	0.86	0.10	1.52	0.09	1.32
CG17234	all	0.06	0.13	0.48	0.23	3.56	0.10	1.56
	PA	0.07	0.12	0.56	0.22	3.41	0.15	2.37
	UG	0.07	0.15	0.46	0.22	3.42	0.10	1.58
CG17239	all	0.07	0.11	0.68	0.13	1.97	0.10	1.53
	PA	0.07	0.11	0.67	0.18	2.81	0.10	1.58
	UG	0.07	0.10	0.69	0.13	1.95	0.10	1.53
CG17240	all	n/a						
	PA	0.06	0.08	0.82	0.11	1.72	0.04	0.57
	UG	n/a						
Gene average	all	0.07	0.11	0.66	0.15	2.36	0.09	1.45
	PA	0.08	0.10	0.74	0.17	2.52	0.07	1.15
	UG	0.08	0.11	0.71	0.16	2.30	0.10	1.48

		all	PA	UG
Genomic region	K_g	0.064	0.063	0.066

K_a:nonsynonymous changes per nonsynonymous site; K_s - synonymous changes per synonymous site; K_u - silent changes in immediately upstream regions; K_g - silent changes in the randomly selected non-coding genomic region of the 2nd chromosome; K_p - changes in predicted promotor regions

Table 3.4 Analyses of the divergence in CG15098 between *D.melanogaster* and *D. simulans*, including non-coding, promoter and coding regions.

		Coding regions			Non-coding regions							
Gene name		**K_a**	**K_s**	**K_a/K_s**	**K_u**	**K_u/K_g**	**K_i**	**K_i/K_g**	**K_d**	**K_d/K_g**	**K_p**	**K_p/K_g**
CG15098	all	0.06	0.07	0.81	0.18	2.77	0.10	1.56	0.08	1.25	0.02	0.28
	PA	0.06	0.07	0.80	0.20	3.11	0.10	1.60	0.08	1.26	0.02	0.31
	UG	0.06	0.07	0.82	0.20	3.04	0.10	1.56	0.08	1.21	0.01	0.22
		all	PA	UG								
Genomic region	K_g	0.06	0.06	0.07								

K_a:nonsynonymous changes per nonsynonymous site; K_s - synonymous changes per synonymous site; K_u - silent changes in immediately upstream regions; K_i - silent changes in intronic regions; K_d - silent changes in regions immediately downstream; K_p - changes in predicted promotor region

Table 3.5: HKA summary statistics for the cluster of proteases and the gene of unknown function. HKA tests results are calculated from a comparison of neutral genomic regions compared to the coding regions of the genes.

Gene	Population	Seg. Sites	Ave. diff.	Div.Time	*P*
CG17012	PA	28	6.43	-0.13	0.93
	UG	19	2.65	-0.43	0.90
CG17234	PA	6	0.71	-0.16	0.70
	UG	29	3.22	-0.49	0.84
CG17239	PA	16	3.63	-0.11	0.92
	UG	18	5.16	-0.26	0.86
CG17240	PA				
	UG				
CG15098	PA	12	2.59	-0.12	0.90
	UG	15	3.37	-0.33	0.94

Gene: *D. melanogaster* CG number; PA: Pennsylvania lines; UG: Uganda lines; Seg. Sites: segregating sites; Ave. diff: average number of differences between *D. melanogaster* and *D. simulans*; P: P-value obtained from a X^2 test (df =1) of the difference between null and selection models.

Table 3.6: McDonal d-Kreittman test statistics for the coding regions of the protease cluster and the gene of unknown function

		Polymorphic		Fixed		
Gene	**Population**	**Silent**	**Replacement**	**Silent**	**Replacement**	**Prob.**
CG17012	PA	9	15	25	56	0.62
	UG	4	8	4	12	0.69
CG17234	PA	2	1	16	28	0.55
	UG	5	7	8	13	1
CG17239	PA	7	9	18	37	0.55
	UG	8	10	17	35	0.37
CG17240	PA	2	6	6	17	1
	UG					
CG15098	PA	13	25	2	11	0.30
	UG	12	26	5	12	1

Gene: D. melanogaster CG number; PA: Pennsylvania; UG: Uganda; Prob: P-values were obtained using a 2-tailed Fisher's exact test

Chapter IV.

Patterns of gene expression in the sperm storage organs of *Drosophila melanogaster*

[Based on the manuscript submitted to *Insect Molecular Biology* by Adrianne M. Prokupek, Stephen D. Kachman, Lawrence G. Harshman]

ABSTRACT

Sperm storage is a dynamic and complex process utilized by female animals across a wide range of taxa. Sperm storage organs play an important role in reproduction and are important in evolutionary processes such as speciation. The function of these organs is poorly understood, especially at the protein level. This study investigated the transcriptome of the two *Drosophila melanogaster* sperm storage organs and identified genes likely to impact sperm storage, maintenance and use. Functional categorization of the genes expressed in the SSOs indicates that each organ play a unique roles in the process of sperm storage. The spermathecae was enriched for proteases, metabolism genes and antioxidants. The seminal receptacle exhibited a number of genes involved in localization, signaling, ion transport and immunity/defense.

INTRODUCTION

Transcriptome analysis using microarrays provides a means to investigate gene expression and identify proteins in organs for which little functional information is available. Genes expressed in these tissues/organs can be grouped by function, allowing for the identification of proteins that play significant functional roles. The female sperm storage organs (SSOs) are important in reproduction and various evolutionary processes, but little is currently known about the genes expressed in these organs for any species. There is a general lack of functional knowledge of SSOs creating a major deficit in the understanding of sperm storage, an important process of reproduction.

Female sperm storage is a reproductive strategy utilized by most internally fertilizing animals. The structure and function of SSOs appear to be both dynamic and complicated (Bloch-Qazi & Wolfner, 2003, Adams & Wolfner, 2007). Sperm storage allows females to acquire sperm to take place days or even months before it is needed to fertilize mature oocytes. The reproductive benefits of this process are potentially great as contact with a suitable male does not necessarily correspond with egg availability. For species which engage in multiple matings, sperm storage can offer an additional selection step via sperm competition and/or cryptic female sperm choice, as well as gives the female a mechanism to compensate for infertile/genetically incompatible males. The duration of time sperm spend in storage is highly variable, extending from a few hours (e.g., mouse) up to decades (e.g., honey bee). Specialized SSOs have evolved in a number of species; in insects SSOs typically appear as sac-like structures (spermathecae) or long tubules (seminal receptacles). These organs are responsible for maintaining sperm

viability, organizing the sperm in storage, and facilitating the proper release of sperm from storage (reviewed in (Neubaum & Wolfner, 1999c, Bloch-Qazi et al., 2003).

Drosophila melanogaster females mate multiply and store sperm in specialized SSOs (Lefevre & Jonsson, 1962, DeVries, 1964, Pitnick et al., 1999). Copulation in *D. melanogaster* results in the transfer of approximately 4000-6000 sperm (Kaplan et al., 1962), approximately a quarter of the transferred sperm are stored, and an estimated 30-80% of stored sperm are used for fertilization (Kaplan et al., 1962, Fowler, 1973). The process of sperm storage in *D. melanogaster* begins before copulation has ended (Fowler, 1973, Gilbert et al., 1981, Lefevre & Jonsson, 1962), and is complete within 6 hours after mating (Lefevre & Jonsson, 1962, Gilbert, 1981, Neubaum & Wolfner, 1999b, Tram & Wolfner, 1999, Bloch-Qazi & Wolfner, 2003). *D. melanogaster* have two types of SSOs; the seminal receptacle (SR) and paired spermathecae (ST). The SR stores the majority of the sperm (65-80%) and is the first SSO to store sperm as well as the first to release sperm (Gilbert, 1981, Neubaum & Wolfner, 1999c). The ST is considered to be the long term SSO of *Drosophila*. Sperm accumulate slower in the ST and are utilized after the sperm in the SR are depleted. Within 1.5 hours of mating ovulation begins and by 3 hours the female is laying fertilized eggs (Heifetz et al., 2000), indicating that the SR is actively storing and releasing sperm simultaneously. By 10 hours post-mating a noticeable decline is seen in the sperm stores of the SR, and by 48 hours the SR sperm stores are ~50% used, whereas the ST sperm stores are only ~15% depleted (Neubaum & Wolfner, 1999b).

It is unclear why *Drosophila* species require both the ST and the SR for sperm storage. It has been suggested that the SR is more efficient at storing sperm, but the

sperm within the SR are more susceptible to displacement by the sperm of other males (Pitnick et al., 1999, Civetta, 1999, Price et al., 1999). The ST may protect sperm from displacement. Secretory cells line the walls of the ST and secrete proteins into the lumen where sperm are stored; these proteins could function to keep sperm viable (Filosi & Perotti, 1975, Anderson, 1945, Pitnick et al., 1999).

Parallels between reproductive proteins of distantly related taxa have been identified, indicating that there is some degree of conservation. Genes encoding proteins have been identified in Dipteran species which closely resemble mammalian reproductive proteins both in sequence and putative function. A study by Allen and Spradling (2008) investigated proteins expressed within the ST in *D. melanogaster*. They discovered that products secreted by the ST are necessary for sperm maturation and function, paralleling the role of the mammalian epididymis and female reproductive tract. The gene hormone receptor 39 (Hr39) in *Drosophila* is a putative hormone receptor closely related to the steroidogenic factor 1 (Sf1) nuclear hormone receptor of mammals. Sf1 is active during embryogenesis and is necessary for proper reproductive tract development (Allen & Spradling, 2008). The Allen and Spradling study revealed a closer connection between Dipteran and mammalian reproductive biology than previously believed and sets the stage for use of *D. melanogaster* as a model for human reproduction.

The current study utilized a transcriptome analysis was conducted for each of the SSOs of *D. melanogaster*, before and after mating. The gene expression data from this project provides insight into the identity and expression patterns of genes expressed in the sperm storage organs. The information is potentially applicable to a wide range of species. This study identified genes differentially expressed within each SSO and

between the ST and SR of *Drosophila melanogaster* at two time points post-mating, providing the most detailed information on sperm storage gene expression patterns available for any species.

MATERIALS and METHODS

Stock maintenance

The Canton-S (CS-C) stock of *D. melanogaster* were used in the present study. Flies were reared in larval density controlled vials, on a standard *Drosophila* diet in a temperature controlled environment (25 degree Celsius) with a standard 12/12 light/dark cycle. Adults were collected as virgins. Virginity was confirmed by observing that no larvae appeared in the vials.

Mating and sample collection

Virgin females between four and seven days of adult age were used for virgin dissection of SSOs and for matings. For mated female collections, a single virgin female was placed with a single male. The female was separated from the male immediately after the observed mating had ended naturally and the timing of post-mating began. Tissues were collected from virgin, 3 hour post-mating, and 6 hour post-mating females. Females were dissected on a cold plate in RNase later (Ambion).

Approximately 80 females were used for each dissection event. The ST and SR were removed from each female and placed into microcentrifuge tubes containing TRIzol (Ambion) for RNA extraction. The ST sample included the spermathecae from each

female and the attached spermathecal duct. From each dissection period 2 samples were obtained, one of ~160 ST and the other of ~ 80 SRs.

RNA extraction

Freshly dissected tissues were ground in TRIzol and liquid nitrogen using mortar and pestle. RNA was extracted and the samples were then cleaned using Qiagen micro RNA clean-up columns and quantified using a NanoDrop (ND-1000) spectrophotometer (NanoDrop Technologies, Inc). To obtain sufficient quantities of RNA from ST and SR, the RNA for each sample was pooled from 10 independent dissection periods and extractions. Each ST sample used for one microarray chip contained the RNA from ~ 1600 ST (80 females X 2 ST per female X 10). Each SR sample used for one microarray chip contained the RNA from ~ 800 SR (80 females X 1 SR per female X 10).

Microarray assays

ST and SR microarray cRNA sample preparation used a two-cycle target and labeling kit (Affymetrix, Santa Clara, CA), which is designed for starting RNA levels of ≤ 100ng for each microarray chip. For both the ST and SR 100ng of total RNA were used for amplification prior to cDNA synthesis followed by cRNA synthesis.
The Gene Chip Drosophila genome 2.0 array (Affymetrix), which includes probe sets to 18880 *D. melanogaster* genes, was used to measure gene expression. Hybridization of cRNA to the Drosophila Gene Chips was performed by the Microarray Core Facility at the University of Nebraska-Lincoln. All microarray hybridizations were performed using three replicate chips for each treatment (virgin ST and SR, 3 hours post-mating ST and

ST, 6 hours post-mating ST and SR) using 18 independent samples of RNA extracted from different sets of flies.

Microarray analysis

The Robust Multichip Averaging (RMA) procedure (Irizarry et al. 2003) implemented in the Affy module of the BioConductor microarray analysis package was used to background correct and normalize the microarray data. The resulting measurements of hybridization intensity were used as a proxy for gene expression levels. Average hybridization intensity was calculated for each gene using the three arrays for each sample. A cut-off value of 100 for average hybridization intensity was selected to help prevent the inclusion of genes which are not truly expressed in the sample. Using a cut-off value of 100 selected for approximately one-third of the genes.

To detect temporal patterns of female-expressed genes we performed pairwise comparisons of the hybridization data from virgins, 3 hour post-mated females, and 6 hour post-mated females using both SR and ST. To detect patterns of gene expression changes between organs, pairwise comparisons were made between the ST and SR at each time point. Differential expression of genes was determined using both fold change (FC), and false discovery rate (FDR). FDR is defined as the expected proportion of false positives among the declared significant results, and provides a more direct interpretation of microarray data in comparison with standard p-values (Pawitan et al., 2005, Benjamini & Hochberg, 1995) The cut-offs used for differentially expressed genes were FC ≥ 2, and a FDR q-value ≤ 0.1.

Functional analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID) (http://david.abcc.ncifcrf.gov) was used to classify gene function for the genes present in all lists using the most relevant gene ontology (GO) associated terms. The associated fourth level molecular function (MF) and biological process (BP) is listed for each of the expressed genes. DAVID was used to determine significant enrichment of known functional annotations within our differentially expressed gene lists. DAVID calculates the probability for representation of genes within a given category for an input list based on the observed representation number of genes, within the same category for a pre-defined background list. The predefined *D. melanogaster* background list consisted of GenBank accession numbers for each unique transcript probe on our arrays. All significantly ($p < 0.05$) enriched functional groups within all levels of biological process (BP) and molecular function (MF) was reported in the Supporting Information tables.

The Protein ANalysis THrough Evolutionary Relationships (PANTHER) (http://www.pantherdb.org) classification system was also used to classify genes into broad functional categories. Genes are classified into families and subfamilies using hidden Markov models (HMMs) based on shared functions, then further classified by molecular function and biological process ontology terms. PANTHER categorized the genes into more generalized categories compared to DAVID. The ontology of PANTHER uses a controlled vocabulary of molecular function and biological process arranged as directed acyclic graphs, similar to the Gene Ontology (GO), but abbreviated and simplified to facilitate high-throughput analysis. The PANTHER database contains annotation data for 14,115 *Drosophila melanogaster* genes.

RESULTS and DISCUSSION

In this study, we investigated gene expression changes in the SSOs of *D. melanogaster* at two time points (3 and 6 hours) after mating. A pre-mating (virgin) state was used as a baseline for the determination of gene expression patterns after mating. Three hour post-mating was used because this is the time needed for the complete entrance of sperm into storage, as well as the start of egg deposition and fertilization (Heifetz et al., 2000, Heifetz et al., 2001). Six hours post-mating was used based on a previous study which showed a spike in gene expression in the lower reproductive tract six hours after the completion of mating (Mack et al., 2006b). We found distinctly different patterns of gene expression for the two types of organs, and for pre- versus post-mating time points. As expected based on the fact that both of these organs store sperm, a substantial overlap in the genes expressed was also found. The distinct patterns produced for the two sperm-storage organs will be valuable for future functional research of sperm storage and studies on the role of SSOs in evolution.

Average expression

Genes having a hybridization intensity of greater than 100 after background correction were considered to be expressed. Average hybridization intensities as well as DAVID gene ontology (GO) categories for fourth level biological processes (BP) and molecular functions (MF) is listed for genes in the ST and SR at virgin, 3 hour and 6 hour time points (Supplementary table 1, 2). The highest expressed gene, based on hybridization intensity, at all three time points in the SR is a gene of unknown function.

Genes of unknown function can be interesting in terms of their potential to play organ-specific roles. Genes classified as unknown could play roles that are unique to sperm storage, or may represent new classes of genes which have not been identified in other, organs of *Drosophila* or other taxonomic groups..

At all three time points the gene in the ST with the highest expression was a serine protease (CG9897 in virgin and CG17239 at both 3 and 6 hours post-mating). Serine proteases have been identified as important molecules in reproduction, both in the seminal fluid and in the female reproductive tract (Ravi-Ram & Wolfner, 2007, Lung et al., 2002a, Lawniczak & Begun, 2004, Swanson et al., 2001a). The hybridization intensity of protease CG17239 remains highest in the ST 3 days after mating (Allen & Spradling, 2008), alluding to the importance of this particular protein for ST function. CG17239 is found on the long arm of chromosome 2, in a cluster of 4 proteases (CG17239, CG17012, CG17234, and CG17240), all of which have secretion signals (Prokupek et al.). ST proteins with secretion signals could directly interact with male seminal fluid proteins or sperm in the lumen of this organ. Three of the four proteases found in this cluster had high expression in the ST at all time points (Table 4.1), suggesting they might be necessary for storing sperm. The four proteases comprising the cluster were discovered in a previous EST study of the ST at three hours post-mating (Prokupek et al) (CG17240 was not found in the current microarray study), and found that these genes are enhanced for expression in the ST. The overlap in the ST of the genes of these two studies, as well as the low level of expression of these genes in the SR, suggests that the proteases in this cluster play roles important to the ST.

All three of the yolk proteins in *D. melanogaster* (YP1-YP3) were highly expressed in the ST at all time points, but only YP1 had high expression in the SR. In a study by Allen and Spradling (2008), YP1, 2 and 3 were found to be highly expressed in the spermathecae three days post-mating. Yolk proteins, found in tissues such as fat bodies and follicle cells, are an integral component of yolk in eggs (Brennan et al., 1982, Barnett et al., 1980). The presence of these proteins in the SSOs may indicate that these organs contribute to yolk production for eggs as suggested by Allen and Spradling (2008). Alternatively, YPs might play a novel role in the SSOs by protecting sperm from oxidative damage. In honeybees, the antioxidant activity of vitellogenins (similar to yolk proteins) apparently promotes individual survival (Seehuus et al., 2006), suggesting YPs might similarly promote the survival of sperm.

Molecular function

PANTHER was used to group the top 100 genes, based on hybridization intensity, into categories of molecular function. Proteases made up a high proportion (almost 10 percent in most cases) of the top expressed genes of the ST. Structural proteins, proteins involved in transport, muscular contraction and ion channels, synthetases and hydrolases were highly expressed in the SR at all time points. Perhaps transport and ion channel activities act to generate a suitable environment for stored sperm before and after mating. Both the ST and SR showed a high number of genes involved in nucleic acid binding (MF) and protein metabolism (BP); most of the nucleic acid binding genes encode ribosomal proteins (Supplementary figure 1).

Differential gene expression

Differential gene expression was measured by pairwise comparisons of virgins to the two time points post-mating for ST and SR. A larger number of genes were differentially expressed within the SR compared to the ST at both time points (Figure 4. 1). In comparison with our study, far fewer genes were differentially expressed in the whole body of *D. melanogaster,* or lower reproductive tract after mating, compared to virgins (McGraw et al., 2004, Lawniczak & Begun, 2004). A possible explanation for the paucity of genes found to be differentially expressed in the whole body after mating is that opposing gene expression patterns in different organs negate each other (McGraw et al., 2004). The high number of differentially expressed genes in the SR compared to the ST (Figure 4.1) can possibly be explained by the fact that the SR is both actively storing and releasing sperm at 3 hours post-mating, and by 6 hours is presumably releasing sperm at a steady rate; the spermathecae is presumably only storing sperm at both of these time points.

Overrepresented gene categories

Up and down regulated genes were categorized using DAVID. Statistically overrepresented gene categories ($p < .05$) are presented for ST and SR in Supplementary table 6 and 7 respectively. Among overrepresented genes upregulated in the ST (at each time point), the functional categories that stand out include genes involved in metabolic pathways, such lipid, carbohydrate metabolism, and genes with catalytic activity such as juvenile hormone catabolism. Downregulated, overrepresented categories of genes of the

ST included metabolism and catabolism pathways, electron transport, and catalytic activity.

Upregulated functional categories of genes overrepresented in the SR included genes involved in immunity and defense, RNA modification, metabolism (peptidoglycan), female gamete generation, and catalytic activity. Downregulated overrepresented gene categories included metabolism (carbohydrate, glycogen), oxoreductase activity, and defense response. An overrepresentation of serine type peptidases in the SR was seen in upregulated genes at three hours ($p = 0.01$), and in down regulated genes at six hours ($p = 0.03$), suggesting a peak in proteolysis at the three hour time point.

The abundance of immunity/defense genes upregulated in the SR is not mirrored in the ST samples. This indicates that while the ST may be the primary SSO for protein modification (high level of proteases), the role of the SR may be more defense from introduced pathogens.

Organ-specific changes in gene expression

To detect patterns of gene expression, we compared hybridization data from the SR and ST. Genes were considered differentially expressed if the resulting fold change was >2, and q-value < 0.1. Comparing virgin SR and ST, a total of 1,794 genes were differentially expressed; 687 genes upregulated and 1107 downregulated in the ST compared to the SR. For 3 hour post-mating SR and ST, a total of 1,589 genes were differentially expressed; 621 genes upregulated and 968 downregulated in the ST compared to the SR. At 6 hour post-mating SR and ST, a total of 2,538 genes were

differentially expressed; 1,054 genes upregulated and 1,484 downregulated in the ST compared to the SR (Supplementary table 5).

Analysis of the data in the preceding paragraph revealed interesting patterns. Overrepresented categories of genes, those statistically enriched compared to a random expectation, upregulated in the ST in comparison to the SR in virgin and both times post-mating include metabolism, defense/immunity, transport, binding, and proteolysis. Genes upregulated in the SR (downregulated in the ST by comparison) fell into the following categories: cell communication, ion transport, development, defense/immunity (different genes from those relatively upregulated in the ST), localization, and homeostasis (Supplementary Figure 4.2). Apparently, there are few defense/immunity genes held in common in the two different sperm storage structures. The proteins involved in cell communication, ion transport, and localization within the SR may be facilitating the proper timing of release of sperm from this storage organ. A complete list of overrepresented categories is presented in Supplementary table 8. The large number of differentially expressed genes is surprising, as one might assume that the organs have fundamentally similar functions. At each time point, the organs may have very different functions; the SR both storing and releasing sperm, and the ST preparing sperm for long term storage.

General discussion

During the time of mating male *Drosophila* pass to females a number of proteins via their seminal fluid. Some of these proteins are transferred to the SSOs where they may elicit responses that are common to both types of organ. These proteins, especially

male accessory gland proteins (Acps), are responsible for eliciting a suite of post-mating responses in the female. Female responses include increases in ovulation, ovipositioning, and mating refractoriness, as well as sperm storage and decreases in female lifespan (Wolfner, 1997, Wolfner, 2002). At least one Acp (Acp36DE) has been associated with proper sperm storage. Acp36DE associates with the sperm mass and localizes at the openings of the SSOs as well as within them (Bertram et al., 1996, Neubaum & Wolfner, 1999b). Another Acp, sex peptide (SP) enters the SSOs bound to the sperm tail, and is released over time. This protein stimulates juvenile hormone synthesis. (Liu & Kubli, 2003, Chapman et al., 2003a). There may be a connection between juvenile hormone catabolism in the ST and juvenile hormone synthesis induced by SP. Specifically, females may be acting to counter a detrimental physiological effect induced by males; juvenile hormone increases eggs production which shortens the life span of females.

Female genes found to be differentially expressed in the present study might directly interact with Acps. For example, the cleavage of SP is accomplished by a trypsin member of the serine protease family (Peng et al., 2005b). Most Acps are no longer detected after 6 hours postmating, but SP continues to be cleaved from sperm for the entire time sperm are in storage (Peng et al., 2005b). The identity of the specific serine protease involved in the release of SP is not known, but it is quite possibly female derived. Four of the seven Acps which enter the SSOs are serine protease inhibitors (serpins) (Mueller et al., 2005b). Serpins bind to serine proteases, blocking their proteolytic function, and they might interact with proteases in the ST. Proteases found within SSOs may also be interacting with male proteins in cascades, similar to those seen

in immune function and blood or sperm coagulation (Overduin & de Beer, 2000, Malm et al., 2000, Kim et al., 2008).

Although ST are found in a wide range of invertebrate and vertebrate taxa (Eberhard, 1996) the function of proteins and other macromolecules associated with this organ are scarcely understood. An exception is social insects - bees and ants - in which the function of ST proteins have been best studied. Reproductively capable females (queens) mate with several males early in life and rely on stored sperm to fertilize eggs over many years. The reproductive success of the queen is directly dependent on quality of sperm in storage, which creates strong selection for both sperm viability as well as efficient utilization of sperm stores (Baer et al., 2006). In the social ant (*Crematogaster opuntiae*), as well as in honey bees, the secretory cells surrounding the ST have ample glycogen, serving as a possible energy reserve for sperm in storage (Wheeler & Krutzsch, 1994). In honeybees the ST fluid contains sugars including glucose, trehalose and fructose, and a high level of trehalase activity (Alumot et al., 1969b). Enzymes with antioxidant activity have been found in the ST of honeybees, presumably acting to protect stored sperm from oxidative damage (Weirich et al., 2002, Collins et al., 2004b). Perhaps YPs found in the SSOs of *D. melanogaster* play a similar antioxidative function.

This study found that differentially expressed (up and down) genes in diverse metabolic pathways were overrepresented. The differential expression of genes related to metabolism may correspond to the changing nutritional/environmental needs of sperm as they are going through steps of sperm storage and release. Genes for carbohydrate/lipid metabolism are expressed at higher levels in the ST when compared to the SR at all time points, indicating a role in long term sperm storage (Supplementary Figure 4.1).

McGraw *et al.* (2004) used microarrays to compare gene expression of *D. melanogaster* females which were virgin, mated to wild-type males, mated to spermless males, or mated to males lacking Acps. A genome-wide comparative microarray analysis was also done on virgin, courted, and two-hour post-mated females (Lawniczak & Begun, 2004). In both studies, immune-related genes and serine proteases were found to be affected by mating. The number of serine proteases influenced by mating was determined to be higher than expected by chance (Lawniczak & Begun, 2004). The role of such proteases within the female is unknown. Induction of both serine proteases and immune/defense related genes was observed in the present study. Comparisons between the SR and the ST showed similar numbers of immune/defense genes, but higher levels of proteases in the ST.

CONCLUSIONS

Microarrays were used to investigate the transcriptome of *Drosophila melanogaster* sperm storage organs in virgin females and at two time points post-mating. A high number of SSO genes were differentially expressed following mating. At each time point high numbers of genes were differentially expressed between the two sperm storage organs, indicating that each SSO plays a unique role in the process of sperm storage. The spermathecae was enriched for genes involved in proteolysis and metabolism. These genes could be interacting with male proteins such as protease inhibitors or encode proteins that provide lipids/carbohydrates for sperm maintenance. Other spermatheca proteins, such as yolk proteins, could protect sperm perhaps by acting as an antioxidant. The seminal receptacle exhibited a number of overrepresented genes

involved in localization, signaling, and ion transport. These genes could be working to maintain a homeostatic environment, as well as serving communication roles between the SR and oocytes to ensure the proper timing of fertilization. More genes were differentially expressed in the seminal receptacle perhaps due to the dual role of sperm storage and sperm release played by this organ at the time points used in the present study. Further investigation of the roles of proteins in these two organs will allow for a more complete understanding of genes playing essential roles in the process of sperm storage, maintenance, and use.

ACKNOWLEDGEMENTS

This research was supported by a National Science Foundation grant (DEB-ESP0346476).

REFERENCES

Adams, E. M. & Wolfner, M. F. 2007. Seminal proteins but not sperm induce morphological changes in the Drosophila melanogaster female reproductive tract during sperm storage. *J Insect Physiol* **53**: 319-31.

Allen, A. K. & Spradling, A. C. 2008. The Sf1-related nuclear hormone receptor Hr39 regulates Drosophila female reproductive tract development and function. *Development* **135**: 311-21.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic local alignment search tool. *J Mol Biol* **215**: 403-10.

Alumot, E., Lensky, Y. & Holstein, P. 1969a. Sugars and trehalase in the reproductive organs and hemolymph of the queen and drone honey bees (Apis mellifera L.Var. Ligustica Spi.),. *Comp Biochem Physiol B* **28**: 1419-25.

Alumot, E., Lensky, Y. & Holstein, P. 1969b. Sugars and trehalase in the reproductive organs and hemolymph of the queen and drone honey bees (Apis mellifera L.Var. Ligustica Spi.),. *Comparative Biochemistry and Physiology* **28**: 1419-25.

Anderson, R. C. 1945. A Study of the Factors Affecting Fertility of Lozenge Females of Drosophila Melanogaster. *Genetics* **30**: 280-96.

Andolfatto, P. 2005. Adaptive evolution of non-coding DNA in Drosophila. *Nature* **437**: 1149-52.

Andres, J. A., Maroja, L. S., Bogdanowicz, S. M., Swanson, W. J. & Harrison, R. G. 2006. Molecular evolution of seminal proteins in field crickets. *Mol. Biol. Evol.* **23**: 1574-84.

Arbeitman, M. N., Fleming, A. A., Siegal, M. L., Null, B. H. & Baker, B. S. 2004. A genomic analysis of Drosophila somatic sexual differentiation and its regulation. *Development* **131**: 2007-21.

Arnqvist, G. & Rowe, L. 1995. Sexual conflict and arms races between the sexes: a morphological adaptation for control of mating in a female insect. *Proc. R. Soc. Lond B* **261**: 123-127.

Arthur, B. I., Hauschteck-Jungen, E., Nothiger, R. & Ward, P. I. 1998. A female nervous system is necessary for normal sperm storage in *Drosophila melanogaster*: a masculinized nervous system is as good as none. *Proc. R. Soc. Lond B* **265**: 1749-53.

Baer, B., Armitage, S. A. O. & Boomsma, J. J. 2006. Sperm storage induces an immunity cost in ants. *Nature* **441**: 872 - 76.

Barnett, T., Pachl, C., Gergen, J. P. & Wensink, P. C. 1980. The isolation and characterization of Drosophila yolk protein genes. *Cell* **21**: 729-38.

Begun, D. J. 2002. Protein variation in Drosophila simulans, and comparison of genes from centromeric versus noncentromeric regions of chromosome 3. *Mol Biol Evol* **19**: 201-3.

Begun, D. J. & Aquadro, C. F. 1993. African and North American populations of Drosophila melanogaster are very different at the DNA level. *Nature* **365**: 548-50.

Begun, D. J. & Whitley, P. 2000. Reduced X-linked nucleotide polymorphism in Drosophila simulans. *Proc Natl Acad Sci U S A* **97**: 5960-5.

Begun, D. J., Whitley, P., Todd, B. L., H.M., W.-D. & Clark, A. G. 2000a. Molecular population genetics of male accessory gland proteins in *Drosophila*. *Genetics* **156**: 1879 - 88.

Begun, D. J., Whitley, P., Todd, B. L., Waldrip-Dail, H. M. & Clark, A. G. 2000b. Molecular population genetics of male accessory gland proteins in Drosophila. *Genetics* **156**: 1879-88.

Benjamini, Y. & Hochberg, Y. 1995. Controlling the false discovery rate: a pratical and powerful approach to multiple testing. *J Roy Stat Soc* **57**: 289 - 300.

Bertram, M. J., Neubaum, D. M. & Wolfner, M. F. 1996. Localization of the Drosophila male accessory gland protein Acp36DE in the mated female suggests a role in sperm storage. *Insect Biochem Mol Biol* **26**: 971-80.

Birkhead, T. R. & Moller, A. P. 1998. *Sperm competition and sexual selection*. Academic Press, London.

Bloch-Qazi, M. C., Aprile, J. R. & Lewis, S. M. 1998. Female role in sperm storage in the red flour beetle, *Tribolium castaneum*. *Comparative Biochemistry and Physiology* **120**: 641-648.

Bloch-Qazi, M. C., Heifetz, Y. & Wolfner, M. F. 2003. The developments between gametogenesis and fertilization: ovulation and female sperm storage in *Drosophila melanogaster*. *Developmental Biology* **256**: 195-211.

Bloch-Qazi, M. C. & Wolfner, M. F. 2003. An early role for the *Drosophila melanogaster* male seminal protein Acp36DE in female sperm storage. *Journal of Experimental Biology* **206**: 3521-3528.

Bork, P. & Beckman, G. 1993. The CUB domain. A widespread module in develomentally regulated proteins. *J. Mol. Biol*: 539-45.

Brennan, M. D., Weiner, A. J., Goralski, T. J. & Mahowald, A. P. 1982. The follicle cells are a major site of vitellogenin synthesis in Drosophila melanogaster. *Dev Biol* **89**: 225-36.

Chapman, T., Arnqvist, G., Bangham, J. & Rowe, L. 2003a. Sexual conflict. *Trends Ecol Evol* **18**: 41-7.

Chapman, T., Bangham, J., Vinti, G., Seifried, B., Lung, O., Wolfner, M. F., Smith, H. K. & Partridge, L. 2003b. The sex peptide of Drosophila melanogaster: female post-mating responses analyzed by using RNA interference. *Proc Natl Acad Sci U S A* **100**: 9923-8.

Chapman, T., Liddle, L. F., Kalb, J. M., Wolfner, M. F. & Partridge, L. 1995. Cost of mating in Drosophila melanogaster females is mediated by male accessory gland products. *Nature* **373**: 241-4.

Chapman, T., Neubaum, D. M., Wolfner, M. F. & Partridge, L. 2000. The role of male accessory gland protein Acp36DE in sperm competition in Drosophila melanogaster. *Proc Biol Sci* **267**: 1097-105.

Civetta, A. 1999. Direct visualization of sperm competition and sperm storage in *Drosophila*. *Current Biology* **9**: 841-44.

Civetta, A. & Clark, A. G. 2000. Correlated effects of sperm competition and postmating female mortality. *Proc Natl Acad Sci U S A* **97**: 13162-5.

Civetta, A. & Singh, R. S. 1995. High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in Drosophila melanogaster and Drosophila virilis group species. *J Mol Evol* **41**: 1085-95.

Clark, A. G., Aguade, M., Prout, T., Harshman, L. G. & Langley, C. H. 1995. Variation in sperm displacement and its association with accessory gland protein loci in Drosophila melanogaster. *Genetics* **139**: 189-201.

Clark, A. G. & Begun, D. J. 1998. Female genotypes affect sperm displacement in Drosophila. *Genetics* **149**: 1487-93.

Clark, A. G., Begun, D. J. & Prout, T. 1999. Female x male interactions in Drosophila sperm competition *Science* **283**: 217-220.

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

Coleman, S., Drahn, B., Peterson, G., Stolorv, J. & Kraus, K. 1995. A *Drosophila* male accessory gland proteins that is a member of the serpin superfamily of proteinase inhibitors is transferred to females during mating. *Insect Biochem Mol Biol* **25**: 203-07.

Collins, A. M., Williams, V. & Evans, J. D. 2004a. Sperm storage and antioxidative enzyme expression in the honey bee, Apis mellifera. *Insect Mol Biol* **13**: 141-6.

Collins, A. M., Williams, V. & Evans, J. D. 2004b. Sperm storage and antioxidative enzyme expression in the honey bee, Apis mellifera. *Insect Molecular Biology* **13**: 141-146.

Collins, A. M., Williams, V. & Evans, J. D. 2004c. Sperm storage and antioxidative enzyme expression in the honey bee, Apis mellifera. *Insect Mol Biol* **13**: 141-146.

Coyne, J. A. & Orr, H. A. 2004. *Speciation*. Sinauer Associates, Sunderland, Massachusetts.

Czesak, M. E. & Fox, C. W. 2003. Genetic variation in male effects on female reproduction and the genetic covariance between the sexes. *Evolution Int J Org Evolution* **57**: 1359-66.

Darwin, C. 1871. *The Descent of Man and Selection in Relation to Sex*. London.

Davies, S. J. & Chapman, T. 2006. Identification of genes expressed in the accessory glands of male Mediterranean Fruit Flies (Ceratitis capitata). *Insect Biochem Mol Biol* **36**: 846-56.

DeVries, J. K. 1964. Insemination and sperm storage in *Drosophila melanogaster*. *Evolution* **18**: 271-82.

Diatchenko, L., Lau, Y., Campbell, A., Chenchik, A., Mogadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. & Siebert, P. 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci USA* **93**: 6025-30.

Diatchenko, L., Lukyanov, K., Lau, Y. & Siebert, P. 1999. Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes. *Method Enzym* **303**: 349-80.

DiBenedetto, A. J., Lakich, D. M., Kruger, W. D., Belote, J. M., Baker, B. S. & Wolfner, M. F. 1987. Sequences expressed sex-specifically in Drosophila melanogaster adults. *Dev Biol* **119**: 242-51.

Eberhard, W. G. 1996. *Female control: sexual selection by cryptic female choice*. Princeton University Press, New Jersey.

Edgar, R. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-97.

Emanuelsson, O., Ren-Brunak, S., vonHeijne, G. & Nielsen, H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**: 953-71.

Fay, J. C. & Wu, C. I. 2000. Hitchhiking under positive Darwinian selection. *Genetics* **155**: 1405-13.

Ferrandon, D., Jung, A., Criqui, M., Lemaitre, B., Uttenweiler-Joseph, S., Michaut, L., Reichhart, J. & Hoffman, J. 1998. A drosomycin-GFP reporter transgene reveals a local immune response in Drosophila that is not dependent on the Toll pathway. *EMBO* **17**: 1217-27.

Filosi, M. & Perotti, M. 1975. Fine structure of spermatheca of Drosophila melanogaster Meig. *J. Submicr. Cytol* **7**: 259-70.

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**: 243-57.

Fiumera, A. C., Dumont, B. L. & Clark, A. G. 2007. Associations between sperm competition and natural variation in male reproductive genes on the third chromosome of Drosophila melanogaster. *Genetics* **176**: 1245-60.

Fowler, G. 1973. Some aspects of the reproductive biology of *Drosophila*: sperm transfer, sperm storage and sperm utilization. *Adv in Genet* **17**: 293-360.

Fu, Y. X. & Li, W. H. 1993. Statistical tests of neutrality of mutations. *Genetics* **133**: 693-709.

Galindo, B., Vacquier, V. & Swanson, W. 2003a. Positive selection in the egg receptor for abalone sperm lysine. *Proc Natl Acad Sci USA* **100**: 4639-43.

Galindo, B. E., Vacquier, V. D. & Swanson, W. J. 2003b. Positive selection in the egg receptor for abalone sperm lysin. *Proc Natl Acad Sci U S A* **100**: 4639-43.

Gavrilets, S. 2000. Rapid evolution of reproductive barriers driven by sexual conflict. *Nature* **403**: 886-9.

Gilbert, D. G. 1981. Ejaculate esterase 6 and initial sperm use by female *Drosophila melanogaster*. *J. Insect Physiol.* **27**: 641 - 650.

Gilbert, D. G., Richmond, R. C. & Sheehan, K. B. 1981. Studies of esterase 6 in Drosophila melanogaster. VII. Remating times of females inseminated by males having active or null alleles. *Behav Genet* **11**: 195-208.

Gilchrist, A. S. & Partridge, L. 2000. Why it is difficult to model sperm displacement in Drosophila melanogaster: the relation between sperm transfer and copulation duration. *Evolution Int J Org Evolution* **54**: 534-42.

Goldman, N. & Yang, Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* **11**: 725-36.

Gromko, M. H., Newport, M. E. A. & Kortier, M. G. 1984. Sperm dependence of female receptivity to remating in *Drosophila melanogaster*. *Evolution* **38**: 1273-1282.

Haerty, W., Jagadeeshan, S., Kulathinal, R. J., Wong, A., Ravi Ram, K., Sirot, L. K., Levesque, L., Artieri, C. G., Wolfner, M. F., Civetta, A. & Singh, R. S. 2007.

Evolution in the fast lane: rapidly evolving sex-related genes in Drosophila. *Genetics* **177**: 1321-35.

Haley, S. & Wessel, G. 2004. Proteolytic cleavage of the cell surface protein p160 is required for detachment of the fertilization envelope in the sea urchin. *Dev. Biol* **272**: 191-202.

Harshman, L. G. & Clark, A. G. 1998. Inference of sperm competition from broods of field-caught Drosophila. *Evolution* **52**: 1334-41.

Harshman, L. G. & Prout, T. 1994. Sperm displacement without sperm transfer in *Drosophila melanogaster*. *Evolution* **48**: 758-766.

Haygood, R., Fedrigo, O., Hanson, B., Yokoyama, K. D. & Wray, G. A. 2007. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. *Nat Genet* **39**: 1140-4.

Heifetz, Y., Lung, O., Frongillo, E. J. & Wolfner, M. F. 2000. The Drosophila seminal fluid protein Acp26Aa stimulates release of oocytes by the ovary. *Current Biology* **10**: 99-103.

Heifetz, Y., Tram, U. & Wolfner, M. F. 2001. Male contributions to egg production: the role of accessory gland products and sperm in Drosophila melanogaster. *Proc Biol Sci* **268**: 175-80.

Holland, B. & Rice, W. R. 1999. Experimental removal of sexual selection reverses intersexual antagonistic coevolution and removes a reproductive load. *Proc Natl Acad Sci U S A* **96**: 5083-8.

Hosken, D. J. 1999. Sperm displacement in yellow dung flies: a role for females. *Trends Ecol Evol* **14**: 251-252.

Hosken, D. J. & Ward, P. I. 2000. Copula in yellow dung flies (Scathophaga stercoraria): investigating sperm competition models by histological observation. *J Insect Physiol* **46**: 1355-1363.

Howard, D. J. 1999. Conspecific sperm and pollen precedence and speciation. *Ann. Rev. Ecol. Syst.* **30**: 109-132.

Howard, D. J., Palumbi, S. R., Birge, L. & Manier, M. K. (in prep) Sperm and speciation. In: *Sperm Biology: an Evolutionary Approach*, (Birkhead, T. R. & Hosken, D. J., eds.). pp. Elsevier Press.

Huang, X. & Madan, A. 1999. CAP3: A DNA sequence assembly program. *Genome Res* **9**: 868-77.

Hudson, R. R., Kreitman, M. & Aguade, M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153-9.

Kalb, J. M., DiBenedetto, A. J. & Wolfner, M. F. 1993. Probing the function of Drosophila melanogaster accessory glands by directed cell ablation. *Proc Natl Acad Sci U S A* **90**: 8093-7.

Käll, L., Krogh, A. & Sonnhammer, E. L. L. 2004. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol* **338**: 1027-36.

Kamei, N. & Glabe, C. 2003. The species-specific egg receptor for sea urchin sperm adhesion is EBR1, a novel ADAMTS protein. *Genes and Devel.* **17**: 2501-07.

Kamei, N., Swanson, W. & Glabe, C. 2000. A rapidly diverging EGF protein regulates species-specific signal transduction in early sea urchin development. *Dev. Biol* **225**: 267-76.

Kaplan, W. D., Tinderhold, V. E. & Gugler, D. H. 1962. The number of sperm present in the reproductive tracts of *D. melanogaster* females. *Dros. Inf. Serv.* **36**: 82.

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 Genet* **3**: e148.

Kern, A. D., Jones, C. D. & Begun, D. J. 2004. Molecular population genetics of male accessory gland proteins in the Drosophila simulans complex. *Genetics* **167**: 725-35.

Kim, C. H., Kim, S. J., Kan, H., Kwon, H. M., Roh, K. B., Jiang, R., Yang, Y., Park, J. W., Lee, H. H., Ha, N. C., Kang, H. J., Nonaka, M., Soderhall, K. & Lee, B. L. 2008. A three-step proteolytic cascade mediates the activation of the peptidoglycan-induced toll pathway in an insect. *J Biol Chem*.

Kubli, E. 1996. The *Drosophila* sex-peptide: A peptide pheromone involved in reproduction. *Adv. in Dev. Biol. Chem* **4**: 99-128.

Lawniczak, M. K. & Begun, D. J. 2004. A genome-wide analysis of courting and mating responses in Drosophila melanogaster females. *Genome* **47**: 900-10.

Lawniczak, M. K. & Begun, D. J. 2005. A QTL analysis of female variation contributing to refractoriness and sperm competition in Drosophila melanogaster. *Genet Res* **86**: 107-14.

Lawniczak, M. K. & Begun, D. J. 2007. Molecular population genetics of female-expressed mating-induced serine proteases in Drosophila melanogaster. *Mol Biol Evol* **24**: 1944-51.

Lefevre, G. & Jonsson, U. B. 1962. Sperm transfer, storage, displacement, and utilization in *Drosophila melanogaster*. *Genetics* **47**: 1719-1736.

Lessells, C. M. 2006. The evolutionary outcome of sexual conflict. *Philos Trans R Soc Lond B Biol Sci* **361**: 301-17.

Liu, H. & Kubli, E. 2003. Sex-peptide is the molecular basis of the sperm effect in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **100**: 9929-33.

Lung, O., Tram, U., Finnerty, C. M., Eipper-Mains, M. A., Kalb, J. M. & Wolfner, M. F. 2002a. The Drosophila melanogaster seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression. *Genetics* **160**: 211-24.

Lung, O., Tram, U., Finnerty, M., Eipper-Mains, M., Kalb, J. & Wolfner, M. 2002b. The Drosophila melanogaster seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression. *Genetics* **160**: 211-14.

Mack, P., Kapelnikov, A., Heifetz, Y. & Bender, M. 2006a. Mating-responsive genes in reproductive tissues of female *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **103**: 10358-63.

Mack, P. D., Kapelnikov, A., Heifetz, Y. & Bender, M. 2006b. Mating-responsive genes in reproductive tissues of female Drosophila melanogaster. *Proc Natl Acad Sci U S A* **103**: 10358-63.

Malm, J., Hellman, J., Hog, P. & Lilja, H. 2000. Enzymatic action of prostate-specific antigen (PSA or hK3): substrate specificity and regulation by Zn (2+). *Prostate* **45**: 132-39.

Marchler-Bauer, A. & Bryant, S. 2004. CD-Search:protein domain annotations on the fly. *Nucleic Acids Research* **32**: 327-31.

McDonald, J. H. & Kreitman, M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351**: 652-4.

McGraw, L. A., Gibson, G., Clark, A. G. & Wolfner, M. F. 2004. Genes regulated by mating, sperm, or seminal proteins in mated female Drosophila melanogaster. *Curr Biol* **14**: 1509-14.

Metz, E. & Palumbi, S. 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Mol Biol Evol* **13**: 397-406.

Miller, G. & Pitnick, S. 2002. Sperm-female co-evolution in Drosophila. *Science* **298**: 1230-33.

Miller, G. & Pitnick, S. 2003. Functional significance of seminal receptacle length in Drosophila melanogaster. *J. Evol Biol* **16**: 114-16.

Monsma, S. A., Harada, H. A. & Wolfner, M. F. 1990. Synthesis of two Drosophila male accessory gland proteins and their fate after transfer to the female during mating. *Dev Biol* **142**: 465-75.

Monsma, S. A. & Wolfner, M. F. 1988. Structure and expression of a Drosophila male accessory gland gene whose product resembles a peptide pheromone precursor. *Genes Dev* **2**: 1063-73.

Moriyama, E. N. & Powell, J. R. 1997. Synonymous substitution rates in Drosophila: mitochondrial versus nuclear genes. *J Mol Evol* **45**: 378-91.

Moshitzky, P., Fleischmann, I., Chaimov, N., Saudan, P., Klauser, S., Kubli, E. & Applebaum, S. W. 1996. Sex-peptide activates juvenile hormone biosynthesis in

the Drosophila melanogaster corpus allatum. *Arch Insect Biochem Physiol* **32**: 363-74.

Mueller, J., Ravi-Ram, K., McGraw, M., Bloch-Qazi, M., Siggia, E., Clark, A., Aquadro, C. & Wolfner, M. 2005a. Cross-species comparison of *Drosophila* male accessory gland protein genes. *Genetics* **171**: 131-43.

Mueller, J. L., Ravi Ram, K., McGraw, L. A., Bloch Qazi, M. C., Siggia, E. D., Clark, A. G., Aquadro, C. F. & Wolfner, M. F. 2005b. Cross-species comparison of Drosophila male accessory gland protein genes. *Genetics* **171**: 131-43.

Neubaum, D. M. & Wolfner, M. F. 1999a. Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. *Genetics* **153**: 857-869.

Neubaum, D. M. & Wolfner, M. F. 1999b. Mated Drosophila melanogaster females require a seminal fluid protein, Acp36DE, to store sperm efficiently. *Genetics* **153**: 845-57.

Neubaum, D. M. & Wolfner, M. F. 1999c. Wise, winsome, or weird? Mechanisms of sperm storage in female animals. *Curr Top Dev Biol* **41**: 67-97.

Newport, K. M. & Gromko, M. H. 1984. The effect of experimental design on female receptivity to remating and its impact on reproductive success in *Drosophila melanogaster*. *Evolution* **38**: 1261-1272.

Otto, S. P. 2000. Detecting the form of selection from DNA sequence data. *Trends Genet* **16**: 526-529.

Overduin, M. & de Beer, T. 2000. The plot thickens: how thrombin modulates blood clotting. *Nat Struct Biol* **7**: 267-9.

Palumbi, S. 1999. All males are not created equal: Fertility difference depend on gamete recognition polymorophisms in sea urchins. *Proc Natl Acad Sci U S A* **99**: 12632-37.

Panhuis, T. M. & Swanson, W. J. 2006. Molecular evolution and population genetic analysis of candidate female reproductive genes in Drosophila. *Genetics* **173**: 2039-47.

Parker, G. A. 1970. Sperm competition and its evolutionary consequences in the insects. *Biol. Rev.* **45**: 525-67.

Parker, G. A. (1979) Sexual selection and reproductive competition in insects. (Blum, M. S. & Blum, N. A., eds.). pp. 123-166. Academic Press, New York.

Pawitan, Y., Michiels, S., Koscielny, S., Gusnanto, A. & Ploner, A. 2005. False discovery rate, sensitivity and sample size for microarray studies. *Bioinformatics* **21**: 3017-24.

Peng, J., Chen, S., Busser, S., Liu, H., Honegger, T. & Kubli, E. 2005a. Gradual release of sperm bound sex-peptide controls female postmating behavior in *Drosophila*. *Current Biology* **15**: 207-13.

Peng, J., Chen, S., Busser, S., Liu, H., Honegger, T. & Kubli, E. 2005b. Gradual release of sperm bound sex-peptide controls female postmating behavior in Drosophila. *Curr Biol* **15**: 207-13.

Pitnick, S., Markow, T. A. & Spicer, G. S. 1999. Evolution of multiple kinds of female sperm-storage organs in *Drosophila*. *Evolution* **53**: 1804-22.

Pitnick, S., Wolfner, M. & Suarez, S. (in prep.) Sperm-female interactions. In: *Sperm Biology: An Evolutionary Perspective*, (Birkhead, T., Hosken, D. & Pitnick, S., eds.). pp. Elsevier Press.

Potempa, J., Korzus, E. & Travis, J. 1994. The serpin superfamily of proteinase inhibitors: structure, function, and regulation. *J Biol Chem* **269**: 15957-60.

Price, C. S. 1997. Conspecific sperm precedence in Drosophila. *Nature* **388**: 663-6.

Price, C. S., Kim, C. H., Gronlund, C. J. & Coyne, J. A. 2001. Cryptic reproductive isolation in the Drosophila simulans species complex. *Evolution Int J Org Evolution* **55**: 81-92.

Price, C. S. C., Dyer, K. A. & Coyne, J. a. 1999. Sperm competition between Drosophila males involves both displacement and incapacition. *Nature* **400**: 449-52.

Prokupek, A. P., Hoffman, F., Eyun, S., Moriyama, E. N. & Harshman, L. G. An evolutionary expressed sequence tag analysis of *Drosophila* spermatheca genes. *In Review*.

Prout, T. & Bundgaard, J. 1977. The population genetics of sperm displacement. *Genetics* **85**: 95-124.

Ravi-Ram, K. & Wolfner, M. 2005. Fates and targets of male accessory gland proteins in mated female Drosophila melanogaster. *Insect Biochem Mol Biol* **35**: 1059-71.

Ravi-Ram, K. & Wolfner, M. F. 2007. Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduction. *Int. Comp. Biol.*: 19.

Rice, W. R. 1996. Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature* **381**: 232-4.

Rice, W. R. 2000. Dangerous liaisons. *Proc Natl Acad Sci U S A* **97**: 12953-5.

Rice, W. R. & Holland, B. 1997. The enemies within: intergenomic conflict, interlocus contest evolution (ICE), and the intraspecific Red Queen. *Behav. Ecol. Sociobiol.* **41**: 1-10.

Ross, J., Jiang, H., Kanost, M. & Wang, Y. 2003. Serine proteases and their homologs in *Drosophila melanogaster* genome: an initital analysis of sequence conservation and phylogenetic relationships. *Gene* **304**: 117-131.

Rowe, L. & Arnqvist, G. 2002. Sexually antagonistic coevolution in a mating system: combining experimental and comparative approaches to address evolutionary processes. *Evolution Int J Org Evolution* **56**: 754-67.

Rowe, L., Arnqvist, G., Sih, A. & Krupa, J. J. 1994. Sexual conflict and the evolutionary ecology of mating patterns: water striders as a model system. *TREE* **9**.

Rowe, L. & Day, T. 2006. Detecting sexual conflict and sexually antagonistic coevolution. *Philos Trans R Soc Lond B Biol Sci* **361**: 277-85.

Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. 2003. DnaSP, DNA polymorphism analysis by the coalescent and other methods. *Bioinformatics* **19**: 2496-2497.

Schlager, G. 1960. Sperm precedence in the fertilization of eggs in *Tribolium castaneum*. *Ann. Entomol. Soc. Am.* **53**: 557-560.

Scott, D. & Richmond, R. C. 1990. Sperm loss by remating in *Drosophila melanogaster* females. *J. Insect Physiology* **36**: 451-456.

Scott, D. & Williams, E. 1993. Sperm displacement after remating in *Drosophila melanogaster*. *J. Insect Physiology* **39**: 201-206.

Seehuus, S. C., Norberg, K., Gimsa, U., Krekling, T. & Amdam, G. V. 2006. Reproductive protein protects functionally sterile honey bee workers from oxidative stress. *Proc Natl Acad Sci U S A* **103**: 962-7.

Simmons, L. W. & Kotiaho, J. S. 2002. Evolution of ejaculates: patterns of phenotypic and genotypic variation and condition dependence in sperm competition traits. *Evolution Int J Org Evolution* **56**: 1622-31.

Snook, R. R. & Hosken, D. J. 2004. Sperm death and dumping in Drosophila. *Nature* **428**: 939-41.

Song, J., Wong, J. & Wessel, G. 2006. Oogenesis: Single cell development and differentiation. *Dev. Biol* **300**: 385-405.

Swanson, W., Nielsen, R. & Yang, Z. 2003b. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.* **20**: 18-20.

Swanson, W. & Vacquier, V. 1998. Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. *Science* **281**: 710-12.

Swanson, W. & Vacquier, V. 2002a. The rapid evolution of reproductive proteins. *Nat Rev Genet* **3**: 137-144.

Swanson, W., Wong, A., Wolfner, M. & Aquadro, C. 2004a. Evolutionary expressed sequence tag analysis of Drosophila female reproductive tracts identifies several genes subjected to positive selection. *Genetics* **168**: 1457-65.

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

Swanson, W. J., Clark, A. G., Waldrip-Dail, H., Wolfner, M. F. & Aquadro, C. F. 2001a-b. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc. Natl. Acad. Sci. U S A* **98**: 7375-7379.

Swanson, W. J., Clark, A. G., Waldrip-Dail, H. M., Wolfner, M. F. & Aquadro, C. F. 2001a. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. *Proc Natl Acad Sci U S A* **98**: 7375-9.

Swanson, W. J., Nielsen, R. & Yang, Q. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.* **20**: 18-20.

Swanson, W. J. & Vacquier, V. D. 2002. The rapid evolution of reproductive proteins. *Nat Rev Genet* **3**: 137-44.

Swanson, W. J., Wong, A., Wolfner, M. F. & Aquadro, C. F. 2004b. Evolutionary expressed sequence tag analysis of Drosophila female reproductive tracts identifies genes subjected to positive selection. *Genetics* **168**: 1457-65.

Swanson, W. J., Yang, Z., Wolfner, M. F. & Aquadro, C. F. 2001b. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc Natl Acad Sci U S A* **98**: 2509-14.

Swanson, W. J., Yang, Z., Wolfner, M. F. & Aquadro, C. F. 2001b. Positive selection droves the evolution of several female reproductive proteins in mammals. *Proc. Natl. Acad. Sci. U S A* **98**: 2509-5214.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585-95.

Takahata, N. & Clark, A. G. 1993. *Mechanisms of molecular evolution : introduction to molecular paleopopulation biology*. Japan Scientific Societies Press; Sinauer Associates, Tokyo

Thornhill, R. 1983. Cryptic female choice and its implications in the scorpionfly *Harpobittacus nigriceps*. *Am Nat* **122**: 1548-1556.

Tram, U. & Wolfner, M. F. 1999. Male seminal fluid proteins are essential for sperm storage in Drosophila melanogaster. *Genetics* **153**: 837-44.

Trivers, R. L. (1972) Parental investment and sexual selection. In: *Sexual selection and the decent of man*, (Cambell, B., ed.). pp. 1871-1971. Aldine-Atherton, Chicago.

Tusnady, G. E. & Simon, I. 2001. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**: 849-50.

Waage, J. K. 1986. Evidence for widespread sperm displacement ability among Zygoptera (Ordonata) and the means for predicting its presence. *Biol. J. Linn. Soc.* **28**: 285-300.

Wagstaff, B. & Begun, D. J. 2004a. Comparative genomics of acessory gland protein genes in *Drosophila melanogaster* and *D. pseudoobscura*. *Mol. Biol. Evol.* **22**: 818-32.

Wagstaff, B. & Begun, D. J. 2004b. Comparative genomics of acessory gland protein genes in *Drosophila melanogaster* and *D. pseudoobscura*. *Mol Biol Evol* **22**: 818-32.

Wayne, M. L. & Simonsen, K. L. 1998. Statistical tests of neutrality in the age of weak selection. *TREE* **13**: 236-240.

Weirich, G. G., Collins, A. M. & Williams, V. P. 2002. Antioxidant enzymes in the honey bee, *Apis mellifera*. *Apidologie* **33**: 3-14.

Wheeler, D. E. & Krutzsch, P. H. 1994. Ultrastructure of the spermathecae and its associated gland in the and *Crematogaster opuntiae* (Hymenoptera: Formicidae). *Zoomorphology* **114**: 203-214.

Wolfner, M. F. 1997. Tokens of love: functions and regulation of Drosophila male accessory gland products. *Insect Biochem Mol Biol* **27**: 179-92.

Wolfner, M. F. 2002. The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in *Drosophila Heredity* **88**: 85-93.

Wolfner, M. F., Harada, H. A., Bertram, M. J., Stelick, T. J., Kraus, K. W., Kalb, J. M., Lung, Y. O., Neubaum, D. M., Park, M. & Tram, U. 1997. New genes for male accessory gland proteins in Drosophila melanogaster. *Insect Biochem Mol Biol* **27**: 825-34.

Wong, A., Turchin, M. C., Wolfner, M. F. & Aquadro, C. F. 2007. Evidence for Positive Selection on Drosophila melanogaster Seminal Fluid Protease Homologs. *Mol Biol Evol*.

Wong, W. S. & Nielsen, R. 2004. Detecting selection in noncoding regions of nucleotide sequences. *Genetics* **167**: 949-58.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**: 555-6.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**: 1586-91.

Yang, Z. & Bielawski, J. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**: 431-49.

Yang, Z. & Nielsen, R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *J.Mol.Evol.* **46**: 409-18.

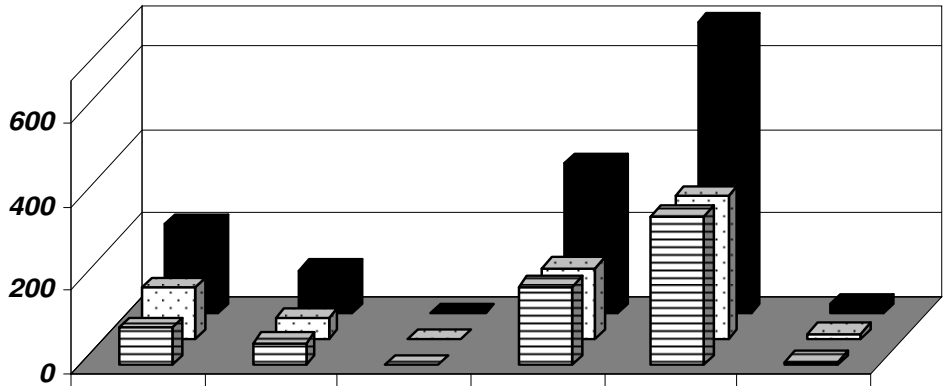
Yang, Z. & Nielsen, R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**: 908-17.

Yang, Z. & Swanson, W. J. 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol Biol Evol* **19**: 49-57.

Yang, Z., Swanson, W. J. & Vacquier, V. D. 2000. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Mol Biol Evol* **17**: 1446-1454.

Zhu, Y., Machleder, E., Chenchik, A. & Siebert, P. 2001. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* **30**: 892-97.

Figure 4.1: Differential regulation of genes expressed in the spermathecae and seminal receptacle at two time points post-mating. Three comparisons are made in this graph for each organ 3h – V, 6h – V and 6h – 3hr.



	Spermathecae			Seminal Receptacle		
	3h -V	6h -V	6h - 3h	3h -V	6h -V	6h - 3h
up	91	54	0	188	359	8
down	124	52	1	172	345	14
Total	215	106	1	360	704	22

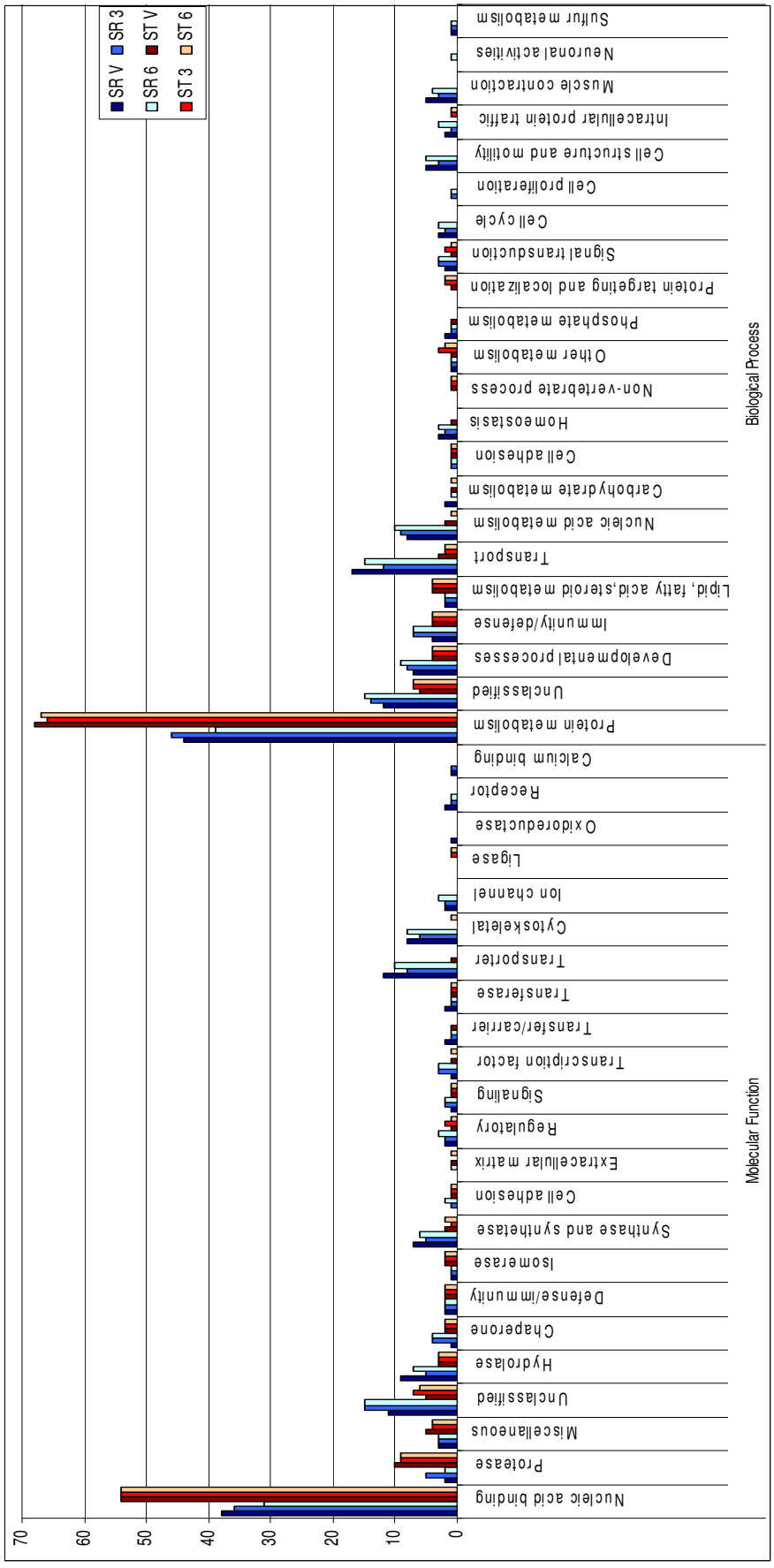
3h - V: difference in expression between in the 3 hour time point compared to virgin. 6h – V: difference in expression between the 6 hour time point compared to virgin. 6h – 3h: difference in expression between the 6 hour time point and 3 hour time point. up: number genes with higher expression in the first comparative category; down: number of genes with lower expression in the first comparative category. For example in 3h-V :up refers to the genes that have higher expression at 3 hours compared to virgin; down refers to genes that have lower expression at 3 hours compared to virgin; total refers to the total number of genes differentially expressed between 3 hours and virgin.

Table 4.1. Rank expression of serine proteases and yolk proteins expressed in the spermathecae and seminal receptacle

	Gene symbol	ST V	ST 3	ST 6	ST 3*	SR V	SR 3	SR 6
Proteases	CG17239	12	1	1	1	553	53	1838
	CG32834	20	26	12	6	2300	1833	6483
	CG31681	6	7	5	7	354	100	1433
	CG32277	27	35	34	11	2027	1026	5507
	CG17012	5	6	9	12	448	161	1930
	CG18125	137	5	9	18	1204	104	2889
	CG9897	1	14	11	19	1552	1095	5793
	CG17234	73	3	3	25	747	28	1143
	CG30371	36	10	16		1144	287	2507
	CG10469	89	138	93		104	297	214
	CG13318	508	590	459		50	15	18
Yolk Proteins	YP1	7	11	7		92	40	44
	YP2	28	41	14		256	119	121
	YP3	34	67	24		286	172	179

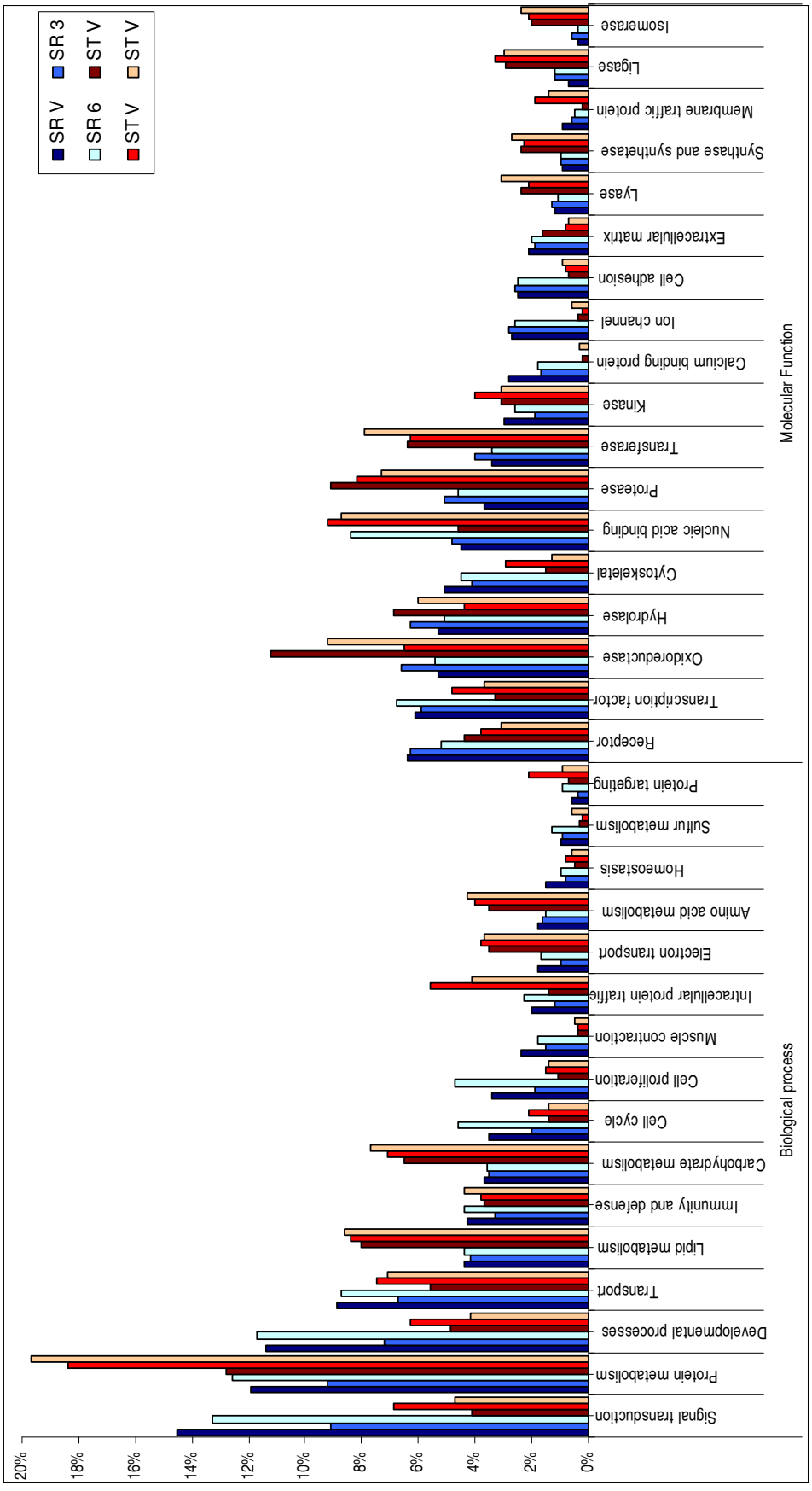
Proteases and yolk proteins ranked by average hybridization intensity. Genes are listed which are in the top 100 intensities in at least one experimental category (note that the rank number can be much lower than 100 in other categories). Numbers indicate their ranking by hybridization intensity. ST = spermatheca, SR = seminal receptacle. *The ranking of the expression level 3 days post-mating as reported in Allen and Spradling, 2008.

Supplementary Figure 4.1: Classification of the top 100 genes based on average hybridization intensity in the ST and SR. Classification of genes by biological process and molecular function was independently performed.



Prokupek 2008

Supplementary Figure 4.2: PANTHER classification of genes upregulated in the SR and ST. Upregulation refers to a higher expression in one SSO organ when compared to the other SSO at each time point. Data is represented as the percentage of genes in each category. Classification of genes by biological process and molecular function was independently performed.


SR V
SR 3
SR 6
ST V
ST V
ST V
20%
18%
16%
14%
12%
10%
8%
6%
4%
2%
0%
Biological process
Signal transduction
Protein metabolism
Developmental processes
Transport
Lipid metabolism
Immunity and defense
Carbohydrate metabolism
Cell cycle
Cell proliferation
Muscle contraction
Intracellular protein traffic
Electron transport
Amino acid metabolism
Homeostasis
Sulfur metabolism
Protein targeting
Molecular Function
Receptor
Transcription factor
Oxidoreductase
Hydrolase
Cytoskeletal
Nucleic acid binding
Protease
Transferase
Kinase
Calcium binding protein
Ion channel
Cell adhesion
Extracellular matrix
Lyase
Synthase and synthetase
Membrane traffic protein
Ligase
Isomerase

Supplementary Table 4.1: Average expression (hybridization intensity) of genes in the spermathecae (transcript level > 100). Genes are classified by DAVID into level 4 biological process and level 4 molecular function gene ontology categories.

Supplementary Table 4.2: Average expression (hybridization intensity) of genes in the seminal receptacle (transcript level > 100). Genes are classified by DAVID into level 4 biological process and level 4 molecular function gene ontology categories.

Supplementary Table 4.3: Differential expression of spermathecae genes at two time points (3 and 6 hours) post-mating.

Supplementary Table 4.4: Differential expression of seminal receptacle genes at two time points (3 and 6 hours) post-mating.

Supplementary Table 4.5: Genes differentially expressed in the spermathecae compared to the seminal receptacle in virgin, 3h and 6h post-mating samples. In this comparison, upregulated (up) refers to genes which are upregulated in the spermathecae in comparison to the seminal receptacle. Downregulated (down) refers to genes which are downregulated (down) in the spermathecae compared to the seminal receptacle.

Supplementary Table 4.6: DAVID analysis of over-represented gene categories of differentially expressed genes in the spermathecae at two time points (3 and 6 hours) post-mating.

Supplementary Table 4.7: DAVID analysis of over-represented gene categories of differentially expressed genes in the seminal receptacle at two time points (3 and 6 hours) post-mating.

Supplementary Table 4.8: DAVID analysis of over-represented categories of differentially expressed genes from a comparison of the spermathecae and seminal receptacle samples at virgin, 3 and 6 hour post-mating time points.