Identification of genes and gene pathways affecting fertility in male *Drosophila*

By Lisa Levesque

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Biochemistry and Medical Genetics

University of Manitoba

Winnipeg

Abstract

Drosophila females are known to remate in wild and laboratory populations generating an opportunity for sperm competition. Normally the second male to mate sires the majority of progeny; however, conspecific sperm precedence is the phenomena whereby the male of the same species as the female fathers the majority of the progeny regardless of mating order. I surveyed *D. simulans* laboratory strains carrying different mapped *D. mauritiana* P-element insertions (IG lines) for their ability to sire progeny when second to mate. I found significant variation in the proportion of progeny sired by IG lines, with some lines showing sperm competitive breakdown (*P2*< 0.5). Taking advantage of the fact that the *D. mauritiana* introgressions have been previously mapped, I have identified two loci that account for conspecific sperm precedence between *D. simulans* and *D. mauritiana*. Using genome resources, I identified 81 candidate genes and narrowed down the list on the basis of differences in male reproductive tract gene expression to five (*P*< 0.05) or eight (*P*<0.1) genes. I found a larger concentration of differentially regulated genes within the 89B position. Using coding sequence data I identified 10 genes as candidate conspecific male precedence genes. Genes in the 89B region come to light as candidates for future functional studies of conspecific male precedence.

Acknowledgements

I would like to thank Dr. Civetta, your guidance and patience was vital to this accomplishment. You are a wonderful teacher and I appreciate all the effort and time you invested into my education.

I wish to acknowledge the members of my committee Dr. Davie and Dr. Whyard for their helpful comments and time.

I wish to thank my family and friends for their constant encouragement, understanding and unconditional love. No girl is luckier than I am. I couldn't have done this without you.

I would like to acknowledge Desiree, a wonderful friend and excellent lab mate. I knew I could always count on you no matter what my problems were. Your support throughout this process was unending and I appreciate it.

I want to acknowledge Barb Brouwers and Vignesh Sundararajan for all of their wonderful help in the lab.

Caroline, it was nice to share ride with someone!

To the members of the Biochemistry and Medical Genetics Department, thank you.

I would like to thank the Manitoba Health Research Council for providing me with partial funding.

Dedication

To my loved ones, those here and passed on: you inspire me to be a better person and motivate me to achieve my dreams.

Table of Contents

List of Abbreviations

ACP	Accessory gland proteins
ANOVA	Analysis of Variance
BLAST	Basic Local Alignment Search Tool
CO_2	Carbon Dioxide
CSP	Conspecific Sperm Precedence
C_T	Threshold cycle, cycles required to reach the threshold in RT-PCR
ddH_20	distilled deionized water
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetraacetate
g	Gravity
HMS	Hybrid male sterility
IG	*D. simulans* with *D. mauritiana* DNA introgression on the 3rd chromosome
$Mgcl_2$	Magnesium chloride
ml	Millilitre
mM	Millimolar
MSL	Male-Spécifique Lethal Complex
ng	Nanogram
P2	Proportion of progeny fathered by the second male to mate
PBS	Phosphate buffered solution
RNA	Ribonucleic acid
RPM	Rotations per minute
RT	Room Temperature
RT-PCR	Reverse transcriptase polymerase chain reaction
SDDW	Sterile Double Distilled Water

TBE	Tris-Borate-EDTA
TRIS	(hydroxymethyl) aminomethane
µl	Microlitre
V/V	Volume to volume

List of Tables

List of Figures

Introduction

1.1 Male infertility

The complexity of the phenomenon of human male infertility is reflected in the fact that compared to female infertility only a small proportion of cases can be successfully treated. Approximately 15% of couples attempting their first pregnancy meet with failure to conceive in India (Poongothai *et al*., 2009) and it is estimated this number is 2-7% worldwide (Chandley, 1998). A standard defined by the World Health Organization (WHO) states that the failure to conceive a pregnancy after two years is defined as primary infertility. A review of research conducted over the past twenty years in India has revealed that male associated factors are at least partly responsible in about 50% of infertile couples (Poongothai *et al*., 2009). It is understood that the genetic regulation of fertility involves inter-related processes including testicular development, spermatogenesis (germ cell mitosis, meiosis and spermatid maturation), and endocrine and paracrine regulation of these events (McLachlan *et al*., 1998). The known causes of male infertility in humans are highly heterogeneous and these include both mechanical differences and endocrine irregularities. Thus, it is recognized that male infertility is multifactorial and encompasses a wide variety of disorders. Diagnosis of male infertility in humans relies heavily on semen analysis. Seminograms of infertile men can reveal a number of sperm abnormalities such as azoospermia, oligospermia, teratospermia, asthenozoospermia, necrospermia, polyspermia and aspermia. However, more than fifty percent of male infertility is idiopathic (cause unknown). It is foreseeable that male infertility could be caused by several identifiable genetic causes without a clear or

distinguishable pathology or phenotype (i.e. normal semen samples). In fact, research to date has revealed numerous genetic causes of male infertility including XXY (Klinefelter syndrome), translocations, inversions, deletions, mutations in the androgen receptor gene, mutations in the cystic fibrosis transmembrane conductance regulator (CTFR) gene, microdeletions in the long arm of the Y chromosome (Yq), and mutations in mitochondrial genes due to the high demand of adenosine triphosphate in sperm (McLachlan *et al*., 1998; Dohle *et al*., 1999; Poongothai *et al*., 2009). Genetic investigations on male reproduction involving humans are both logistically and ethically problematic. Historically, model organisms have been invaluable to research benefiting humans. It is relevant to note that the *Drosophila* model system has contributed to the identification of several human gene homologs, including genes affecting male fertility (Malicki *et al*., 1992; Eberhart *et al*., 1996; Xu *et al*., 2003).

The broad aim of this thesis is to contribute to the current body of knowledge on male reproduction by identifying genes involved in male reproductive pathways that are responsible for differences in male sperm competitive ability. Genotype- phenotype associations for male sperm competitive ability are vague to date; however, this research may help to elucidate some genetic factors involved.

1.2 Fitness and Sexual Selection

An organism's fitness is defined as the success of a genotype as measured by its ability to survive in a given environment, productiveness in mating, and the ability to produce viable offspring (Freeman & Herron, 2001 and references there in). Natural selection is an evolutionary process whereby organisms most adapted to their environment contribute the majority of genes to the gene pool. Sexual selection is a form of selection that adequately explains the occurrence of sexual dimorphism within species and the presence of elaborate secondary sexual characteristics that occur in nature (Freeman & Herron, 2001 and references there in). Sexual selection is differential reproductive success based on the individual's ability to contend for and secure access to mates. As the personal energetic cost of producing gametes can be very high, mechanisms have evolved to avoid gamete wastage. These mechanisms help to ensure that reproductive energies are focused on matings with highest probability of reproductive success (Snook & Markow, 2002). In anisogamous species, the maternal and paternal investments are not equivalent. Aniosogamy (also known as heterogamy) refers to the reproductive strategy involving the union of two dissimilar gametes differing in either size alone and/or form. As sperm are relatively cheap for males to produce, a male can produce a high volume of gametes and therefore should be limited in reproductive capacity by his access to female's mates (Bateman, 1948; Simmons, 2001). When a male fails to mate and produce offspring the result is that none of his genes will be propagated into the gene pool; therefore, it is logical that a male's ultimate goal should be to

maximize the number of eggs he fertilizes (Singh *et al*., 2002). Females eggs are more costly to produce and maternal investment is significantly higher; therefore, we expect a female will be limited in her reproductive capacity not by the number of mates she has access to but rather by the genotype of her mates. As a result of this existing asymmetry, we expect females to be the choosy sex as her fitness is increased by maximizing the genotype of her offspring. Thus, we expect males to invest energy into making themselves appealing to females (Simmons, 2001). In essence sexual selection is comprised by the sum of the interactions that take places between the sexes known as intersexual selection but also the interactions that take place between members of the same sex, known as intrasexual selection. Intrasexual selection can manifest itself in many ways from the use of elaborate secondary sexual traits including weaponry to more subtle competition between sperm.

1.3 Postmating Prezygotic Barriers

From an evolutionary biology perspective, male reproductive success has received extensive attention owing to the fact that breakdowns in male reproductive ability can lead to the establishment of barriers to gene flow between populations and eventually cause speciation. Historically, evolutionary research has focused on premating (prezygotic) mechanisms of isolation; including geographic and behavioural barriers to mating or postmating (postzygotic) mechanisms including hybrid inviability and sterility (Mayr, 1942; Mayr, 1963; Dobzhansky, 1970; Carson, 1985; Coyne & Orr, 2004). In

alignment with the Dobzhansky-Muller model, incompatible gene interactions between diverging populations are recognized as a basis for postzygotic reproductive isolation. The Dobzhansky-Muller model assumes that there is selective pressure against combinations of alleles when they are present in the heterozygous state, as a result of deleterious interactions between the alleles. Gametic isolation barriers that affect interbreeding after mating and before the formation of zygotes have been known for years (Patterson, 1946; Stebbins, 1950); however, non-competitive and competitive gametic isolation have only recently received heightened attention and been more formally included in studies of speciation (Hewitt *et al*., 1989; Gregory & Howard, 1993; Gregory & Howard, 1994; Wade *et al*., 1994; Price, 1997; Howard, 1999; Simmons, 2001; Fricke & Arnqvist, 2004; Coyne & Orr, 2004; Ludlow & Magurran, 2006). In the absence of or working together with other reproductive barriers, postmating prezygotic sexual selection may contribute greatly to the separation of populations into closely related species (Fricke & Arnqvist, 2004).

1.4 Sperm Competition

It has become clear that, within populations, males who are successful at securing copulation are not guaranteed reproductive success. This is directly related to the fact that females of numerous species are known to multiply mate with more than one male (Birkhead & Møller, 1998). To date, the body of knowledge generated from research firmly supports the belief that polyandrous (the mating pattern where a female mates

with multiple males) females generate selective pressures ultimately leading to coevolution between males and females. The affect of these stresses are broad and are thought to influence several aspects of male x female interactions including, premating signal-response exchanges needed for species mate recognition (Liimatainen & Hoikkala, 1998; Greenspan & Ferveur, 2000; Saarikettu *et al*., 2005; Nickel & Civetta, 2009), male and female primary genitalia morphology (Eberhard, 1985; Arnqvist, 1998; Cordoba-Aguilar & Siva-Jothy 2004; Jagadeeshan & Singh, 2006), sperm and female sperm storage organ morphology (Pitnick *et al.,* 1999; Miller & Pitnick, 2002; Pitnick *et al*., 2003) and also postmating chemical cues (Wolfner, 2007; Wolfner, 2009; Wigby *et al*., 2009).

Sexual selection continues after copulation through postmating prezygotic mechanisms including sperm competition and female cryptic choice (preferential use of sperm by females). For males, securing a female and copulating does not guarantee reproductive success. In essence any characteristics that will make a male more efficient at not only securing a female and mating but also producing progeny will be favored under high selective pressure.

In *Drosophila*, females willingly remate both in laboratory and wild populations (Harshman & Clark, 1998; Snook, 1998; Jones & Clark, 2003). A female *D. melanogaster* will store around twenty five percent of the sperm she receives in a single mating and use approximately half of this amount to fertilize eggs as they are oviposited over the following 2 week period (Suarez, 2002; Fiumera *et al*., 2007). This storage of the sperm within the female reproductive tract generates an opportunity for sperm competition to

affect a male's fertility given that in multiply mating species, sperm from more than one male will be present. The final outcome of sperm competition in *Drosophila* is strongly influenced not only by complex male X male (Clark *et al*., 2000; Birkhead *et al*., 2004) and male X female interactions (Clark *et al.*, 1999; Nilsson *et al*., 2003) but also trade offs with other postmating reproductive traits (Bjork *et al*. 2007; Civetta *et al*., 2008). This alludes to the intricacy of the mechanisms underlying sperm competition outcomes and also the complicated nature of male – female coevolution within the genus. In *Drosophila* there is extensive evidence for variability among males in sperm competitive ability. Using a set of 20 chromosome substitution lines to quantify the effect of chromosome replacement in sperm competition in *D. melanogaster* Civetta and Clark (2000) found evidence that the third chromosome likely harbours genes related to differences in both sperm offence and defence. From this research it was also evident that there was not a single chromosomal effect observed for female influence on sperm competition outcomes, implying that there is a different and intricate process for female cryptic choice. Despite the complex genetic basis of a 'sperm competition' phenotype, there are well-characterized genes in *D. melanogaster* known to affect a male's ejaculate ability to compete against rival sperm. The best examples are genes producing proteins secreted by the male's accessory glands during ejaculation. Secreted accessory gland proteins (Acps) are transferred from males to females along with sperm during copulation and are known to have a variety of effects on female reproductive physiology (Swanson *et al*., 2001). Using a loss of function mutant *D. melanogaster* strain for *Acp29AB* gene it was shown that the seminal protein plays an important role in the maintenance of sperm during

storage in female organs (Wong *et al.,* 2008). This agreed with previous association studies that had shown a correlation between *Acp29AB* genotypes and sperm competitive ability (Clark *et al*., 1995). A knock out of *Acp62F* by targeted deletion has shown that this protein affects a male's ability to place sperm in storage (Mueller *et al*., 2008). While the knockout phenotype might not be necessarily representative of genetic variants segregating in a natural population, a population survey of sequence variation at *Acp62F* has also established significant associations between polymorphisms at this gene and both second male paternity success and female induced fecundity (Fiumera & Clark, 2007). Acps are also known to induce oogenesis and ovulation, decrease remating behaviour and influence egg hatchability (Swanson *et al*., 2001; Chapman *et al*., 2001). Thus, Acps play an important role in a male's overall fertility.

1.5 Conspecific Sperm Precedence

Competitive pressures imposed by multiply-mating females can spawn postmating prezygotic reproductive isolation and subsequently speciation (Gavrilets, 2000). It is important to note female remating is not exclusive to *Drosophila*; remating occurs in a wide variety of species including plants, insects, and mammals (Hewitt *et al*., 1989; Wade *et al*., 1994; Gregory & Howard, 1997; Howard, 1999). Among closely related species of *Drosophila*, evidence of intraspecies diversification is present in sperm competition and the phenomenon of conspecific sperm precedence (CSP). When a female multiply mates with two conspecific males, the majority of progeny are sired by the second male, above

60% (Gromko *et al*., 1984; Clark *et al*., 1995; Fiumera *et al*., 2005; Friberg *et al*., 2005; Fiumera *et al*., 2007). However, second-male advantage disappears in favour of CSP when a female mates with both a conspecific male and a heterospecific male regardless of mating order (Price, 1997; Price, 2000; Price, 2001; Dixon *et al,.* 2003). For clarification, a conspecific male is defined as a male of the same species, whereas a heterospecifc male is defined as a male of a different species that is able to generate progeny when mated. CSP is a known form of postmating prezygotic isolation and is defined as the preferential utilization of conspecific sperm or gametes for fertilization when females have been inseminated by both a conspecific and a heterospecific male (Howard, 1999). Mechanisms of CSP have been studied considerably in a variety of insects, vertebrates, and plants (Hewitt *et al*., 1989; Robinson *et al*., 1994; Rieseberg *et al*., 1995; Price, 1997; Howard *et al*., 1998; Howard, 1999; Dixon *et al.,* 2003; Geyer & Palumbi, 2003, 2005; Chang, 2004). The cricket genus *Allonemobius* represents one of the finest demonstrations of CSP occurring in natural populations; two closely related species are capable of mating among heterospecifics in nature but are isolated by preferential fertilization by conspecifics as well as non-competitive gametic isolation (Gregory & Howard, 1993; Howard *et al*., 1998; Howard, 1999). CSP is a competitive form of isolation that is only observable after multiple matings.

Very little is known about the genetic basis of CSP as only two studies have previously attempted to map genes responsible for this phenotype. Using *D. simulans* - *D. sechellia* introgression lines, significant quantitative trait loci (QTL) were detected on the

second and third chromosome only when using low stringency statistical thresholds (Civetta *et al*., 2002). Another QTL approach using reciprocal F2 backcross females of crosses between two crickets, *Allonemobius fasciatus* and *A. socius*, mated to males of the two species found several unlinked markers associated with either enhancing or reducing conspecific male paternity success (Britch *et al*., 2007).

1.6 Objectives and Aims

To date, only two studies have attempted to resolve the genetic basis of CSP (Civetta *et al*., 2002; Britch *et al.,* 2007). I was able to take advantage of the naturally occurring CSP phenotype among species of the *Drosophila simulans* clade to develop an approach that allowed for detection and identification of genes that contribute to variability in paternal success. The *D. simulans* clade is made up of three sibling species *D. simulans*, *D. sechellia*, and *D. mauritiana*. These three species are close relatives of each other and diverged from a common ancestor, *D. melanogaster,* approximately 0.5-1 million years ago (Kilman & Hey, 1993) with *D. mauritiana* separating from *D. simulans* approximately 0.26 million years ago (Kilman *et al*., 2000) (Figure 1). These species within the clade are able to hybridize and produce viable fertile females in the laboratory despite acquired differences in morphology of genitalia and divergence in proteins expressed within the gonad tissue (Thomas & Singh, 1992; Civetta & Singh, 1995; 1998). The male hybrids produced, though viable, are sterile (Kulathinal & Singh, 1998), which may relate to the fact that genes related to male reproduction are highly divergent between the

species. Here, I used a set of sixty *D. simulans* lines with single and unique mapped introgressions of *D. mauritiana* DNA into their third chromosome (IG lines) (Tao *et al*., 2003) in a double mating experiment to identify recent genetic changes that associate with breakdown in sperm competitive ability. Males from each of the IG lines were competed as second males against *D. simulans* males, and as a result, a subset of IG lines were identified as poor sperm competitors (*P2* <0.5). Males from these IG lines share two distinct *D. mauritiana* introgressed regions on their third chromosome, one in the 77B to 84B and the other in the 88B to 92E cytogenetic map range, suggesting a phenotype with a complex genetic basis. Eighty-one candidate genes were identified within this mapped region based on expression in male reproductive tissue and/or association with male sex and reproduction. Candidate genes were narrowed down by assaying differences in gene expression between *D. simulans* and *D. mauritiana.* By associations between differences in gene expression and poor sperm competitive ability, I identified 5 genes of interest, three of them located within the 89B chromosomal position. Using an evolutionary approach that fits models specifying different rates of non-synonymous and synonymous substitutions within coding sequences and along the *D. simulans* and *D. mauritiana* lineages, I identified an additional ten candidate genes that have likely undergone species-specific adaptations. There is limited functional information available for all fifteen candidate genes; however, *Mst89B* is particularly interesting given the fact that this gene has been shown to indirectly interact with *Acp62F* (Giot *et al*. 2003), a gene known to influence male sperm competitive ability in *D. melanogaster* (Fiumera *et al*. 2007; Mueller *et al*. 2008).

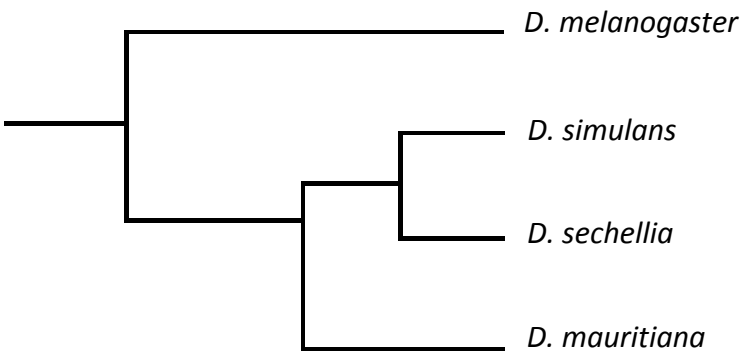

Figure 1: Phylogeny of *Drosophila simulans* clade in relation to *D. melanogaster*

Members of the *Drosophila simulans* clade diverged from the common ancestor *D. melanogaster* approximately 0.5-1 million years ago. *D. simulans* and *D. mauritiana* diverged from a common ancestor approximately 0.26 million years ago.

Methods

2.1 Fly Stocks and Maintenance

I use males from a set of 60 *D. simulans* strains that contain *D. mauritiana* mapped genetic introgression into their third chromosome (IG lines). Each generation, the IG lines were maintained by selecting orange-eyed males carrying a *D. mauritiana* P-element insert and crossing them to virgin females from a *D. simulans* B strain (white eyes) (see Tao *et al*., 2003). A stock of *D. simulans* ebony mutant flies (*e*/*e*) (black body) was acquired from the *Drosophila* Species Stock center (Tucson, Arizona: stock 14021-0251.033). *D. simulans*, *D. simulans* B, *D. simulans* (e/e), and *D. mauritiana* (*Drosophila* Species Stock Center, 14021-0241.01) were maintained in bottles containing standard cornmeal-molasses media on a 12 hour light/ dark schedule at 22˚C. Every generation, parental flies were collected, transferred into new bottles and left to mate. After seven days the adult flies were discarded. Prior to setting up crosses for the experiments, males and females were collected from the stock on a 5-hour cycle to ensure virginity (Greenspan, 1997). Collection and sexing of the flies was carried out under CO_2 gas anaesthetic. Males and females were separately aged to 3-6 days old in vials containing cornmeal molasses media and containing no more than 20 flies.

Table 1: Protocol for Standard Cornmeal Molasses Agar Media

Ingredient	Quantity
Cornmeal	65 g
Yeast	13 g
Agar	6.5 g
Cold water	170 ml
Boiling Water	760 ml
Molasses	45.5 ml
99% Propionic Acid	5 ml
10% Tegosept **(50g methyl hydroybenzoate/500ml 95% ethanol)**	20 ml

Cornmeal, yeast and agar were mixed in cold water. Water was brought to a boil in a pot on hot plate. The cold mixture was added to the pot of boiling water and stirred well. The mix was then brought to a boil. Molasses was added to and mixed well. Media was removed from heat and cooled to 60˚C. Tegosept and propionic acid were added. A peristaltic pump was used to aliquot food into vials or bottles, these were then covered with cheese cloth and left to set overnight. Vials were plugged with cotton balls or bottles capped with paper lids, these were then placed in a 4˚C fridge to store. Media was removed 4-8 hours before use (so media was at RT).

2.2 Anesthetising Flies

Flies were anesthetised using slow flow CO_2 gas. CO_2 gas was briefly vented into either food vials or food bottles to incapacitate the flies. Flies were then dumped onto a permeable cellulose membrane and examination of flies was done with a low volume of CO_2 flowing through the membrane to ensure that flies remained immobile. Every effort was made to use the lowest volume of CO_2 possible for the shortest duration of time.

2.3 Counting and Sexing Flies

Examination of flies was done using a binocular Nikon dissecting microscope under 20X magnification. Flies were manipulated with a fine tip paintbrush to prevent injury to the flies and their wings. Flies were sexed based on appearance of the sexes (Figure 2) and differences in external genitalia (Figure 3).

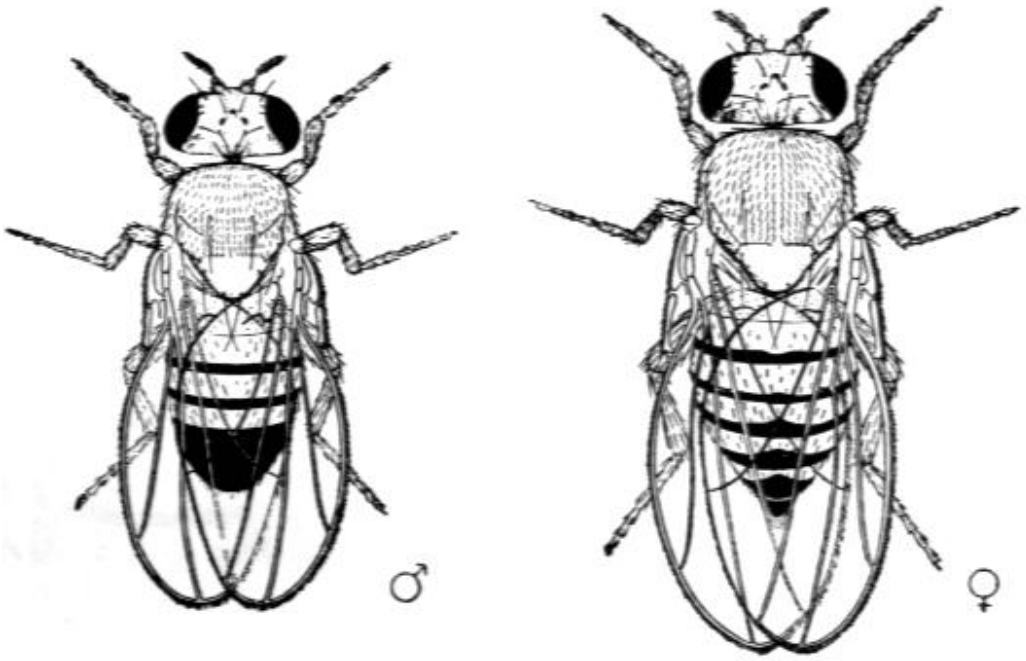

Figure 2: Male and Female *Drosophila*

www.chs.k12.nf.ca/science/b3201/Flylab/male-female.jpg

Figure 3: Morphology of *Drosophila* Genitalia

(Illinois State University, Edwards Lab. photo by Kevin Edwards)

2.4 Mouth Aspirating Flies

Flies were transferred between bottles using a mouth operated aspirator (mouth-pooter). These were hand constructed in our lab using a piece of clear flexible plastic tubing, a glass pasture pipette, pippetting tips, cheese cloth, and tape. The glass pipette tip was broken off with a glass file and flamed using a Bunsen burner to round the edge. The end of the pipette was covered with a piece of cheese cloth and inserted into a piece of clear flexible tubing. A plastic pipette tip was covered at the end with a piece of cheese cloth and this was inserted into the other end of the flexible plastic tubing acting as a mouth piece.

2.5 Dissection of Male Reproductive Tracts

Male reproductive tracts were dissected out of flies aged 3-6 days. Flies were decapitated using fine forceps and placed into 20µl of Phosphate Buffered Solution (PBS) on a glass slide under 50X magnification using a binocular Nikon dissecting microscope. Using a pair of fine forceps and a dissecting needle the abdomen was opened and the male reproductive tract was carefully removed. Using the dissecting needle, excess biological material from the abdomen was removed from the male reproductive tract and discarded. The male reproductive tissue including the seminal vesicles, testicular duct, ejaculatory duct, ejaculatory bulb, testis and accessory glands (Figure 4) were transferred

to a 1.5ml Axygen RNase free tube containing 400µl of RNAlater® and stored at -20°C until RNA extraction.

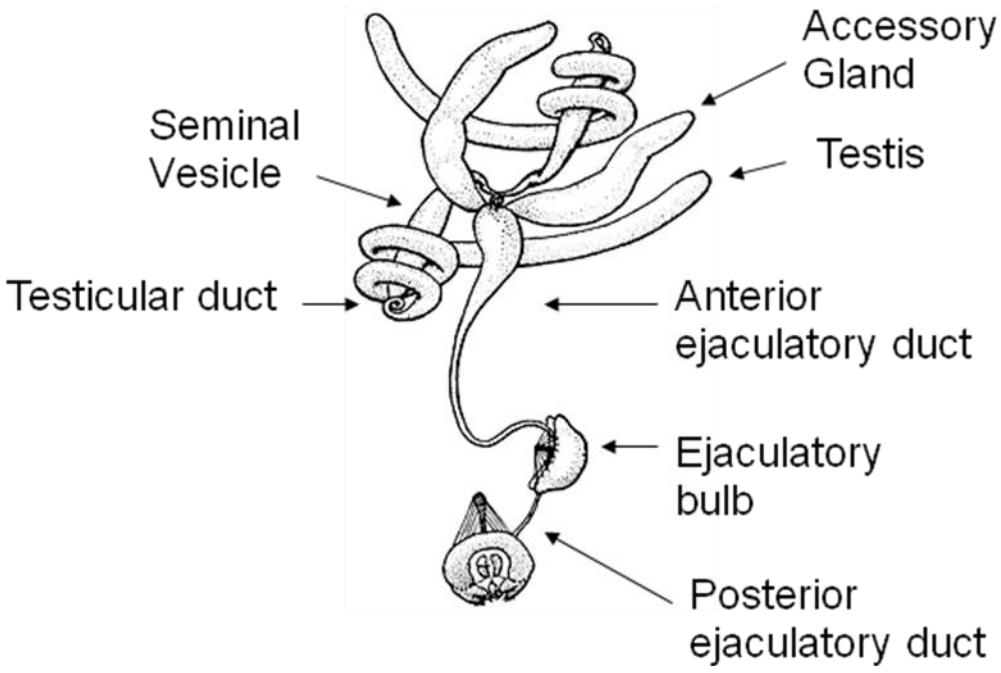

Figure 4: Male Reproductive Tract of *Drosophila* Species

Drosophila protocols (Sullivan *et al*., 2000)

2.6 Sperm Competition Assays

Virgin 3-6 day old *Drosophila simulans* females homozygous for the *ebony* (*e/e*) mutation were mated to same-aged *D. simulans* (*e*/*e*) males. The mating was done *en masse* for a period of two hours in a vial containing 10 females and 20 males. Females were then individually aspirated (using a mouth-pooter) into separate vials (vial 1) and males were discarded. Two days later, the known remating interval for *D. simulans*, two males from an IG line were aspirated into vial 1 containing the female. Vials containing media and flies were placed horizontally against a white background so that the flies were easily visualized. Mating was observed every 15 minutes for a total period of eight hours. Observed copulations were recorded. Males were discarded and females were aspirated into new vials (vial 2). Four days later each female was aspirated with a mouth-pooter and individually transferred to vial 3. Progeny from vials 1, 2 and 3 were counted on the 23rd day after the beginning of ovipositioning and scored based on phenotypic body coloration (wild type *vs*. ebony) (see Figure 5). Females that did not produce ebony progeny in vial 1 were discarded from further analysis (i.e. no evidence of first mating). The fraction of wild-type progeny in vials 2 and 3, sired by the IG male, was designated as *P2* (Boorman & Parker, 1976). The 60 strains were tested over time in 9 blocks (groups of IG lines tested at one time) with partial replicas (IG lines tested in more than 1 block).

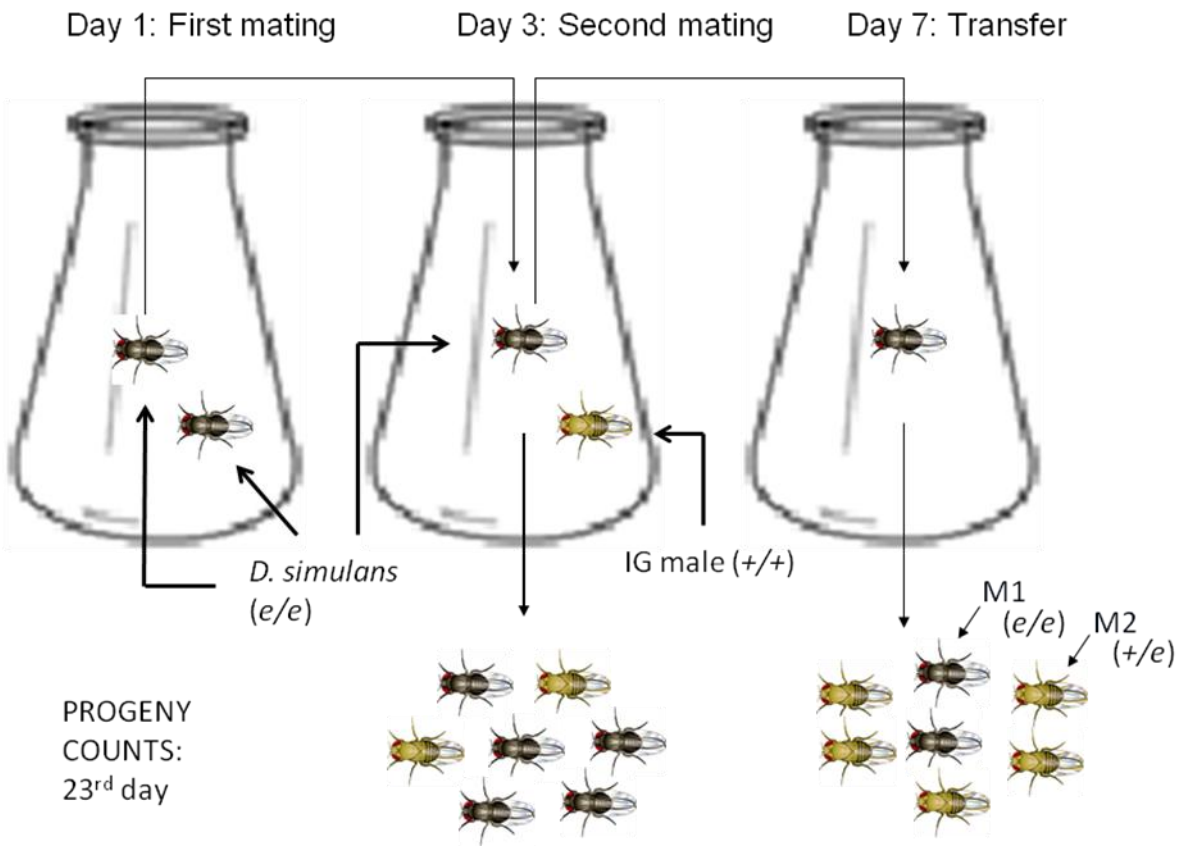

Figure 5: Double Mating Model

Diagrammatic representation of the double mating experiment used to test for sperm competitive breakdown in the IG lines. Note that *D. simulans* females and males have ebony body color (e/e) and the IG lines are wild type (+/+). M1= sired by male 1 (*e/e*) M2= sired by male 2 (*+/+*).

2.7 Fecundity Assays

Virgin *D. simulans* (*e/e*) females aged 3-6 days were mated *en masse* in a vial containing cornmeal-molasses media with males from one of twelve IG lines selected on the basis of their average *P2* scores. Ten females were placed in a vial with 15-20 males from an IG line and left together for a period of eight hours, at the end of which time the females were individually aspirated to a new vial (vial 1). Females were left to oviposit in vial 1 for 7 days and then were individually aspirated into a new vial (vial 2). Counts of progeny were taken on the 23rd day after the beginning of ovipositioning in both vials 1 and 2. Any female who failed to produce progeny, who produced less than 10 progeny or who died prematurely, was discounted from the analysis (due to the possibility of injury or illness in the female).

2.8 Copulation Duration Assay

Virgin *D. simulans* ebony females were collected and aged to 3-6 days in vials containing no more than 20 flies. Females were individually aspirated into fresh food vials and then were presented with one or two males from one of the twelve IG lines tested for fecundity. Fly vials were surveyed every 5-10 minutes and upon initiation copulations were timed with a lab timer. After copulation was complete, duration was recorded and the males were removed using a mouth aspirator.

2. 9 Statistical Analysis

I used angular transformation (Sokal & Rohlf, 1995) of *P2* in the analysis of variance (ANOVA) among the 60 IG strains tested, with both strain and block as variables. Males from IG lines showing average *P2* values lower than 0.5 were considered poor sperm competitors; as *P2* equal to or higher than 0.6 are commonly observed in intraspecific tests of sperm competition. Because the *D. mauritiana* introgressions have been previously mapped (Tao *et al*., 2003), we were able to establish associations between mapped *D. mauritiana* introgressions and *P2* lower than 0.5.

Comparisons of means between two groups were done using a randomization test that calculates the mean value and the difference between two treatments. The entire dataset is then randomly reallocated to two groups of equal size as the samples and a new difference between groups is estimated. This step is repeated 25,000 times to obtain a random distribution of differences. If the observed difference between the two samples is beyond the 5% tail of the random distribution ($P < 0.05$), then the two samples or treatments are considered significantly different.

2.10 Candidate Genes

Candidate genes within mapped introgressed regions that associated with average *P2* scores lower than 0.5 were identified using the genetic map of *D. melanogaster* available at Flybase (http://flybase.bio.indiana.edu/). Candidate genes were selected

based on their location, a section of the third chromosome that showed significant effects on sperm competitive ability, and narrowed down by focusing on genes expressed in male reproductive tissue. Tissue of gene expression was determined using a gene expression search via term link available at Flybase. Termlink categorizes genes by anatomy followed by organ systems. Within organ systems, it was possible to narrow down the search to the male reproductive system. I also conducted a reverse search by searching for all genes within the male reproductive system and then narrowing down the search by chromosome map position. Additional candidate genes were identified, on the basis of mapped chromosome position, from the *Drosophila melanogaster* sperm proteome paper (Dorus *et al*., 2006), from a study that examined differences in gene expression between closely related *Drosophila* sister species (Haerty & Singh., 2006) and from candidate accessory gland proteins (Haerty *et al*., 2007).

2.11 DNA Extraction

DNA was extracted from whole flies. 12 Flies were collected using CO_2 gas and placed in a 1.5µl Axygen tube. 100µl of homogenizing buffer (10M Tris-HCL pH 8.0- 100 ml, 0.5M EDTA - 2 ml, 10% SDS- 1 ml, SDDW to complete total volume to 10ml) was added to the tube, tissue was then homogenized using the OMNI tissue homogenizer on low (variable speed) for 20 seconds. Samples were incubated in a VWR heating block at 70°C for 30 minutes. 14µl of 8M potassium acetate was added to samples and placed on ice for 30 minutes. Samples were centrifuged at 14,000 RPM using an eppendorf centrifuge

at 4°C for 20 minutes. Supernatant was then transferred to a new 1.5ml tube and 50µl of 100% isopropanol was added; samples were then inverted several times to ensure adequate mixing and left at room temperature for 10 minutes. Tubes were placed at room temperature and spun at 14,000 RPM for 10 minutes to pellet DNA. The supernatant was discarded and the remaining pellet was rinsed twice with 50µl of 70% ethanol by flicking and inverting the tube several times. Any traces of ethanol were carefully removed by briefly spinning the tube at room temperature to pull all remaining alcohol to the bottom of the tube and by using an eppendorf pipette for alcohol removal. Finally, the tube was inverted and DNA pellets were dried for 30 minutes followed by 30 minutes right side up (total dry time 1 hour). 40µl of SDDW was added to each tube to re-suspend the DNA, the tube was placed in a heating block at 37°C and checked (tube flicked) periodically until entire pellet was re-suspended.

2.12 Primer design

Oligonucleotide primers for PCR and sequencing (Table 8) were designed using Primer3 software (Rozen & Skaletsky, 2000) on the basis of conserved sequence regions in the *D. melanogaster*, *D simulans* and *D. sechellia* alignments found at the UCSC Genome Browser (http://genome.ucsc.edu/).

2.13 PCR

PCR was used to amplify genes of interest in *D. mauritiana* and *D. simulans*. PCR was carried out using a MJ Research PTC-200 Peltier Thermal Cycler using reaction conditions listed in Table 2. PCR conditions were optimized for each pair of primers using both a gradient of $MgCl_2$ (1.0mM-3mM) and also a T_M (annealing temperature 50˚C- 65˚C) gradient in the thermocycler program (Table3).

Table 2: Reaction Mixture for PCR

Volume	Ingredient
14.6 µl	SDDW
1.5 µl	$MgCl_2$[a] (3 mM concentration)
2.5 µl	PCR Buffer
0.6 µl	DNTPs
0.2 µl	TAQ DNA polymerase
19.4 µl Total	Master mix*
1.8 µl	Forward Primer (100X dilution)
1.8 µl	Reverse Primer (100X dilution)
2 µl	DNA
25 µl Total	Per Reaction

*Master Mix was mixed together for all reactions in a single tube, and then 19.4 µl was aliquoted per reaction into individual reaction tubes. Volume of primers and DNA were added to each individual reaction.

[a]- Volumes of SDDW and $MgCl_2$ were modified depending of concentration of $MgCl_2$ needed.

Table 3: Thermocycler Program

Step	Condition
1	95°C for 2 minutes
2	95°C for 1 minute
3	55-65°C for 30 seconds*
4	72˚C for 1 minute 30 seconds
5	Go to step 2, 29 times
6	72˚C for 10 minutes

* annealing temperature was adjusted based on optimization for each set of primers

2.14 Evaluating DNA for Non-Specific Products

4 µl of cleaned PCR product sample was loaded with 2 µl of tracking dye into a well of a 1% agarose gel (Table 4). The resolution of the samples was compared with that of the1µl of 1Kb Plus DNA Ladder™ (Invitrogen, Cat. No. 10787-018) with 2µl of tracking dye to validate size of PCR products and to check for non-specific bands.

Table 4: Protocol for 1% Agarose gel

Volume	Ingredient
0.75 grams	**1% OmniPur Agarose**
75 ml	**1X TBE**
3.75 µl	**Ethidium Bromide**

Agarose was weighed out and placed in a 250 ml Erlenmeyer flask, 1X TBE was added and swirled to mix. The mix was placed in a microwave oven and heated until it boiled. The liquid was cooled to 45˚C; ethidium bromide was added and swirled until mixed. The liquid was then poured into a gel box and left to set for 1 hour.

2.15 Purifying PCR Products

PCR products were cleaned using E.Z.N.A. ® Cycle-Pure Kit (Q-Spin Column) (Omega, Cat.D6493-02) according to the manufacturer's recommendations. PCR products were transferred into a clean 1.5 ml microphage tube and 4 volumes of Buffer CP (cycle pure) were added to the tube. Each tube was briefly vortexed to mix buffer and PCR product and pulsed in the centrifuge at RT. HiBind® DNA columns were placed in the provided 2 ml collection tubes, 100 µl of equilibration buffer was added to the Hi bind column, incubated at RT for 4 minutes and pulse spun at MAX speed for 20 seconds. The PCR sample was added to the HiBind® column and centrifuged at 10,000 x g for 1 minute at RT. Flow through was discarded. The HiBind® column was washed with 100 µl of DNA wash (diluted according to manufacturer's specifications with absolute ethanol) and centrifuged at 10,000 x g for 1 min at RT. Flow through liquid was discarded and the HiBind® column was washed a second time using 500 µl of DNA wash. Flow through liquid was discarded and the empty HiBind® DNA column was spun at 13,000 x g for 2 minutes at RT to dry the column matrix. HiBind® column was placed into a clean 1.5 ml microfuge tube and 40 µl of elution buffer was added to the HiBind® column matrix. Tubes were centrifuged at 13,000 x g for 1 minute at RT to elute DNA from the column.

2.16 Quantification of DNA

Prior to sequencing, aliquots of cleaned amplified DNA were quantified by running 2 µl of each DNA sample against 2 µl of Low DNA Mass™ Ladder (Invitrogen, Cat. No. 10068-013) in a 1% agarose gel. Gels were run submerged in 1X TBE at 120 volts in a CBS Scientific Co. (model MGU420T) gel box. Then, using a Bio-Rad Gel Doc 2000, images were captured and analyzed using Quantity One® (Bio-Rad version 4.3.0*)* software and approximate concentration of DNA ng/ µl of cleaned PCR product were calculated. Images were captured using 'automatic exposure' and selecting 'freeze'. I selected 'detect lanes' and adjusted lane width until just wider than bands. Standards were defined using the quantity standards option, I selected create new and entered the value in ng of each band of a Low Mass™ Ladder (invitrogen, Cat. No. 10068-013). I then selected 'calibrate' and chose 'gel' to calculate the values for all of the bands in the gel.

2.17 Sequencing

Cleaned and quantified PCR products were sequenced on a Beckman Coulter CEQ 2000XL automated sequencer available in Dr. Civetta's lab at the University of Winnipeg using the GenomeLab™ DTCS Quick Start Kit (Beckman Coulter, Cat. No. 608120) according to protocols provided by the manufacturer (Table 5) using the recommended thermocycler program (Table 6).

Table 5: Reaction Mixture for Sequencing (per single reaction)

Volume	Ingredient
2 µl	Quick start
1.2 µl	Reaction Buffer
0.25 µl	Pellet Paint
4.55 µl	SDDW
8 µl Total	Master Mix*
2 µl	Primer (100X dilution) (Forward or Reverse)
10 µl	(DNA template[a] + SDDW)
20 µl Total	Reaction Mixture

*-Master Mix was mixed together for all reactions in a single tube, and then 8 µl was aliquoted per reaction into individual reaction tubes. Volume of primer, DNA and SSDW were added to each individual reaction.

[a]- amount was determined by calculating approximate ng/µl of DNA in cleaned PCR product based on manufacturer's recommendations accounting for fragment length and optimal DNA concentration.

Table 6: Thermocycler Program for DNA Sequencing Reaction

Step	Condition
1	96°C for 20 seconds
2	50°C for 20 seconds
3	60°C for 4 minutes
4	Go to step one 29 more times
5	Hold at 4°C

After the sequencing reaction was complete in the thermocycler, DNA was precipitated using an ethanol precipitation protocol as directed in the GenomeLab™DTCS Quick Start Kit (Beckman Coulter, Cat. No. 608120). Sterile 0.5 ml microfuge tubes were labelled and used for each sample. Fresh Stop Solution (Table 7) was prepared and added to the tube after the sequencing reaction was completed.

Table 7: Protocol for Stop Solution

Volume	Ingredient
2 µl	3M Sodium Acetate (pH 5.2)
0.4 µl	100mM Na2-EDTA (pH 8.0)
1.6 µl	SDDW
1 µl	20mg/ml Glycogen
5 µl Total	Per Reaction

5 µl of stop solution was added to each of the labelled tubes. The sequencing reactions were then transferred to appropriately labelled 0.5 ml microfuge tubes and mixed thoroughly. 60 µl of ice cold 95% (v/v) ethanol/dH_2O from the -20°C freezer was added to each tube and mixed thoroughly. Tubes were then immediately centrifuged at 14,000 RPM at 4°C for 15 minutes. The supernatant was carefully removed with a micropipette. The pellet was then rinsed a further 2 times with 200 µl 70% (v/v) ethanol/ dH_2O from -20°C freezer. For each rinse, the tube was centrifuged immediately at 14,000 RPM at 4°C for 2 minutes. After centrifugation all of the supernatant was removed with a micropipette and tubes were inverted to dry for 30 minutes. After 30 minutes the tubes

were turned right side up and dried for another 10 minutes. The pellets were re-suspended in 40 µl of Sample Loading Solution (provided with kit). Once the pellet was re-suspended in the sample loading solution, the samples were transferred to a previously designated well of the sample plate using a pipette. Each sample in each well of the sample plate was then overlaid with 1 drop of mineral oil (provided in kit).

Table 8: PCR/ Sequencing Primers for genes sequenced in *D. mauritiana* and *D. simulans*

Gene	Primers
CG5178	TGTGTGAGTGCGACCTCAAT* TACTCGACATGGAGCACAGC
CG9391	GACAATGCATGGTTTTTCCAT* CGTCCACGTTGTGTGACATT TGGTCACTTTGCCCATCATA* AGTTCCTTCTGGCCGCTAAC GCTGGGCTTGGTCTACAATC* ACTAGGCGTAGTCAATTTGTTTG
CG14891	CCACATGACTCACCTCTTCG* ATTCGTAGATCGTCTTGAATC CGCCAGTTGTAAAAGATTCA* ATGTCGGTAACCTCGAGAACC
CG1041	TGTTGAAGGAAACGCAAGTG* TTCCTCCACATACATGGCG GGGAFAATCAACGACATGG* TCCCTGAAGAGCCAGCTTAG TTAAAATGGGCAAGAACGAG* GCAGGACGTGGTACTTCAGC
CG7362[A]	TTTGAGGCCCTTGAACACTC* ACTCAGCGCTCACATGACTC GCTTGTGCTCAAGGTGTGTT* CTGGCAATGGCAATCGTG CATTCACGCTGGGATGAAG* CCAAAGTGACCCCATTCAAG CAAGTGCAACAAGGTTGGAA* GCATTGACCAGCACCAAAGT

CG9063[B]	AGTCCTTCTCCTGCGACTCA* AGGTACAGGCAGTCGTACC GCTTCTCTCTCTGGCAATGG CAACACGACTGTGCCATTTC GGCCGTACCTATTTTCGATG* CGTGCATGTGATCGGTCT AGCAAGCAATATGCCACCA* CCCCCGGTACAAAGATCC

[A] gene sequenced in *D. simulans* and *D. mauritiana*

[B] gene sequenced in *D. simulans*

*denotes a forward primer

The quality of the sequences was assessed by examining the sequences by eye and the current level throughout the sequencing was checked for drops or changes in current which indicates problems with sample purity. Sequences were assembled by aligning *D. mauritiana* trace files to alignments of *D. simulans*, *D. sechellia* and *D. melanogaster* using ClustalW (Thompson *et al*., 1994) in MEGA (Kumar *et al*. 2001). This allowed for visual identification and clarification of ambiguous nucleotides (N).

2.18 Candidate Genes: DNA Sequence Data Analysis

The DNA sequence of candidate genes was first retrieved from the genome data available for *D. melanogaster*, *D. simulans* and *D. sechellia* (http://genome.ucsc.edu/*).* For each gene, amino acid sequence alignments were performed using the ClustalW program (www.ebi.ac.uk/Tools/clustalw2/index.html) within the DAMBE software package (Xia & Xie, 2001) and used as reference to generate nucleotide sequence alignments. I tested for

significant variation in rates of evolution along branches leading to each of the species by comparing the likelihood of a free-ratios model of evolution (M1) to the likelihood of constant ratio model of evolution (M0) using the phylogenetic analysis by maximum likelihood (PAML) software package (Yang, 1997). Difference in rates of evolution can be indicative of species-specific adaptations. Genes demonstrating significant variation in rates of evolution across lineages were further evaluated for evidence of acceleration and/or deceleration of evolutionary rates along the *D. simulans* lineage. This was accomplished by comparing the likelihood of a model that assumes a free ω (d_N/d_S per codon) value for the foreground *D. simulans* branch and a different fixed ω for the background branches (M2) to the likelihood of the constant (M0) ratio model.

I tested for evidence of positive selection along the *D. simulans* branch using the mixed branch-site model (model = 2; NSsites = 2) within codeml in PAML (Yang, 1997; Yang & Nielssen, 2002). The log-likelihood of the branch-site model is compared to the same model; however, the ω value was fixed to 1 in the *D. simulans* foreground branch so that significant variation in ω between the foreground and background branches is attributable to positive selection only (eliminating differences in selective constraints) (Zhang *et al*. 2005). It is possible that rapid evolution and positive selection could occur along other branches within the *Drosophila* phylogeny; however, the purpose of my thesis is to identify genes as candidates for CSP between *D. simulans* and *D. mauritiana.* Thus, my analysis was restricted to the *D. simulans* lineage in *D. melanogaster*, *D. simulans* and *D. sechellia* comparisons.

According to Flybase, multiple *D. melanogaster* genes lacked orthologs in *D. simulans* and/or *D. sechellia*. Genes lacking orthologs in *D. simulans* were partially sequenced in *D. simulans* (strain 'sim2', Winters, CA; From Dr. A.G. Clark) to confirm nucleotide changes and/or indels found in the published sequences (http://genome.ucsc.edu/index.html) that lead to stop codons. Genes for which orthologs are reported missing in *D. simulans* were partially sequenced in our D. simulans strain. I also sequenced *D. mauritiana* for genes showing evidence of positive selection suggesting *D. simulans* species specific adaptations.

2.19 RNA Extraction

Male reproductive-tract tissue samples stored at -20°C in RNA*later*® (Ambion, Cat. No. 7024) were pelleted by adding 400µl of ice cold PBS to the tube. The tubes were then spun at 7,000 RPM at 4˚C for 7 minutes to pellet male reproductive tracts. All liquid was carefully removed and discarded using a pipette. 750µl of TRIzol® was added to the 1ml RNase free Axygen Scientific tube containing the pelleted male reproductive tracts. Tissue was then homogenized using an OMNI tissue homogenizer (OMNI international) with RNase free tips at medium speed for 15-30 seconds under a fume hood. Tubes were then placed in an eppendorf centrifuge at 4˚C and spun at 11,000 RPM for 4 minutes. The supernatant was then transferred to a new 1.5ml RNase free tube and left to sit at RT for 5 minutes. After this time, 150µl of chloroform was added to the tube; the tube was then closed and shaken vigorously by hand for 15 seconds then left to sit for 3 minutes at RT.

The microfuge tubes were then centrifuged at 11,000 RPM for 15minutes at 4˚C. The top aqueous layer was carefully transferred to a new 1.5ml RNase free tube. 1µl of 20mg/ml glycogen and 375µl of isopropanol were added to the aqueous phase and mixed by inverting 15 times and then the tubes were left to sit at RT for 10 minutes. The tubes were then centrifuged at 11,000 RPM for 10 minutes at 4˚C. The supernatant was then removed using a pipette and discarded. The pellet was washed with 1ml 75% ethanol (the tube was gently flicked to release the pellet from the bottom) and centrifuged at 9,000 RPM for 5 minutes at 4˚C twice. A pipette was then used to remove the ethanol from the tube and the pellet was dried inverted for 10 minutes at RT. Once the pellet was dry, it was re-suspended by adding 50µl of nuclease free water to the tube and placing the tube in a heating block at 55˚C for 10 minutes (gently flicking the tube occasionally to ensure the pellet was dissolving). The labelled tubes were stored at -70˚C until further use.

2.20 Candidate Genes: Gene Expression Data Analysis

All candidate genes were tested for differences in gene expression between *D. simulans* and *D. mauritiana* using the MiniOpticon Real-Time detection system from Bio-Rad. Oligonucleotide primers were designed by aligning sequences between sibling species *D. melanogaster*, *D simulans* and *D. sechellia*. Attempts were made to design primers over introns, for ease of detection of DNA contamination, and within conserved regions between the 3 species. Primers were designed using Primer3 software (Rozen & Skaletsky, 2000) to produce amplicons of similar sizes (200-250 base pairs), similar CG

content and similar T_M (melting temperature). Primers were designed using the *D. melanogaster* mispriming library to ensure sequences for primers were as unique as possible in an attempt to decrease the likelihood of amplifying non-specific products. An iScript™ One-Step RT-PCR Kit with SYBR® Green was used for qRT-PCR according to manufacturer's suggested protocols; however, reactions volumes of 12.75µl were used to conserve reagent (Table 9).

Table 9: Reaction Mixture for RT-PCR

Volume	Ingredient
6.5 µl	SYBR® Green
2 µl	RNAse free H_2O
1 µl	RNA
0.25 µl	iScript™
9.75 µl Total	Master Mix*
1.5 µl	Forward Primer (10X dilution)
1.5 µl	Reverse Primer (10X dilution)
12.75 µl Total	Per Reaction

*-Master Mix was mixed together for all reactions in a single tube, and then 8 µl was aliquoted per reaction into individual reaction tubes. Volume of primer, DNA and SSDW were added to each individual reaction.

C_t values were normalized to a known *Drosophila* housekeeping gene which is routinely used in relative expression RT analysis, *RP49* (aka *RpL32*). Relative expression was calculated between *D. simulans* and *D. mauritiana* species by calculating $2^{-\Delta\Delta Ct}$ (Livak

& Schmittgen, 2001). Two biological replicates were carried out for each gene in both *D. simulans* and *D. mauritiana*. Technical replicates were carried out on a subset of genes in both species to ensure accuracy in pipetting technique. To test for significant differences in gene expression and to control for experiment wise type I errors for the large number of multiple tests, we generated an experiment wise statistical threshold by using the five and ten percent tails of a population of 1,000 average values obtained by randomly sampling from the data with replacement.

Table 10: List of RT-PCR primers used for relative gene expression analysis between *D. simulans* and *D. mauritiana* for 81 candidate genes

Gene	Primers
Abdominal A	CACTACTTAACTCGGCGAAGG*
	GATGATCGAGTGGTGGTCCT
Abdominal B	GGGATTCGAGACGGACAC*
	AGTAGGTGGGCGGCTCATAG
Actin 79B	GTAATCACCATTGGCAACG*
	TTGGAGATCCACATCTGCTG
Actin 88F	GATCACCATTGGCAACGAG*
	TTCGAGATCCACATTTGCTG
a-Tubulin at 84B	ATGCGGCCAACAACTACG*
	ATCAGACGGTTCAGGTTGGT
CG10284	AAGCTGGTCACCCAGCTCC*
	CCTCAAGACACAATCCCTCA
CG10317	GCGCAAAAGGATGGACAC*
	AGTGTGTCGAATTGTGGTTTTGC
CG10326	ACCTGGACCTGTCGCTGATT*
	TGGCCACATGAGGATGAGTA
CG1041	AGGCGAAGAGACCGACGA*
	GAGAGCATCGGCAAGTTTGTC
CG10510	GGAAAACTACCTCCAAGAACTGG*
	AACGCACGTGCTGATCCT
CG10589	TGGTAAGGATTAGTGGCTTCG*
	TACTGTTCGGCCTGGTTGAT
CG11779	ATGGTGATGATCGGCTTCTT*
	TGTGGCGAAACATGTCGTAT
CG12162	GGCCGCCTTTGAGGCGAAC*
	GGAGCGTGCGTAAGGAACT
CG14448	AGTGGTTGCGAGCTCCATAG*
	AAGGCTTGCAGTCTGTCCAC
CG17122	TCACAAGGAGGCGCTCTATAA*
	CGGACCAGCTCGTTCTTG
CG17387	AAGGCTGAGAAGGAAGTGGTC*
	CCTTCGCCTTCTCTATTTTGG
CG17556	ACCTGGACTTGTCGCTGATT*
	TGGCCACATGAGGATGAGTA
CG31287	GCAGAGCTATCCGAGTCCGG*

	AGCTGCTCGAACTCCTTTTG
CG31294	GCGACATTCGCAAGAAAT*
	AATGGGCCGGATCGAGATGG
CG31418	GTGGATCGCGAGATCTTCC*
	ACCCTTGTCGTGAACAATCC
CG31542	GTGTGGCAAGTGCAACAACT*
	AGTGGCACTCCGTCTTGA
CG31546	GGAGGGCCTCATATGCGTGAT*
	AGCGGGTGAACTGGTCCAC
CG32436	TTGGAGAAAGGAGGATGATGA*
	CAATCCTTCTCGCGCATT
CG32444	ATGACCGTGTCGGTGATCCA*
	TAGGTGATGAGGCACGTCAG
CG32445	AAGCTGGAGGACGTGTGC*
	AGGGTGACACCGTCCTTGT
CG3321	GTTTCCCCGCTGATCAAGT*
	TGCTTGGGAGTGGGCTTC
CG34053	AGCCCAGACAACGGATTGTT*
	GACAGCGACTTCCGAATGAC
CG3517	CATGCTCCTTTGGGCATAAT*
	GCCTCGTAGGCCATCTGTTC
CG3610	CAAACCGATGAGGATGTTAC*
	GGCCACAGCTATTCGATGAT
CG3731	GCTGCCACCCTACAGAAGAC*
	GCAGGTTGCTCTCCACTTC
CG4390	TCCGCTTTATCCTCCTTTTTC*
	AACAGGATATTGCGGCTTTG
CG4546	CATTCGCTTTGTACGCAAGT*
	GTCGTGATGGTCGAATCC
CG4686	GGCAGCCACTACCACTTCAT*
	GCCATCAGGCAGAATCCTC
CG4836	TGCAGACCAAATACCCCATC*
	GATAATGCCGATTCCATGAC.
CG5265	GAAGGAGGGCGCTAAGCTAC*
	TCCGGTAGATAACCACCACAT
CG6125	AGCACACGTTCGGCTTGT*
	CCGTACTCCGAGGAGAGTCC
CG6136	CCATTTACTGCATGCTGAGG*

	AGCTGTGCCAGGCAATCTAC
CG6255	CCACTCGGAGGAGTCCATAA*
	GAGACTATGCCCACCACTCC
CG7131	GCAGCACGGTGTACAAGC*
	GGGAAACAGGTGACATCCAC
CG7145	ATCGAGAAGGCCATCAAAAC*
	AGCTGGCGTGTAGGACAAGT
CG7342	GTCTGCCCATTCTGATCGTG*
	TCAGGCACGATGAGGTAGAA
CG7362	CTCGGGAAGAACATCCAACT*
	GTAGGAAAAGCGTTCGTGAG
CG9063	AATGCGCCTATGAATTGGAC*
	CAGTATTGGAAGCCATACGC
CG9389	GGAGTACACCACCAAGAAGC*
	GCCCGAGCTGTACTCCTGTA
CG9391	AACAATGAGCAGCGACAGG*
	AATGCGTGCACAAAGTTCAT
CG9593	AGCGACTGAACCAGGAAGAA*
	ACTCGTTGCCAAACCAGAAA
CG9772	TTTAAGTGGCTCCCGAAGAA*
	AGTGTCAGCAGTGAGCTTCG
Cyclin H (CG7405)	GCGACCTCAATGAGCACTT*
	ACTATGTCTGTGGCCTTGTT
dj like	TATCGGTGCCTTCAGACTTC*
	GCCTGCATGGTGATTTCTTT
don juan	AAGGAAGGGAATCAGGATGAA*
	TTTTCCTTTTCGGCCAAT
Doublesex	CCACGACCCTGTTGAGTC*
	AGCGGCTGCCATGTGGAC
dynein light chain 90F	GCGAAGAAAGCCAGTTCATT*
	CGATGCAGTACATGGTCTTGTT
Effete	CAGCCCTTATCAAGGAGGTGT*
	GCATACTTTCTAGTCCACTCTCG
Fruitless	GACTTTGGCCAACTGGAAA*
	CGGCACTCAATAGTGTCCAA
Fumble	ACATGGGCAACTTTCTGTCG*
	TTGTAGTTGTCGGGACCGTA
Gilgamesh	CGATGATCTGGAGGCATTG*

	TGTCTTCCCTGTCCAGTCG
Grappa	ACGGACAGGACAAGCACGAC*
	ATAGCTCGTCTCGCCGTAAA
Histone H4 replacement	AGCGTCATCGTAAGGTGCTT*
	TTGAGGGCATAGACCACGTC
Kokopelli	CCTGTGAACTCGCAGAGC*
	AGGTGCCTTTTGCAGTACCA
Male-specific transcript 77F	CCTCATCGTCAGGCTTTGTT*
	CAGCCACTTATTTCAGCGTGC
Maternal transcript 89Bb	GACCACCAGCTCTTCGATTC*
	TAGCACCGACGCTACTGATG
Mediator complex subunit 17	AGCCCATTGTGCAGGTGTA*
	AGCTCCGCTGGAACAATCAC
meiosis I arrest	GCTCTCAAAGCCATCTCCAC*
	GTCGAACTGGCGCACTATGT
Miranda	GTTTCAACGACGTGGATGTG*
	GGCTGACTTGGCCTCCTTTA
Mst89B	TTATTGTGCTCACCCAGCAG*
	CACAGTTGTCTGCGTTTTGG
Noisette	GCCCAACGAGTTCAATGAGT*
	GGTGCAGGTAGTCGTTGAGG
Nanos	GCTCCTTCTGCAGCGACAGT*
	CGAGTGGCATGGGCATTAAG
Polo	CAGCCTTAACCATCCGAACA*
	AGAGATTGCCCAGCTTCAGA
Qm	GGATTTTCCCCTGTGCGTTC*
	TAAACTTAGCACGCCGCAAA
Rhodopsin 2	GAGACGCCATTCGACCTG*
	CCAGGTTGAGAACGAGCAA
ripped pocket	GACTCTGGAAACGGGCTATG*
	TGGCAATTTGATTCGGTGTA
SAK	CGCAGGACCTGATAAACAAA*
	CAGTTGAAGCTTGCCCAAAT
Sex combs reduced	GGCCTATACGCCCAACCTGT*
	TGGTGGACAGATCCTGTGG
Skuld	GTCAATCAACACGCACATGG*
	CCTCGTCGGTGTATCCCTTA
Spase 18/21-subunit	AGCATGTTGCAGATCGACGA*

	ACGCGATGCACAATGGGTAT
spermatocyte arrest	GCCACGTTTTGGATCTAGG*
	TCCATTTCCACGTCTTCCTC
spindle E	ATGGCCACGGAGTCCAAATC*
	TGATTACCGGATGAGCATTG
Sungrazer	GTACGCCCTGAACAACAGAT*
	TACGCCACCAGTGATCTGTC
Tim17b1	ATACCGATTGAGCGGAGGAT*
	TTCGCATATTCGTGCGATAA
CG14891	AGATCGGGATGGGTTCTTTT*
	TCTTGAATCTTTTACAACTGG
CG6040	TGCATTCGATGATGTGATGA*
	CAGCGCTGAAGGTCTAGCTC

*denotes forward primer

Results

3.1 Double Mating Experiments/ Sperm Competition

Sixty male IG lines were evaluated in a double mating experiment for evidence of sperm competitive breakdown. For all 60 IG lines a total of 2,635 *D. simulans* ebony females were set up to doubly mate. Counts were taken from vial 2 and vial 3 on the 23rd day after the beginning of ovipositioning in that vial. Progeny were determined to be *P1* (progeny produced by the first male to mate) or *P2* (progeny produced by the second male to mate) based on phenotype (ebony vs. wild type).

P e/e♀ x e/e♂
x +/+♂
↓
F_1 e/e vs. e/+
(ebony) (wild type)
Father e/e +/+

Figure 6: Double Mating Experiment using *ebony* Mutation to Determine Paternity

During data analysis, females were discarded if they failed to produce progeny with ebony body coloration (no verification of successful 1st mating) or failed to produce progeny with wild type body coloration (no verification of successful 2nd mating). Therefore, the data from 913 females were removed from final analysis, leaving a remaining sample population of 1,722 females to be analyzed. Angular transformation of

P2 values adequately fit the assumption of homoscedastisity and normality satisfying the requirements for ANOVA (analysis of variance). ANOVA revealed significant variation among average *P2* scores of the 60 IG lines tested ($F_{59,1600}$= 6.92; P< 0.001) and 12 lines had *P2* scores lower than or equal to 0.5. A *P2* value of 0.5 would be expected if there was free sperm mixing and equal sperm usage; however, as mentioned earlier we expect to see *P2* values greater than 0.5 owing to the normal second male precedence that is typically observed. Therefore these 12 IG lines demonstrated *P2* values contradictory to second male precedence observed in conspecific crosses. As all IG lines were not tested at the same time the data were analyzed for differences in the time blocks in which they were tested (lines were replicated in multiple blocks). A significant block effect ($F_{8,1600}$= 5.95; P< 0.001) was observed. This indicates that some of the differences between IG males are attributable to differences in environmental conditions between blocks. However, there was no significant IG line x block interaction ($F_{54,1600}$= 1.15; P= 0.210) and the rank order of average score of IG lines was consistent over blocks indicating that environmental factors equally effect all strains. The data were re-analyzed using only IG lines in which 10 females or more were successfully mated to IG males and produced 10 or more progeny per female (Civetta *et al*., 2005). The results were similar in that this analysis revealed significant variation between IG lines and a significant block effect with a non-significant IG line x block interaction. When these conditions were satisfied, only 10 IG lines were shown to have *P2* values lower than 0.5 (Figure 7).

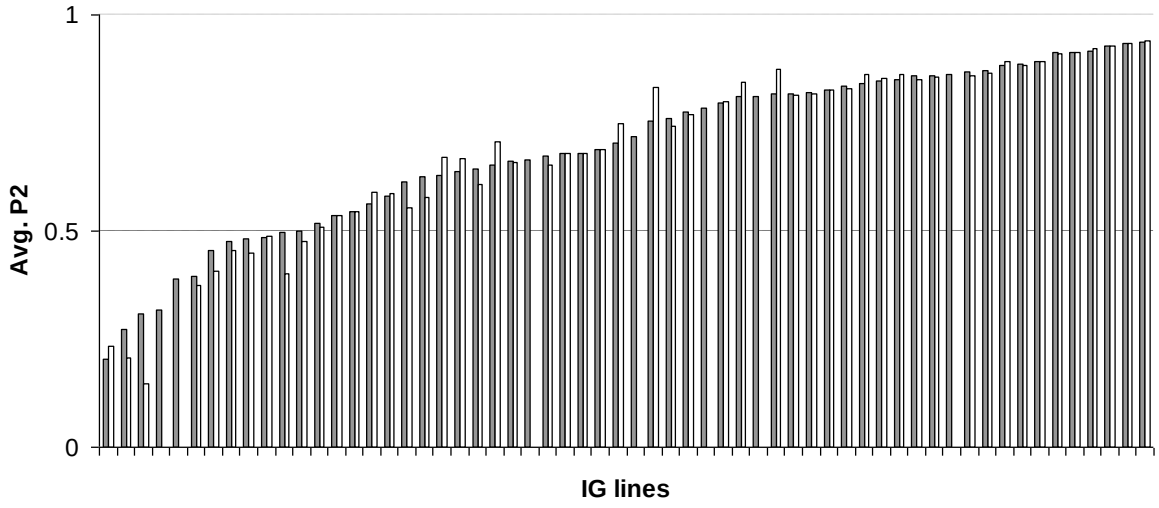

Figure 7: Average second male paternity success (*P2*) for males from 60 different *D. simulans* introgressed (IG) lines. For each line we show averages when all data are included (grey bars) and when data are analyzed by removing IG lines where less than 10 males remate and females produce less than 20 offspring (white bars).

Second male paternity success scores were positively and significantly correlated with measures of female induced fecundity (where induced fecundity reflects the males ability to produce progeny) and remating (Pearson Correlations for *P2* with fecundity and remating: 0.119 and 0.140 respectively; P< 0.001). I therefore addressed whether low *P2* values were fully explained by low induced fecundity. A single mating assay was conducted using *D. simulans* females and a subset of males from 12 IG lines with *P2* values both lower than or greater than 0.5 (Figure 8). Significant variation in single-mating fecundity among females mated to the different IG lines tested ($F_{12,205}$= 9.94; P< 0.001) was observed and fecundity was positively correlated with *P2* (0.420; P< 0.001). Therefore, second male paternity success of IG males is affected by low induced fecundity. To investigate whether this low induced fecundity was entirely responsible for low *P2* values, estimates of competitive male paternity success were calculated by combining estimates of female fecundity in a single mating assay involving *D. simulans* ebony and IG males. This was done by dividing the number of progeny produced by *D. simulans* ebony females singly mated to males from different IG lines (Fec_{IG}) by the number of progeny produced by *D. simulans* ebony females mated to *D. simulans* ebony males (Fec_e) plus the number of progeny produced by females of the same ebony strain mated to males from the 12 IG lines tested (Fec_{IG}). The expected *P2* score then becomes $P2_{exp}= Fec_{IG} / (Fec_{IG} + Fec_e)$. For all IG lines evaluated, the estimate of the IG males' competitive paternity success turns out to be greater than 0.5 and significantly different (Randomization test $D_{12,12}$= -0.245; P < 0.001) from the observed *P2* values obtained in double mating experiments.

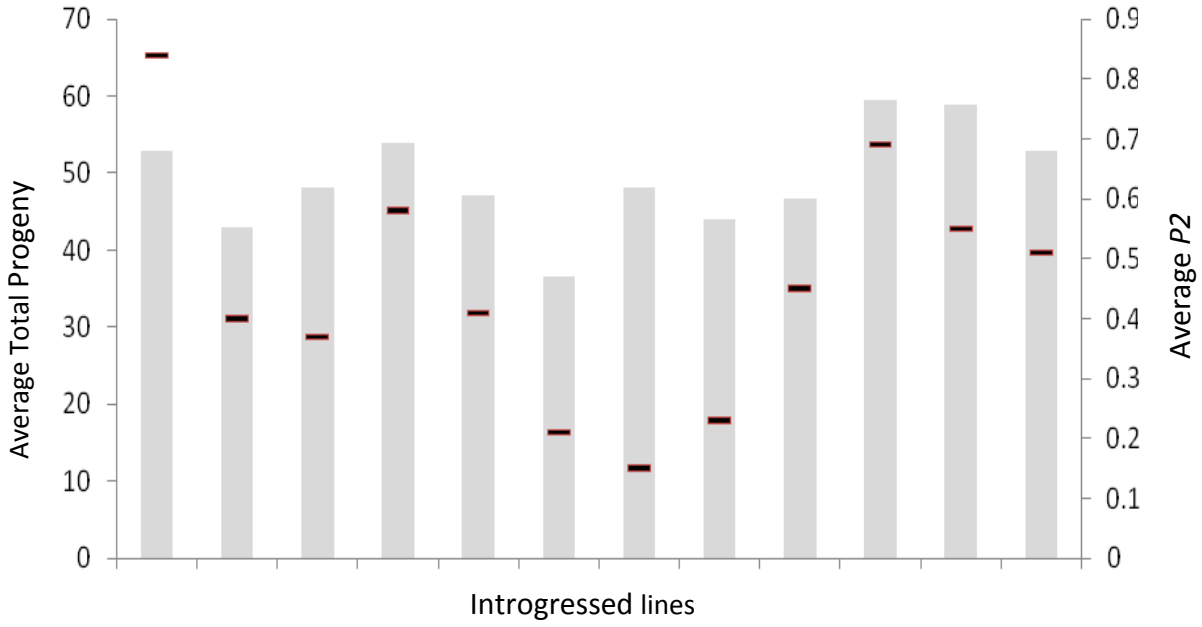

Figure 8: Overall Fecundity

Average total progeny in single matings for 12 randomly selected IG lines with *P2* values ranging from 0.15- 0.84. Average total progeny represented by grey bars. Average *P2* values represented by black lines.

Average *P2* scores lower than 0.5 could be influenced by copulation duration. Copulation duration was evaluated as a possible confounding variable affecting *P2*. Copulation duration was measured for a subset of 12 IG lines with varying average *P2* values (both greater than and lower than 0.5) and found no significant variation among IG lines ($F_{11,188}$= 1.13; P= 0.341) with all IG lines having copulation duration times more comparable to conspecific *D. simulans* x *D. simulans* matings than heterospecific *D. simulans* x *D. mauritiana* matings (Figure 9).

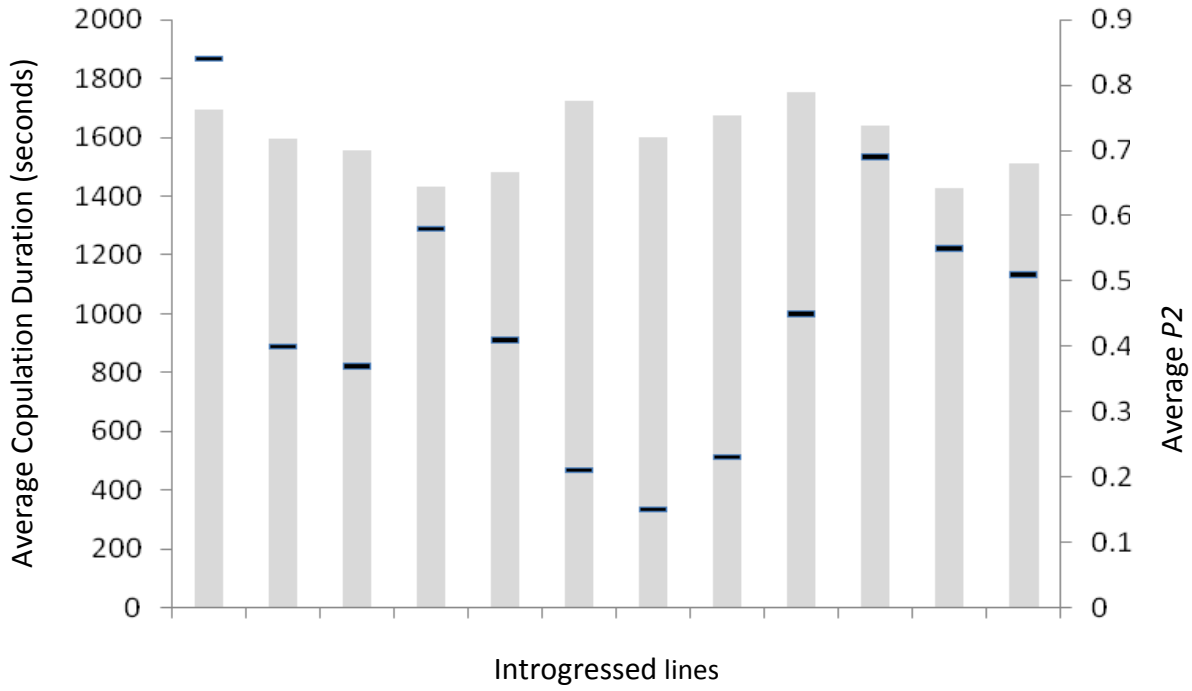

Figure 9: Copulation Duration

Average copulation duration in seconds of 12 randomly selected IG lines with P2 values ranging from 0.15- 0.84. Average copulation duration represented by grey bars. Average *P2* values represented by black lines.

3.2 Identification of Broad-Sense Candidate Genes

Since the map position of the introgressions is well defined in the IG lines, it was possible to determine loci responsible for sperm competitive breakdown (*P2* less than or equal to 0.5). The twelve IG lines showing average *P2* scores lower than 0.5 have a *D. mauritiana* introgression within one broad region of the third chromosome. This region can be divided into two loci, corresponding to map positions 77B to 84B and 88B to 92E, in *D. melanogaster*. When the *D. mauritiana* introgression spanned both of these chromosome regions, there was evidence of sperm competitive breakdown (loss of second male precedence) whereas if only one region was involved in the introgression, *P2* values were unaffected (Figure 10). This demonstrates that gene interactions underlie the manifestation of CSP. Using the chromosome map information gathered from the double mating experiments and the 12 IG lines showing sperm competitive breakdown 81 broad-sense (preliminary) candidate genes were identified on the basis of reproductive function and/or male reproductive tissue of expression.

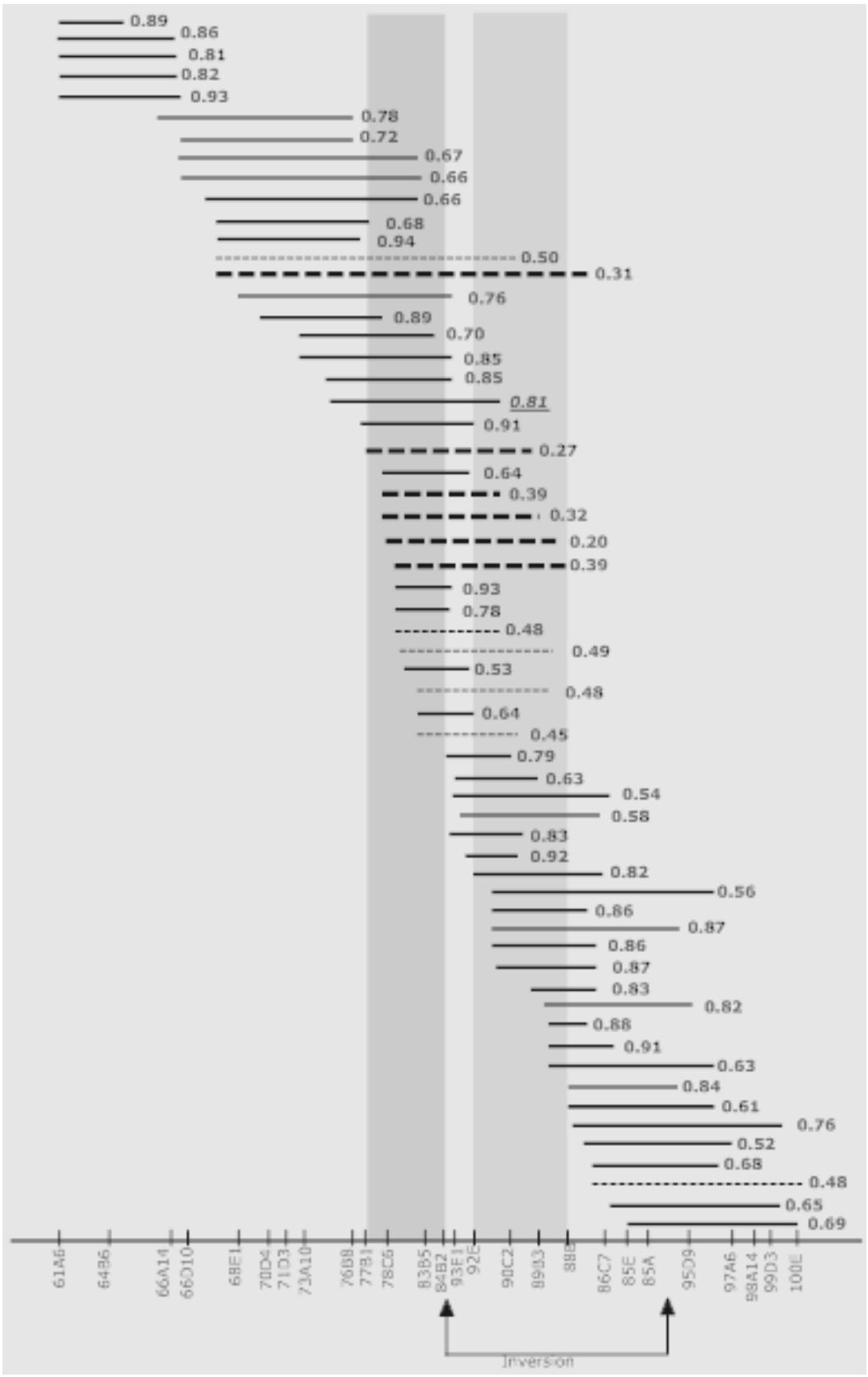
0.89
0.86
0.81
0.82
0.93
0.78
0.72
0.67
0.66
0.66
0.68
0.94
0.50
0.31
0.76
0.89
0.70
0.85
0.85
0.81
0.91
0.27
0.64
0.39
0.32
0.20
0.39
0.93
0.78
0.48
0.49
0.53
0.48
0.64
0.45
0.79
0.63
0.54
0.58
0.83
0.92
0.82
0.56
0.86
0.87
0.86
0.87
0.83
0.82
0.88
0.91
0.63
0.84
0.61
0.76
0.52
0.68
0.48
0.65
0.69
61A6
6486
66A14
66D10
68E1
70D4
71D3
73A10
7688
7791
78C6
8385
84B2
93F1
92E
90C2
8983
88B
86C7
85E
85A
9509
97A6
98A14
99D3
100E
Inversion

Figure 10: Map position of 60 *D. mauritiana* introgressions within the *D. simulans* third chromosome (modified from Tao *et al*. 2003).

Two shaded areas represent the two loci introgressions required for a breakdown in second male paternity success (*P2* < 0.5). Average *P2* values are given beside the line denoting the position of the introgression. Dashed lines are used for IG males with average *P2* equal or lower than 0.5, with those having a *P2* between 0.45 and 0.5 shown as thinner lines. Notice one value higher than 0.5 (underlined) containing the two candidate loci suggestive of the possible existence of a suppressor somewhere between 73A10 and 77B map position. Other introgressions outside the mapped loci with average *P2* lower than 0.5 is suggestive of other loci responsible for second male paternity breakdown. Arrows indicate an inversion present in the *Drosophila* simulans clade when compared to *D. melanogaster*.

3.3 Candidate Genes: Relative Gene Expression

Candidate genes were identified by using data gathered from double mating experiments using IG lines. However, ultimately the question is whether these genes are responsible for CSP as a postmating prezygotic isolating barrier between *D. simulans* and *D. mauritiana*. Therefore, it is important to identify differences in expression of the candidate genes in pure species as opposed to the IG lines. For this reason, real-time PCR (qRT-PCR) analysis of the 81 broad sense candidate genes (preliminary based on fulfilling criteria of tissue of expression and cytolocation) was carried out in *D. simulans* and *D. mauritiana* species. This allowed for identification of differences in expression of the candidate genes between the two species. Using RNA samples extracted from male-

reproductive tracts of both *D. mauritiana* and *D. simulans* quantitative real-time PCR (qRT-PCR) from reverse transcribed products corresponding to our 81 broad-sense candidate genes was carried out. Data analysis revealed five (CG10317, CG14891, Mst89B, CG6040 and CG4836) and eight (same as before plus CG3610, CG17387 and CG31287) candidate genes with significant differences in gene expression between the two species using either a five or ten percent threshold level respectively (Table 11, Figure 11). This result is not qualitatively different when using two-fold average differences in gene expression and its 95% confidence interval as threshold. Only one (CG17387) of the eight genes is located in the 77B to 84B region and three of the five differentially regulated genes (CG10317, CG14891 and Mst89B) as well as CG31287 are located in the 89B position. This suggests that the evolution of species-specific co-regulation patterns of this gene cluster could be critical during species diversification and the evolution of CSP.

Table 11: Differentially regulated genes between *D. simulans* and *D. mauritiana*

Gene	**Cytolocation**	**Known/Predicted Function**	**Up-regulated**
CG10317	89B13-89B14	Unknown	*D. mauritiana*
CG14891	89B20-89B20	Nucleic acid binding	*D. mauritiana*
Mst89B	89B9- 89B9	Unknown, found to interact with 16 other proteins by yeast 2 hybrid (Giot *et al*., 2003)	*D. simulans*
CG6040	91F2-91F2	Unknown	*D. mauritiana*
CG4836	92C1-92C1	Oxidation reduction	*D. simulans*
CG3610*	88D5-88D5	Unknown, found to interact with 25 other proteins by yeast 2 hybrid (Giot *et al*., 2003)	*D. simulans*
CG17387*	82D5-82D5	Nucleotide binding, found to interact with 1 other protein by yeast 2 hybrid (Giot *et al*., 2003)	*D. mauritiana*
CG31287*	89B7-89B7	Protein folding	*D. mauritiana*

Note: *denotes significant at 10% threshold.

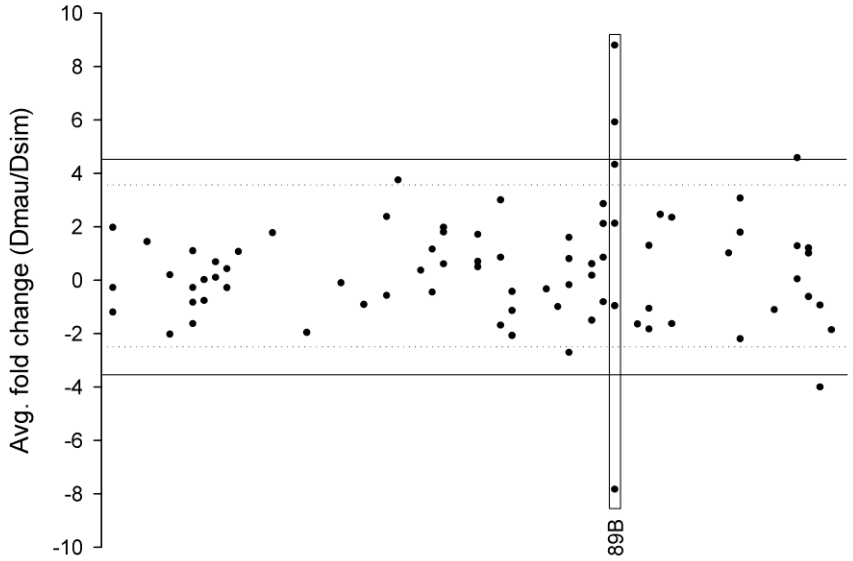

Figure 11: Assessment of Diversification

Average fold difference in expression from male reproductive tract RNA extractions for 81 candidate genes between *D. simulans* and *D. mauritiana*. The differences in gene expression are shown as *D. mauritiana* relative to *D. simulans* (Dmau/Dsim). The X axis shows the cytogenetic map position with the 89B location boxed. Experiment-wise statistical threshold at P< 0.05 and P< 0.1 are shown by solid and dotted lines respectively.

3.4 Candidate Genes: DNA Sequence Data Analysis

If species specific interactions are broken down in heterospecific crosses due to the presence of translated products that differ in function, we expect to see species-specific signals in phylogenetic lineages leading to *D. simulans* and/or *D. mauritiana*. I analyzed all 81 broad-sense candidate genes for evidence of variation in rates of evolution among lineages in *D. melanogaster*, *D. simulans* and *D. sechellia* comparisons using currently available sequence data from the 12 *Drosophila* species genome project. Seventeen genes, equally distributed along the two mapped loci, revealed variable evolutionary rates among lineages based on the comparison of the free-ratios model to the one-ratio model. Seven out of the seventeen genes showed significant acceleration in the *D. simulans* lineage relative to the other two background lineages, with another four genes showing significant deceleration. One gene (CG14307 *fruitless*) showed both acceleration and deceleration depending on the *D. melanogaster* alternative translation product used for analysis. Fruitless is known to have alternate translation products in males and females and is found near the bottom of the sex determination hierarchy. The gene is important in developing males specific musculature and male courtship behaviour and is believed to be involved in neural aspects of sexual dimorphism. Particularly interesting are genes that demonstrate not only evidence of change in rates of evolution but also species-specific adaptive diversification. Genes CG7478, CG31542, CG1984, CG3158, CG14307 and CG6255 exhibited both accelerated evolution and positive selection in the *D. simulans* lineage. Four other genes, CG9389, CG15179, CG31287 and

CG4836, did not show evidence of a significant acceleration or deceleration but show evidence of positive selection in the *D. simulans* lineage (Supplementary Table 1).

Seven genes lacked *D. melanogaster* orthologs in *D. sechellia* and/or *D. simulans* due to the presence of indels and/or nucleotide substitutions leading to the occurrence of stop codons within the coding sequence (missing in *D. sechellia*: CG9391, CG34357, CG1041, CG7362, CG5178, CG14891; missing in *D. simulans*: CG9063, CG34357, CG7362). With the exception of CG34357, whose gene region spans 64Kb (Flybase), we partially sequenced all other six gene coding sequences in *D. simulans* and/or *D. mauritiana* and found that the lack of orthology is either restricted to *D. sechellia* or simply the result of sequencing errors in the genome database entry (Supplementary Table 1). We therefore used our *D. mauritiana*, and in some cases *D. simulans*, partial sequences to test for variable rates of evolution and positive selection along the *D. simulans* lineage in *D. melanogaster*, *D. simulans* and *D. mauritiana* sequence alignments. Genes CG7362 and CG14891 showed evidence of positive selection along the *D. simulans* branch (Table 12)

With the exception of CG14307, we also partially sequenced *D. mauritiana* for genes showing evidence of positive selection along the *D. simulans* lineage and tested them using both PAML branch and branch-site models using the *D. simulans*, *D. sechellia* or *D. melanogaster*, *D. mauritiana* trio (Supplementary Table 1). In the 77B-84B locus, we detected evidence of positive selection in either *D. simulans* and/or *D. mauritiana* for CG7478, CG31542, CG1984 and CG1041. In the 88B-92E locus, CG7362, CG3158, CG31287, CG14891, CG6255 and CG4836 all showed evidence of positive selection. Only two

(CG31287 and CG14891) of these ten genes are located within a common cytogenetic map position, 89B. This result reinforces our previous observation, based on gene expression analysis, that the 89B position might have been critical during species diversification and the evolution of species-specific adaptations to postmating prezygotic reproductive challenges.

Table 12: Testing for adaptive diversification in *D. simulans* and *D. mauritiana*

Ten of twelve genes previously genes previously detected as experiencing positive selection along the *D. simulans* branch show evidence of adaptive diversification in either *D. simulans* or *D. mauritiana*.

		D. simulans[a]				***D. mauritiana***[a]			
Gene[b]	**Map**	**Model 1 ℓ (ω=1)**	**Model 2 ℓ (ω)**	**2Δℓ**[c]	**ω**	**Model 7 ℓ (ω=1)**	**Model 8 ℓ (ω)**	**2Δℓ**[c]	**ω**
CG9389	78C3	-2413.9	-2413.9	0		-2413.9	-2413.9	0	
CG7478	79A6	-1212.7	-1206.4	12.6***	713.4	-1213.8	-1213.8	0	
CG31542	83A1	-837.1	-822.5	29.2***	62.8	-837.3	-837.3	0	
CG1041[d]	83E4	-2386.2	-2382.2	0		-2382.2	-2341.1	82.2***	999
CG15179	84A1	-747.2	-747.2	0		-747.1	-747.1	0	
CG1984	84B2	-1861.6	-1841.2	40.8***	20.6	-1865.5	-1865.5	0	
CG7362[d]	88D2	-2202.4	-2199.2	6.4*	999	-2202.4	-2199.2	6.4*	999
CG3158	89A5	-4249.9	-4232.3	35.2***	999	-4250.2	-4250.2	0	
CG31287	89B7	-1041.3	-1021.5	39.6***	999	-1041.3	-1041.3	0	

CG14891[d]	89B20	-2691.4	-2688.1	6.5*	17.4	-2691.4	-2688.5	5.7*	15.7
CG6255	92A5	-1229.1	-1189.7	78.7***	999	-1230.1	-1230.1	0	
CG4836	92B4	-5881.8	-5881.8	0		-5880.8	-5864.5	32.6***	999

[a]Foreground branch being tested.

[b]Candidate genes previously detected as experiencing positive selection along the *D. simulans* branch in comparisons with *D. melanogaster* and *D. sechellia* or with *D. melanogaster* and *D. mauritiana* (see Supplementary table 1).

[c]For each gene tested, we compared the likelihood (ℓ) of the branch-site model (Model=2; NSsite=2) with the same model but fixing the ω value of the foreground branch to 1. The test statistics follows a chi-square distribution with one degree of freedom. *** P< 0.001; ** P< 0.01; *P< 0.05.

[d]Comparisons are *D. melanogaster – D. simulans – D. mauritiana*.

Discussion

The third chromosome has been previously identified to be associated with differences in second male paternity success (sperm competition) in the *Drosophila simulans* clade (Civetta *et al*., 2002). In a double mating experiment, *D. simulans* females were mated to a *D. simulans* male and also a male from one of 23 *D. simulans* lines with a small *D. sechellia* genetic introgression on one of the three major chromosomes. Significant variation in second male paternity success, fecundity, and mating discriminations were observed (Civetta *et al*., 2002). A QTL on the third chromosome, *D. melanogaster* map position 69D-77A, was found to be associated with interspecific differences in male fecundity between *D. simulans* and *D. sechellia* (Civetta *et al*., 2002). As suggested from this work, fine mapping would be needed to identify candidate genes within the QTLs as the resolution achieved by QTL mapping is limited by the number of molecular markers used to construct the genetic map. Subsequent to the identification of a third chromosome QTL with a role in differential male fecundity, it was a natural follow up to design an experiment using a series of *D. simulans* x *D. mauritiana* introgressed (IG) lines. Dr. Y. Tao (Tao *et al.,* 2003) created the IG lines for the purpose of examining hybrid male sterility (HMS), to evaluate the role of actual genes located on chromosome three. The growth of available resources made it possible to also survey genes within mapped loci based on functional annotations that relate to male reproductive physiology. By using a set of 60 IG lines, with a *D. simulans* background and single well mapped introgressed segments of *D. mauritiana* DNA, which overlapped and spanned the entire length of the

third chromosome, I was able to find two loci on the third chromosome responsible for differences in sperm competitive ability, more specifically CSP, between *D. simulans* and *D. mauritiana*.

A correlation between *P2* scores and male induced female fecundity was observed between the IG lines. This result is not entirely unexpected and it has been previously demonstrated in other experiments in which *Drosophila* species were used (Price *et al*., 2000; Civetta *et al*., 2002, Fiumera *et al*., 2007). *P2* values are in part related to a male's overall ability to induce progeny production by the female and are therefore related to the IG male's overall fecundity. However, it is important to determine whether low *P2* values observed in double mating experiments are truly representative of sperm competitive breakdown and not solely attributable to lower overall fecundity. This is particularly important to clarify in the case of my study as the overall fecundity of these IG lines when the introgression was present in a homozygous state had been previously tested and it was determined that 23 lines were sterile (no progeny), 11 were quasi-sterile (1-10 progeny), 9 were subfertile (11-45 progeny), and 15 were fertile (>45 progeny), 2 lines were not tested (Tao *et al*., 2003b). Thus, although the *D. mauritiana* introgressions in my experiments were maintained in the heterozygous state, the possibility remained that low *P2* values were the result of differences in fertility or the inability of the IG males to induce progeny production in females. There is also the possibility that some IG lines had lower vigour than other lines due to decreased health of the stock. While single mating fecundity did correlate with IG lines sperm competitive ability, the variation

between IG males in average progeny production from single mating assays ranged from 32 - 53 (i.e. no sterility). Moreover, the low number of progeny produced in a competitive setting by the IG lines showing *P2* lower than 0.5 could not solely be explained as a consequence of reduced ability to induce progeny production in females (see results).

In fact, prior studies have shown that reduced fertility in single matings between different species of *Drosophila* cannot fully explain CSP. Price (1997) analyzed the number of progeny produced from both single matings using *D. simulans* females and either *D. simulans* or *D. mauritiana* (heterospecific) male. She also analyzed double matings with *D. simulans* females and both a conspecific male and either *D. mauritiana* or *D. sechellia* male (as both first or second to mate). The number of progeny produced in single *D. simulans* x *D. mauritiana* matings (mean ± standard error; 104 ± 12.5 progeny) was lower than in conspecific *D. simulans* matings (256 ± 16.8 progeny). Of note, the number of progeny produced in the heterospecifc mating was greater than in *D. mauritiana* conspecific matings (81 ± 14.9 progeny). However, in a competitive setting regardless of mating order, the heterospecific males suffered a very drastic reproductive loss demonstrating CSP. For *D. simulans* females doubly mated to *D. mauritiana,* mean progeny production was 20 ± 11.6 (14% of progeny compared to a single heterospecific mating) when second to mate and 1 ± 0.6 (1%) when first to mate (Price, 1997). Similar results were obtained when the same experiments were repeated using *D. simulans* females and *D.* sechellia males (Price, 1997). Therefore, reduced female fecundity

observed in single heterospecifc matings when compared to conspecific matings is not enough to account for the CSP observed in the double mating trials (Price, 1997, 2000).

While the extent of variation exerted by the introgressions on the phenotype measured emphasizes the complex genetic basis of male reproductive success, I have been able to map two loci that appear to be co-required for a breakdown in male sperm competitive ability and paternity success. Moreover, while previous studies have established associations between single genes and variation in first and second male paternity success in *D. melanogaster* (Fiumera *et al*., 2005, 2007), this is the first study to ascertain such associations in crosses between closely related species of *Drosophila*. We therefore provide a genetic basis for a well characterized postmating prezygotic isolation barrier between species of the *Drosophila simulans* clade that has been elusive in an earlier quantitative trait loci (QTL) study between *D. simulans* and *D. sechellia* as previously mentioned (Civetta *et al*., 2002).

There has been a recent surge of interest in determining the genetic basis of reproductive isolating barriers due to the evolutionary implications that the answers will have for outlining the process of speciation and evolutionary divergence. Effort to find genes represents an attempt at understanding the genetic basis of functional changes between species (Wu & Palopoli, 1994). Research using *D. mauritiana* and *D. simulans* sibling species has attempted to reveal the number of genes responsible for HMS. To date the research has revealed that complex epistatic interactions are likely involved in this type of reproductive isolating barrier. An analysis of the affect of *D. mauritiana*

introgressions on overall fertility of the IG lines revealed that not all of the introgressions had negative affects on male fertility as previously thought, and that some of the longer introgressions were, in reality, more fertile than lines that contained shorter introgressions (Tao *et al*., 2003). The interest of Tao *et al*. (2003) was in deciphering the loci responsible for hybrid male sterility between *D. simulans* and *D. mauritiana*. They noted that only one map 'region' of *D. mauritiana* DNA on the third chromosome was capable of eliciting a high level of infertility when introgressed into *D. simulans* by itself (Tao *et al*. 2003). However, at least two to three different 'factors' within the mapped region could elicit partial sterility, suggesting that, in an isolated population, speciation requires multiple allelic fixations over time, with each producing an incompatibility of relatively minute effect (Tao *et al*. 2003). This agrees with the polygenic view of reproductive isolation; however, this does not prohibit a sporadic incompatibility of considerable effect owing to a single pair of allelic fixations (Orr & Coyne, 1992; Orr, 1998); however, these are expected to be uncommon. Similarly, my mapping results identified a minimum of two loci required for the CSP phenotype; notably, this does not rule out additional loci that could cause and/or suppress the phenotype.

The possibility of suppressor loci is suggested by males of one IG line showing no breakdown in second male paternity despite the presence of the two causative loci (see Figure 10, underlined IG line). There is also one introgression outside the mapped area (average *P2* = 0.48) that could represent the presence of additional loci influencing second male paternity success. The mapping indicates that multiple genetic elements and gene

interactions likely underlie the genetic basis of conspecific male precedence in *Drosophila*. This observation is in line with the view that complex epistasis plays a major role during evolution, species differentiation, and isolation. The identification of an epistatic basis of conspecific sperm precedence does not rule out the potential existence of genes of a large effect within our two mapped loci. However, given the complex and epistatic nature of the genome and our relative infancy in understanding gene dynamics, it is difficult to ascertain whether a single gene is solely or mainly responsible for a given phenotype and such endeavours must be approached with caution. Often what appears to be a single locus can be one of multiple genes necessary for the desired phenotype. The mapping and misidentification of *Odysseus* (*Ods*) located in the X chromosome at cytological position 16D (~500kb) as a major reproductive isolation factor responsible for complete hybrid male sterility in interspecific hybrids in *Drosophila* by Perez *et al*. (1993) is a classic example. The experimental design involved introgressing small segments of *D. mauritana* (or *D. sechellia*) into the X chromosome of a *D. simulans* genetic background in order to assess which introgessions were associated with HMS and furthermore to determine whether HMS was the result of a single gene with large effect or multiple genes with lesser effects. They employed multiple techniques to accomplish this including recombination analysis, RFLP (restriction fragment length polymorphism) analysis and SSCP (single strand conformation polymorphism). Originally *Ods* was determined to be solely responsible for HMS between *D. simulans* and D*. mauritiana* by fulfilling three criteria: (1) complete penetrance of sterility, (2) complimentary recombination analysis (should map the gene to the same loci from both ends) and (3) physical demarcation (by a

series of DNA markers) (Perez *et al*., 1993). Interestingly, *D. sechellia* introgressions containing *Ods* did not cause HMS. Further research revealed that the introgression of Ods^{mau} into *D. simulans* by itself was insufficient to cause HMS in *D. simulans* and that it required the co-introgression of at least one or more genes to cause full hybrid male sterility (Perez & Wu, 1995). When working with model organisms containing introgressions as part of the experimental design, one must be aware that the introgression may contain more than one gene affecting the phenotype of interest. The introgression approach is useful for identifying candidate genes; however, these candidates must be carefully evaluated. *Odysseus* is now known as a gene that functions to increase sperm production during spermatogenesis (Sun *et al*. 2004). Sun *et al*. (2004) sought to evaluate *Ods* function to further understand the evolution of postmating isolation. They attempted to answer two questions, firstly what is the normal function of the gene involved in reproductive isolation and secondly how does that normal function relate to its role in incompatibility in hybrids? They concluded that the observable phenotype, in this case HMS, and the underlying genetic process, spermatogenesis, can be weakly coupled in the sense that a gene with a normally weak effect on phenotype has a very robust effect on HMS. This result affirms that gene function can be drastically different in a new genetic environment, as is the case for the normal function of *Ods* and its unpredictable manifestation in hybrids (Sun *et al*., 2004).

I identified candidate genes within the mapped loci on the basis of gene regulatory differentiation and changes in coding sequences driven by adaptive diversification

between *D. simulans* and *D. mauritiana*. Five of the eight candidates identified on the basis of differential gene expression had been previously shown to be differentially expressed in *D. simulans* and *D. mauritiana* using microarray analysis of testes gene expression (Haerty & Singh, 2006) and the other three are genes coding for sperm proteins (Dorus *et al*., 2006). The work by Haerty and Singh (2006) is a good source for validation of my results because the study completed comparative gene expression analysis on testes tissue from hybrids including *D. simulans* crossed with *D. mauritiana*, and also the parental species. Haerty and Singh (2006) used RNA from all hybrids and pure species to hybridize a microarray chip that had been spotted with a cDNA library from *D. melanogaster*. This is a potentially confounding variable as previous research has identified that the estimated sequence difference between *D. simulans* and *D. melanogaster* is approximately 4-8% (Coyne & Krietman, 1986; Caccone *et al*., 1988). However, the authors recognized this potential bias and addressed it by analyzing any possible association between sequence divergence and differences in gene expression. By aligning coding sequences for 354 genes between *D. melanogaster* and *D. simulans* (286 genes) and also *D. sechellia* (68 genes), they found no significant correlation between sequence divergence and differences in gene expression. Therefore the identification of the same five genes in both experiments suggests the differences in expression are real.

An interesting observation is that four out of the eight differentially expressed genes are located within the 89B cytogenetic map position, and so are two of the ten genes showing evidence of positive selection in the *D. simulans* and/or *D. mauritiana*

lineage. It is possible that selection on protein coding genes and coevolution with DNA binding regulatory elements in this particular mapped position could play a major role during the evolution of postmating prezygotic isolation barriers. Selection driven coevolution has been demonstrated for X chromosome dosage compensation and the misregulation of X-linked genes in *Drosophila* hybrids that can lead to inviability (Bachtrog, 2008). In *Drosophila*, dosage compensation occurs by doubling the rate of transcription on the X chromosome in males. A ribonucleoprotein complex, male-specific lethal (MSL), binds to hundreds of sites along the male X chromosome and alters chromatin structure, thereby mediating two-fold hypertranscription of the male X (Bachtrog, 2008 and references therein). It is interesting that recent work has revealed adaptive evolution in all five (or 4/5) genes involved in the MSL complex in *D. melanogaster* but not in two closely related species *D. simulans* and *D. yakuba* (Bachtrog, 2008). Changes in the MSL1 and MSL2 protein domain, which target to the X chromosome, show evidence of positive selection; therefore, represent selection driven c-evolution among DNA-protein interactions (Bachtrog, 2008). As a result, it is posited that the rapid evolution of gene components of the MSL complex may be responsible for incompatibilities between species and postmating isolation barriers. Similarly, it is possible that a rapidly evolving cluster of genes at 89B could represent a source of postmating prezygotic isolation.

The cluster of genes at the 89B third chromosome position spans 13Kb and includes four maternally expressed genes and one testes specific gene. In *D. melanogaster*

MSt89B is known to be expressed in testes tissue and is located within 6kb of two maternally expressed ovarian genes, representing a clustering of genes involved in reproduction. Two 12-bp sequences were identified in the 5'UTR of *Mst89B* with a strong similarity to translational control elements (TCEs) originally identified in the CGP sperm tail protein family (*Mst98Cb, Mst98Ca* and *Mst87F)* genes; however, in a different position relative to the transcription start site. This suggests a potential functional importance either in translational control or other areas of RNA metabolism of genes expressed in the testis (Stebbings *et al*., 1998).

Ten candidate genes were identified on the basis of adaptive diversification along the *D. simulans* and/or *D. mauritiana* lineage. The information available for these genes from studies in *D. melanogaster* reveal little more than the fact that they are linked to male reproduction on the basis of their expression in testes. In species like *Drosophila*, where females multiply mate, it is logical to assume that the adaptive diversification detected for these reproductive genes might be driven by their role in competition for fertilization through male x female and/or male x male interactions. It is suggested that the rapid evolution of male reproductive genes is accounted for by sexual selection pressures acting through sperm competition, female cryptic choice, and sexually antagonistic coevolution (Rice, 1996; Parker & Partridge, 1998; Swanson & Vacquier, 2002). It is interesting to note that one of the ten candidate reproductive genes, CG3158, possibly exerts an effect through genetic conflict. In *D. melanogaster*, mutations at CG3158 (*spnE*), disrupt Piwi interacting RNA (piRNA) formation and therefore increases

the activity of retrotransposons (Aravin *et al*., 2004; Vagin *et al*., 2006). Selfish genetic elements, like transposons, can affect sperm function and impair sperm competitive ability, rendering males who carry them less fertile (Jaenik, 2001; Price & Wedell, 2008). The effect of selfish genetic elements is usually suppressed by modifiers; however, it is possible to speculate that adaptive divergent evolution detected at CG3158 could render sperm unable to compete within a heterospecific female reproductive tract due to the lack of conspecific suppressors acting on it. Two other candidate genes, CG7362 and CG7478, have been characterized as members of the phagocytosis innate immunity system used to internalize pathogens into organelles for destruction (Stroschein-Stevenson *et al*., 2006; Stuart *et al*., 2007). Several studies have shown evidence of tradeoffs between immune function and male reproductive success. Given that the male *Drosophila* mating strategy is to maximize his progeny output, it would be expected that his investment into parasitic/ immune defence will be lower as the pressing selective pressure on the male is to produce progeny (McKean & Nunney, 2001). *D. melanogaster* lines selected for increased sexual activity become more susceptible to infection; when the males are presented with an increasing number of females there is less investment in immune response (McKean & Nunney, 2001). The trade off between male sexual activity and humoral immune response demonstrated coincides with the view that immune function and disease susceptibility are traits governed by tradeoffs (McKean & Nunney, 2001). Similarly, polyandrous *S. stercoraria* males exhibit a lower immune response than monandrous males (Hosken, 2001). While most of these studies report a phenotypic link between immunity and reproduction, here we are able to establish a link between genes

regulating immunity and reproduction. It is therefore possible that the divergent evolution of CG7362 and CG7478, while benign in their own species, could be problematic for foreign heterospecific sperm that need to fight female species specific reproductive tract pathogens.

The identification of more than a single locus effect highlights the importance of disrupted gene interactions causing breakdowns in competitive male paternity success. Giot *et al*. (2003) used a yeast two-hybrid system in *D. melanogaster* to reveal evidence of interactions between proteins. One of my candidate genes, *Mst89B*, is particularly interesting in terms of its protein interactions. Mst89B interacts with Cdlc2, a microtubule motor activity protein expressed in the sperm, as well as the transcription regulator Brinker (*Brk*) (Giot *et al*. 2003). In turn, both Cdlc2 and Brk interact with Acp62F, an accessory gland protein shown to affect a male's ability to place sperm in storage when the gene is knocked out by targeted deletions (Mueller *et al*. 2008). A population survey of sequence variation at *Acp62F* has also established significant associations between polymorphisms at this gene and both second male paternity success and female induced male fecundity (Fiumera & Clark, 2007).

There are no human homologues known for the genes identified as candidates in *Drosophila*; however, BLAST (Basic Local Alignment Search Tool) searches revealed that select candidates exhibit partial homology to human genes. The human genes identified code for proteins that regulate gene expression at the level of transcription, through

protein-protein interactions, as well genes involved in post-translational modifications of other proteins (ubiquitination). The common theme for all homologues identified appears to be affecting gene expression; however, there is not a distinctive stage predominantly targeted for disruption. In fact the variability in types of genes identified is not surprising given that human male sterility, particularly idiopathic sterility, is decidedly complex and its underlying genetic cause remains elusive.

Conclusion

I identified at least two loci responsible for conspecific male precedence. The power of the associations established in my thesis is in its capacity to narrow down, by testing the effect of a sizeable number of genetically manipulated lines on phenotypic variation, a large number of genes to a manageable number of candidate genes. A third of the candidate genes located within these two loci, showing differential gene expression or signature of adaptive diversification between parental species, are located in the 89B map position. My thesis highlights a potential major role for this chromosome position during the evolution of species specific adaptations to differential male fertility and postmating prezygotic reproductive challenges.

Works Cited

ARAVIN, A.A., KLENOV, M.S., VAGIN, V.V., BANTIGNIES, F., CAVALLI, G., and GVOZDEV, V.A. 2004. Dissection of a natural RNA silencing process in the *Drosophila melanogaster* germ line. Mol. Cell Biol. **24**: 6742-6750.

ARNQVIST, G. 1998. Comparative evidence for the evolution of genitalia by sexual selection. Nature **393**: 784-786.

ARTIERI, C.G., HAERTY, W., and SINGH, R.S. 2007. Association between levels of coding sequence divergence and gene misregulation in *Drosophila* male hybrids. Mol Evol. **65**: 697-704.

BACHTROG D., 2008. Positive selection at the binding sites of the male-specific lethal complex involved in dosage compensation in *Drosophila*. Genetics **180**: 1123-1129.

BATEMAN, A.J. 1948. Intra-sexual selection in *Drosophila*. Heredity. **2**(3): 349-368.

BETANCOURT, A.J. and PRESGRAVES, D.C. 2002. Linkage limits the power of natural selection in Drosophila Proc Natl Acad Sci U S A. **21**:13616-13620.

BIRKHEAD, T. R. and MOLLER, A. P. 1998. Sperm Competition and Sexual Selection. Academic Press, New York.

BIRKHEAD, T.R., CHALINE, N., BIGGINS, J.D., BURKE, T., and PIZZARI, T. 2004. Nontransitivity of paternity in a bird. Evolution **58**: 416-420.

BJORK, A., STARMER, W.T., HIGGINSON, D.M., RHODES, C.J., and PITNICK, S. 2007. Complex interactions with females and rival males limit the evolution of sperm offence and defense. Proc. R. Soc. B. **274**: 1779-1788.

BOORMAN, E., and PARKER, G.A., 1976. Sperm (ejaculate) competition in *D. melanogaster* and the reproductive value of females to males in relation to female egg and mating status. Ecol. Entomol. **1**: 145-155.

BRITCH, S. C., SWARTOUT, E.J., HAMPTON, D.D., DRANEY, M.L., CHU, J., MARSHALL, J.L., and HOWARD, D.J. 2007. Genetic architecture of conspecific sperm precedence in *Allonemobius fasciatus* and *A. socius*. Genetics **176**: 1209-1222.

CACCONE, A., AMATO, G.D., and POWELL, J.R. 1988. Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melano*- subgroup. Genetics **118:** 671–683.

CARSON, H. L. 1985. Unification of speciation theory in plants and animals. Syst. Bot. **10**: 380–390

CHANDLEY, A.C. 1998. Genetic contribution to male infertility. Hum. Reprod. 3: 76-83; 84-88

CHANG, A. S. 2004. Conspecific sperm precedence in sister species of Drosophila with overlapping ranges. Evolution **58**: 781-789.

CHAPMAN, T., HERNDON, L.A., HEIFETZ, Y., PARTRIDGE, L., and WOLFNER, M.F. 2001. The Acp26Aa seminal fluid protein is a modulator of early egg hatchability in Drosophila melanogaster. Proc Biol Sci. **268**: 1647-1654.

CHAPMAN, T. and DAVIES, S.J. 2004. Functions and analysis of the seminal fluid proteins of male Drosophila melanogaster fruit flies. Peptides. **25**: 1477-1490.

CIVETTA, A. and 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**: 1085-1095.

CIVETTA, A. and SINGH, R.S. 1998. Sex-related genes, directional sexual selection, and speciation. Mol Biol Evol. **15**: 901-909.

CIVETTA, A. and CLARK, A.G. 2000. Chromosomal effects on male and female components of sperm precedence in Drosophila. Genet Res. 75:143-151.

CIVETTA, A., WALDRIP-DAIL, H.M. and CLARK, A.G. 2002. An introgression approach to mapping differences in mating success and sperm competitive ability in *Drosophila simulans* and *D. sechellia*. Genet. Res. **79**: 65-74.

CIVETTA, A., MONTOOTH, K.L. and MENDELSON, M. 2005. Quantitative trait loci and interaction effects responsible for variation in female postmating mortality in Drosophila simulans and D. sechellia introgression lines. Heredity **94**: 94-100.

CIVETTA, A., ROSING, K.R., and FISHER, J.H. 2008. Differences in sperm competition and sperm competition avoidance in *Drosophila melanogaster*. Anim. Behav. **75**: 1739-1746.

CLARK, A.G., AGUADÉ, M., PROUT, T., HARSHMAN, L.G., and 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., and PROUT, T. 1999. Female × male interactions in Drosophila sperm competition. Science **283**: 217-220.

CLARK, A.G., DERMITZAKIS, E.T., and CIVETTA, A. 2000. Nontransitivity of sperm precedence in Drosophila. Evolution **54**: 1030-1035.

COYNE, J.A. and KREITMAN, M. 1986. Evolutionary genetics of two sibling species, *Drosophila simulans* and *D. sechellia.* Evolution **40:** 673–691.

COYNE, J.A. and ORR, H.A. 2004. *Speciation*. Sinauer Associates, Inc., Sunderland. MA.

CÓRDOBA-AGUILAR, A. and SIVA-JOTHY, M.T. 2004. Sperm displacement ability in *Calopteryx haemorrhoidalis* (Vander linden): Male and female roles, male limits in performance, and female neural control (Zygoptera : Calopterygidae). Odonatologica **33**: 245-252.

DIXON, S.M., COYNE, J.A. and NOOR, M.A. 2003. The evolution of conspecific sperm precedence in Drosophila. Mol. Ecol. **12**: 1179-1184.

DOHLE, G.R., VEEZE, H.J., OVERBEEK, S.E., VAN DEN OUWELAND, A.M., HALLEY, D.J., WEBER, R.F., and NIERMEIJER, M.F. 1999. The complex relationships between cystic fibrosis and congenital bilateral absence of the vas deferens: clinical, electrophysiological and genetic data. Hum Reprod **14**: 371-374

DOBZHANSKY, T. 1970. Genetics of the Evolutionary process. New York: Columbia Univ. Press

DORUS, S., BUSBY, S.A., GERIKE, U., SHABANOWITZ, J., HUNT, D.F., and KARR, T.L. 2006. Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. Nat. Genet. **38**: 1440-1445.

EBERHARD, W.G. 1985. *Sexual Selection and Animal Genitalia*. Harvard University Press, Cambridge.

EBERHART, C.G., MAINES, J.Z., and WASSERMAN, S.A. 1996. Meiotic cell cycle requirement for a fly homologue of human Deleted in Azoospermia. Nature. **381**: 783-785.

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

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

FREEMAN, S. and HERRON, J. 2001. Evolutionary Analysis SECOND EDITION Upper Saddle River, NJ. USA. ©Prentice Hall.

FRIBERG, U., LEW, T.A., BYRNE, P.G., and RICE, W.R. 2005. Assessing the potential for an ongoing arms race within and between the sexes: selection and heritable variation. Evolution **59**: 1540-1551.

FRICKE, C. and ARNQVIST, G. 2004. Conspecific sperm precedence in flour beetles. Animal Behaviour. **67**: 729-732.

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

GEYER, L.B. and PALUMBI, S.R. 2003. Reproductive character displacement and the genetics of gamete recognition in tropical sea urchins. Evolution International Journal of Organic Evolution. **57**: 1049-1060

GEYER, L.B., and PALUMBI, S.R. 2005. Conspecific sperm precedence in two species of tropical sea urchins. Evolution **59**: 97-105.

GIOT, L., BADER, J.S., BROUWER, C,, CHADHURI, A., KUANG, B., LI, Y., HAO, Y.L., OOI, C.E., GODWIN, B., VITOLS, E., VIJAYADAMODAR, G., POCHART, P., MACHINENI, H., WELSH, M., KONG, Y., ZERHUSEN, B., MALCOLM, R., VARRONE, Z., COLLIS, A., MINTO, M., BURGESS, S., MCDANIEL, L., STIMPSON, E., SPRIGGS, F., WILLIAMS, J., NEURATH, K., IOIME, N., AGEE, M., VOSS, E., FURTAK, K., RENZULLI, R., AANENSEN, N., CARROLLA, S., BICKELHAUPT, E., LAZOVATSKY, Y., DASILVA, A., ZHONG, J., STANYON, C.A., FINLEY, R.L. Jr., WHITE, K.P., BRAVERMAN, M., JARVIE, T., GOLD, S., LEACH, M., KNIGHT, J., SHIMKETS, R.A., MCKENNA, M.P., CHANT, J., ROTHBERG, J.M. 2003 A protein interaction map of *Drosophila melanogaster*. Science **302**: 1727-1736.

GREENSPAN, R.J. 1997. Fly pushing: the theory and practice of Drosophila genetics. Cold Springs Harbor Laboratory Press. Plainview, NY. 19-21

GREENSPAN, R.J. and FERVEUR, J.F. 2000. Courtship in Drosophila. Annu Rev Genet. **34**: 205-232

GREGORY, P.G. and HOWARD, D.J. 1993. Laboratory hybridization studies of *Allonemobius fasciatus* and *A. socius* (Orthoptera: Gryllidae). Annals of the Entomological Society of America. **86**: 694-701

GREGORY, P.G. and HOWARD, D.J. 1994. A postinsemination barrier to fertilization isolates 2 closely-related ground crickets. Evolution **48**: 705-710.

GROMKO, M. H., D. GILBERT, AND R. C. RICHMOND. 1984. Sperm transfer and use in the multiple mating system of *Drosophila*. pgs. 371–426 *in* R. L. Smith, ed. Sperm competition and the evolution of animal mating systems. Academic Press, London.

HAERTY, W. and SINGH, R.S. 2006. Gene regulation divergence is a major contributor to the evolution of Dobzhansky-Muller incompatibilities between species of Drosophila. Mol. Biol. Evol. **23**: 1707-1714.

HAERTY, W., JAGADEESHAN, S., KULATHINAL, R.J., WONG, A., RAVI RAM, K., SIROT, L.K., LEVESQUE, L., ARTERI, 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**: 1321-1335.

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

HEWITT, G.M., MASON, P., and NICHOLS, R.A. 1989. Sperm precedence and homogamy across a hybrid zone in the alpine grasshopper *Podisma pedestris*. Heredity **62**: 343-353.

HOSKEN, D.J. 2001. Sex and death: microevolutionary trade-offs between reproductive and immune investment in dung flies. Curr. Biol. **11**: R379-R380.

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

HOWARD, D. J., GREGORY, P., CHU, J., and CAIN, M. 1998. Conspecific sperm precedence in an effective barrier to hybridization between closely related species. Evolution. **52**: 511-516.

JAENIKE, J. 2001. Sex chromosome meiotic drive. Ann. Rev. Ecol. Syst. **32**: 25-49.

JAGADEESHAN, S. and SINGH, R.S. 2006. A time-sequence functional analysis of mating behaviour and genital coupling in Drosophila: role of cryptic female choice and male sex-drive in the evolution of male genitalia. J. Evol. Biol. **19**: 1058-1070.

JONES, B. and CLARK, A.G. 2003. Bayesian sperm competition estimates. Genetics **163**: 1193-1199.

KLIMAN, R.M. and HEY, J. 1993. Reduced natural selection associated with low recombination in Drosophila melanogaster. Mol Biol Evol. **10**: 1239-1258

KLIMAN, R.M., ANDOLFATTO, P., COYNE, J.A., DEPAULIS, F., KREITMAN, M., BERRY, A.J., MCCARTER, J., WAKELEY, J., and HEY, J. 2000. The population genetics of the origin and divergence of the Drosophila simulans complex species. Genetics. **156**: 1913-1931

KULATHINAL, R.J. and SINGH, R.S. 2004. The nature of genetic variation in sex and reproduction-related genes among sibling species of the Drosophila melanogaster complex. Genetica. **120**: 245- 252

KUMAR S, TAMURA K, JAKOBSEN IB, NEI M. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics. **17**: 1244-1245

LIIMATAINEN, J. and HOIKKALA. A. 1998. Interactions of the males and females of three sympatric *Drosophila virilis* group species, *D. montana, D. littoralis* and *D. lummei,* in intra- and interspecific courtships in the wild and in the laboratory. J. Insect Behav. **11**: 399–417.

LIVAK, K.J. and SCHMITTGEN, T.D. 2001 Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}{}_{T}$ method. Methods **25**: 402-408.

LUDLOW, A.M. and MAGURRAN, A.E. 2006. Gametic isolation in guppies (Poecilia reticulata). Proc Biol Sci. **273**: 2477-2482

MALICKI J, CIANETTI LC, PESCHLE C, MCGINNIS W. 1992 A human HOX4B regulatory element provides head-specific expression in Drosophila embryos. Nature. **358**: 345-347.

MAYR, E. 1942. *Systematics and the Origin of Species*. Columbia University Press, New York.

MAYR, E. 1963. Animal Species and Evolution. Cambridge, MA: Belknap Press.

MCKEAN, K. A. and NUNNEY, L. 2001. Increased sexual activity reduces male immune function in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **98**: 7904-7909.

MCLACHLAN, R.I., MALLIDIS, C., MA, K., BHASIN, S., and DE KRETSER, D.M. 1998. Genetic disorders and spermatogenesis. Reprod Fertil Dev. 10: 97-104.

MILLER, G.T. and PITNICK, S. 2002. Sperm-female coevolution in *Drosophila*. Science **298**: 1230-1233.

MUELLER, J.L., LINKLATER, J.R., RAVI RAM, K., CHAPMAN, T., and WOLFNER, M.F. 2008. Targeted gene deletion and phenotypic analysis of the *Drosophila melanogaster* seminal fluid protease inhibitor Acp62F. Genetics **178**: 1605-1614.

NICKEL, D. and CIVETTA, A. 2009. An X chromosome effect responsible for asymmetric reproductive isolation between male Drosophila virilis and heterospecific females. Genome **52**: 49-56.

NILSSON, T., FRICKE, C., and ARNQVIST, G. 2003. The effects of male and female genotype on variance in male fertilization success in the red flour beetle (*Tribolium castaneum*). Behav. Ecol. Sociobiol. **53**: 227-233.

ORR, H.A. and COYNE, J.A. 1992. The genetics of adaptation: a reassessment. Am Nat. **140**: 725-742.

ORR, H.A. 1998. Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. Genetics. **149**: 2099-2104.

PARKER, G.A. and PARTRIDGE, L. 1998. Sexual conflict and speciation. Philos Trans R Soc Lond B Biol Sci. **353**: 261-274.

PATTERSON, J.T. 1946. A New Type of Isolating Mechanism in Drosophila. Proc Natl Acad Sci U S A. **32**: 202-208.

PEREZ, D.E., WU, C.I., JOHNSON, N.A., and WU, M.L. 1993. Genetics of reproductive isolation in the *Drosophila simulans* clade: DNA marker-assisted mapping and characterization of a hybrid-male sterility gene, *Odysseus* (*Ods*). Genetics **134**: 261-275.

PEREZ, D.E. and WU, C.I. 1995. Further characterization of the *Odysseus* locus of hybrid sterility in Drosophila: one gene is not enough. Genetics **140**: 201-206.

PITNICK, S., MARKOW, T., and SPICER, G.S. 1999. Evolution of multiple kinds of female sperm-storage organs in *Drosophila*. Evolution **53**: 1804-1822.

PITNICK, S., MILLER, G.T., SCHNEIDER, K., and MARKOW, T.A. 2003. Ejaculate-female coevolution in *Drosophila mojavensis*. Proc. R. Soc. London B **270**: 1507-1512.

POONGOTHAI, J., GOPENATH, T.S., and MANONAYAKI, S. 2009. Genetics of human male infertility. Singapore Med J. **50**: 336-347.

PRICE, C.S. 1997. Conspecific sperm precedence in Drosophila. Nature **388**: 663-666.

PRICE, C.S., KIM, C.H., POSLUSZNY, J., and COYNE, J.A. 2000. Mechanisms of conspecific sperm precedence in Drosophila. Evolution **54**: 2028-2037.

PRICE, C.S., KIM, C.H., GRONLUND, C.J., and COYNE, J.A. 2001. Cryptic reproductive isolation in the Drosophila simulans species complex. Evolution **55**: 81-92.

PRICE, T.A., and WEDELL, N. 2008. Selfish genetic elements and sexual selection: their impact on male fertility. Genetica **134**: 99-111.

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

RIESEBERG, L.H., DESROCHERS, A.M., and YOUN, S.J. 1995. Interspecific pollen competition as a reproductive barrier between sympatric species of helianthus (asteraceae). Am. J. Botany **82**: 515-519.

ROBINSON, T., JOHNSON, N.A., and WADE, M.J. 1994. Postcopulatory, prezygotic isolation: intraspecific and interspecific sperm precedence in Tribolium spp., flour beetles. Heredity **73**: 155-159.

ROZEN, S. and SKALETSKY, H. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. **132**: 365-386.

SAARIKETTU M, LIIMATAINEN JO, HOIKKALA A. 2005. The role of male courtship song in species recognition in *Drosophila montana*. Behav Genet. **35**: 257-263.

SIMMONS, L. 2001. *Sperm Competition and Its Evolutionary Consequences in the Insects*. Princeton University Press, USA.

SINGH, S.R., SINGH, B.N. & HOENIGSBERG, H.F. 2002. Female remating, sperm competition and sexual selection in *Drosophila*. Genetics and Molecular Research. **1**: 178-215.

SNOOK, R.R. 1998. The risk of sperm competition and the evolution of sperm heteromorphism. Animal Behaviour. **56**: 1497-1507.

SNOOK, R.R. and MARKOW, T.A. 2002. Efficiency of gamete usage in nature: sperm storage, fertilization and polyspermy. Proc Biol Sci. **269**: 467-473.

SOKAL, R.R. and ROHLF, F.J. 1995. *Biometry 3rd edition*. W.H. Freeman and Company, New York, NY

STEBBINGS, L., GRIMES, B.R., and BOWNES, M. 1998. A testis-specifically expressed gene is embedded within a cluster of maternally expressed genes at 89B in Drosophila melanogaster. Dev Genes Evol. **208**: 523-530.

STEBBINS, G.L. 1950. *Variation and Evolution in Plants*. Columbia University Press, New York

STROSCHEIN-STEVENSON, S. L., FOLEY, E., O'FARRELL, P.H., and JOHNSON, A.D. 2006. Identification of Drosophila gene products required for phagocytosis of *Candida albicans*. PLoS Biol. **4**: e4.

STUART, L.M., BOULAIS, J., CHARRIERE, G.M., HENNESSY, E.J., BRUNET, S., JUTRAS, I., GOYETTE, G., RONDEAU, C., LEARTE. S., HUANG, H., YE, P., MORALES, F., KOCKS, C., BADER, J.S., DESJARDINS, M., EZEKOWITZ, R.A. 2007. A systems biology analysis of the Drosophila phagosome. Nature **445**: 95-101.

SUAREZ, S.S. 2002. Formation of a reservoir of sperm in the oviduct. Reprod Domest Anim. **37**: 140-143.

SULLIVAN, W., ASHBURNER, M. & HAWLEY, S. 2000. Drosophila Protocols. Cold Springs Harbour Laboratory Press, New York. Pg 637.

SUN, S., TING, C.T., and WU, C.I. 2004. The normal function of a speciation gene, Odysseus, and its hybrid sterility effect. Science **305**: 81-83.

SWANSON, W.J., CLARK, A.G., WALDRIP-DAIL, H.M., WOLFNER, M.F. and AQUADRO, C.F. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proceedings of the National Academy of Science. **98**: 7375-7379.

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

TAO, Y., CHEN, S., HARTL, D. L., and LAURIE, C. C., 2003b. Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. I. Differential accumulation of hybrid male sterility effects on the X and autosomes. Genetics **164**: 383-397.

TAO, Y., ZENG, Z.B., LI, J., HARTL, D.L., and LAURIE, C.C. 2003. Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. II. Mapping hybrid male sterility loci on the third chromosome. Genetics. **164**: 1399-1418.

THOMAS, S. and SINGH, R.S. 1992. A comprehensive study of genic variation in natural populations of Drosophila melanogaster. VII. Varying rates of genic divergence as revealed by two-dimensional electrophoresis. Mol Biol Evol. **9**: 507-525.

THOMPSON, J.D., HIGGINS, D.G., and GIBSON, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**: 4673-4680.

VAGIN, V.V., SIGOVA, A., LI, C., SEITZ, H., GVOZDEV, V., and ZAMORE, P.D. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. Science **313**: 320-324.

WADE, M.J., JOHNSON, N.A., and WARDLE, G. 1994. Analysis of autosomal polygenic variation for the expression of Haldane's rule in flour beetles. Genetics. **138**: 791-799

WIGBY, S., SIROT, L.K., LINKLATER, J.R., BUEHNER, N., CALBOLI, F.C., BRETMAN, A., WOLFNER, M.F., and CHAPMAN, T. 2009. Seminal fluid protein allocation and male reproductive success. Curr. Biol. **19**: 751-757.

WOLFNER, M.F. 2007. "S.P.E.R.M." (seminal proteins (are) essential reproductive modulators): the view from Drosophila. Soc. Reprod. Fertil. Suppl. **65**: 183-199.

WOLFNER, M.F. 2009. Battle and ballet: molecular interactions between the sexes in Drosophila. J. Hered. **100**: 399-410.

WONG, A., ALBRIGHT, S.N., GIEBEL, J.D., RAM, K.R., JI, S., FIUMERA, A.C., and WOLFNER, M.F. 2008. A role for Acp29AB, a predicted seminal fluid lectin, in female sperm storage in Drosophila melanogaster. Genetics. **180**: 921-931.

WU, C.I., and PALPOLI, M.F. 1994 Genetics of postmating reproductive isolation in animals. Annu. Rev. Genet. **28**: 283-308.

XIA, X. and XIE, Z. 2001. DAMBE: software package for data analysis in molecular biology and evolution. J Hered. **92**: 371-373.

XU, E.Y., LEE, D.F., KLEBES, A., TUREK, P.J., KORNBERG, T.B., and REIJO PERA, R.A. 2003. Human BOULE gene rescues meiotic defects in infertile flies. Hum. Mol. Genet. **12**: 169-175

YANG, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13:** 555-556.

YANG, Z. and NIELSEN, R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. **19**: 908-917.

ZHANG, J., NIELSEN, R., and YANG, Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. **22**: 2472-2479.

Supplementary Table

Table 1: Test of orthology in *D. simulans* and *D. mauritiana*. Nucleotide alignments between *D. simulans* (Dsim) *D. sechellia* (Dsec) and *D. mauritiana* (Dmau) are shown. Dmau and DsimL are sequences generated in our lab. Stop codons due to indels (see carat) or nucleotide changes are bolded and underlined. CG34357 spans approximately 64Kb (Flybase) and was not sequenced.

CG5178

```
Dsim          ATGTGTGACGATGATGCGGGTGCATTAGTTATCGACAACGGATCGGGCATGTGCAAAGCC
Dmau          ------------------------------------------GGCATGTGCAAAGCC
Dsec          ATGTGTGACGATGATGCGGGTGCATTAGTTATCGACAACGGATCGGGCATGTGCAAAGCC

Dsim          GGCTTCGCCGGTGATGACGCTCCCCGTGCTGTCTTCCCCTCGATTGTGGGTCGTCCCCGT
Dmau          GGCTTCGCCGGTGATGACGCTCCCCGTGCTGTCTTCCCCTCGATTGTGGGTCGTCCCCGT
Dsec          GGCTTCGCCGGTGATGACGCTCCCCGTGCTGTCTTCCCCTCGATTGTGGGTCGTCCCCGT

Dsim          CACCAGGGTGTGATGGTGGGCATGGGTCAGAAGGACTCGTACGTGGGCGACGAGGCGCAG
Dmau          CACCAGGGTGTGATGGTGGGCATGGGTCAGAAGGAATCGTACGTGGGCGACGAGGCGCAG
Dsec          CACCAGGGTGTGATGGTGGGCATGGGTCAGAAGGACTCGTACGTGGGCGACGAGGCGCAG

Dsim          AGCAAGCGCGGTATCCTGACGCTGAAGTACCCCATCGAGCACGGCATCATCACGAACTGG
Dmau          AGCAAGCGCGGTATCCTGACGCTGAAGTACCCCATCGAGCACGGCATCATCACGAACTGG
Dsec          AGCAAGCGCGGTATCCTGACGCTGAAGTACCCCATCGAGCACGGTATCATCACGAACTGG

Dsim          GACGACATGGAGAAGATCTGGCATCACACCTTCTACAACGAGCTGCGCGTGGCCCCCGAG
Dmau          GACGACATGGAGAAGATCTGGCATCACACCTTCTACAACGAGCTGCGCGTGGCCCCCGAG
Dsec          GACGATATGGAGAAGATCTGGCATCACACCTTTTACAACGAGCTGCGCGTGGCCCCCGAG

Dsim          GAGCATCCAGTATTATTGACCGAGGCACCCCTGAACCCCAAGGCCAATCGCGAGAAGATG
Dmau          GAGCATCCAGTATTATTGACCGAGGCACCCCTGAACCCCAAGGCCAATCGCGAGAAGATG
Dsec          GAGCATCCAGTATTATTGACCGAGGCACCCCTGAACCCCAAGGCCAATCGCGAGAAGATG

Dsim          ACCCAGATCATGTTCGAGACCTTCAACTCGCCGGCCATGTACGTGGCCATCCAGGCCGTG
Dmau          ACCCAGATCATGTTCGAGACCTTCAACTCGCCGGCCATGTACGTGGCCATCCAGGCCGTG
Dsec          ACCCAGATCATGTTCGAGACCTTCAACTCGCCGGCCATGTACGTGGCCATCCAGGCCGTG

Dsim          CTCTCCC-TGTACGCCTCTGGTCGTACCACCGGTATTGTGCTGGACTCCGGCGATGGTGT
Dmau          CTCTCCC-TGTACGCCTCCGGTCGTACCACCGGTATTGTGCTGGACTCCGGCGATGGTGT
Dsec          CTCTCCCCTGTACGCCTCCGGTCGTACCACCGGTATTGTGCTGGACTCCGGCGATGGTGT
                     ^

Dsim          CTCCCACACCGTACCCATCTATGAGGGCTTCGCCCTGCCCCACGCCATCCTGCGTCTGGA
Dmau          CTCCCACACCGTGCCCATCTATGAGGGCTTCGCCCTGCCCCACGCCATCCTGCGTCTGGA
Dsec          CTCCCACACCGTGCCCATCTA**TGA**GGGCTTCGCCCTGCCCCACGCCATCCTTCGTCTGGA

Dsim          TCTGGCTGGTCGCGATCTGACCGATTACCTGATGAAGATCCTGACGGAGCGCGGCTACAG
Dmau          TCTGGCTGGTCGCGATCTGACCGATTACCTGATGAAGATCCTGACGGAGCGCGGCTACAG
Dsec          TCTGGCTGGTCGCGATCTGACCGATTACCTGATGAAGATCCTGACGGAGCGCGGCTACAG

Dsim          CTTCACCACCACCGCCGAGCGTGAGATCGTGCGCGACATCAAGGAGAAGCTGTGCTACGT
Dmau          CTTCACCACCACCGCCGAGCGTGAGATCGTGCGCGACATTAAGGAGAAGCTGTGCTACGT
Dsec          CTTCACCACCACCGCCGAGCGTGAGATCGTGCGCGACATCAAGGAGAAGCTGTGCTACGT

Dsim          GGCTCTGGACTTCGAGCAGGAGATGGCCACCGCTGCCGCCTCCACCTCGCTGGAGAAGTC
Dmau          GGCTCTGGACTTCGAGCAGGAGATGGCCACCGCTGCCGCCTCCACCTCGCTGGAGAAGTC
Dsec          GGCTCTGGACTTCGAGCAGGAGATGGCCACCGCTGCCGCTTCCACCTCGCTGGAGAAGTC
```

```
Dsim          GTACGAGTTGCCCGACGGCCAGGTGATCACCATTGGCAACGAGCGCTTCCGCTGCCCCGA
Dmau          GTACGAGTTGCCCGACGGCCAGGTGATCACCATTGGCAACGAGCGCTTCCGCTGCCCCGA
Dsec          GTACGAGTTGCCCGACGGCCAGGTGATCACCATTGGCAACGAGCGCTTCCGCTGCCCCGA

Dsim          AGCCCTGTTCCAGCCCTCGTTCCTGGGCATGGAGTCGTGCGGCATCCACGAGACCGTCTA
Dmau          GGCCCTGTTCCAGCCCTCGTTCCTGGGCATGGAGTCGTGCGGCATCCACGAGACCGTCTA
Dsec          GGCCCTGTTCCAGCCCTCGTTCCTGGGCATGGAGTCGTGCGGCATCCACGAGACCGTCTA

Dsim          CAACTCGATCATGAAGTGCGACGTGGACATCCGCAAGGATCTGTATGCCAACTCCGTGCT
Dmau          CAACTCGATCATGAAGTGCGACGTGGACATCCGCAAGGATCTGTATGCCAACTCCGTGTT
Dsec          CAACTCGATCATGAAGTGCGACGTGGACATCCGCAAGGATCTGTATGCCAACTCCGTGCT

Dsim          GTCCGGCGGTACCACCATGTACCCTGGTATTGCCGATCGTATGCAAAAGGAGATCACTGC
Dmau          GTCCGGCGGTACCACCANGTACCCTGGT--------------------------------
Dsec          GTCCGGCGGTACCACCATGTACCCTGGTATTGCCGATCGTATGCAGAAGGAGATCACTGC
```

CG9391

```
Dsim          ATGTCACACAGCGTGGACGTGGAAAAGTGCTTAGAGGTGGCCAGCAACCTGGTTTCAGAA
Dmau          ATGTCACANNNCGTGGACGTGGAAAAGTGCTTAGAGGTGGCCAGCAACCTGGTTTCAGAA
Dsec          ATGTCACACAGCGTGGACGTGGAGAAGTGCTTAGAGGTGGCCAGCAACCTGGTTTCAGAA

Dsim          GCTGGAAGGCTCATCGCTCGCAACAATGAGCAGCGACAGGACTTCGTTTGCAAGAGCAAT
Dmau          GCTGGAAGGCTCATCGCTCGCAACAATGAGCAGCGACAGGACTTCGTTTGCAAGAGCAAT
Dsec          GCTGGAAGGCTCATCGCTCGCAACAATGAGCAGCGACAGGACTTCGTTTGCAAGAGCAAT

Dsim          GACATCGACTTGGTGACCCAAACAGACAAGGATGTGGAGCAGCTACTGATGGACGGCATT
Dmau          GACATCGACTTGGTGACCCAAACAGACAAGGATGTGGAGCAGCTACTGATGGACGGCATT
Dsec          GACATCGACTTGGTGACCCAAACAGACAAGGATGTGGAGCAGCTACTGATGGACGGCATT

Dsim          CGCCGCCACTTTCCGGAGCACAAGTTCATCGGCGAGGAGGAGAGTAGCGGCGAGGAGGGT
Dmau          CGCCGCCACTTTCCGGAGCACAAGTTCATCGGCGAGGAGGAGAGTAGCGGCGGGGAGGGT
Dsec          CGCCGCCACTTTCCGGAGCACAAGTTCATCGGCGAGGAGGAGAGTAGCGGCGAGGAGGGT

Dsim          GTCAAGAAGCTTACCGACGA-GCCCACCTGGATCATTGATCCCGTGGACGGCACCATGAA
Dmau          GTCAAGAAGCTTACCGACGA-GCCCACCTGGATCATTGATCCCGTGGACGGCACCATGAA
Dsec          GTCAAGAAGCTTACCGACGAAGCCCACCTGAATCAT<u>**TGA**</u>TCCCGTGGACGGCACCATGAA
                                 ^

Dsim          CTTTGTGCACGCATTTCCGCACTCTTGCATCT-CCGTGGGTCTGAAGGTGAACAAGGTCA
Dmau          CTTTGTGCACGCATTTCCGCACTCTTGCATCT-CCGTGGGTCTGAAGGTGAACAAGGTCA
Dsec          CTTTGAGCACGCATTTCCGCACTCTTGCATCTTCCGTGGGTCTGAAGGTGAACAAGGTCA
                                            ^

Dsim          CGGAGCTGGGCTTGGTCTACAATCCCATCCTGGAGCAGCGCTTCACTGCGCGACGTGGGC
Dmau          CGGAGCTGGGCTTGGTATACAATCCCATCCTGGAGCAGCGCTTCACTGCGCGACGTGGGC
Dsec          CGGAGCTGGGCTTTGTCTACAATCCCATCCTGGAGCAGCGCTTCACTGCGCGACGAGGGC

Dsim          ACGGAGCCTTCTACAACGGGCGCAGGATCCACGTGAGCGGCCAAAAGGAACTGGG-CAAA
Dmau          ACGGAGCCTTCTACAACGGGCGCAGGATCCACGTGAGCGGCCAAAAGGAACTGGG-CAAA
Dsec          ACGGAGCCTTCTACAACGGGCGCAGGATCCACGTGAGCGGCCAAAAGGAACTGGGGCAAA
                                                                      ^

Dsim          GCGCTGGTCACCAGTGAATTCGGTACCACCCGGGACGAGGCCAAGATGAAGGTCGTGCAT
Dmau          GCGCTGGTCACCAGTGAATTCGGTACCACCCGGGACGAGGCCAAGATGAAGGTCGTGCAT
Dsec          GCGCTGGTCACCAGTGAATTCGGTACCACCCGGGACGAGGCCAAGATGAAGGTCGTGCAT

Dsim          GAGAACTTCGAGAAGATGGCCAAAAAGGCGCATGGCCTACGGGTCCTGGGTTCGGCAGCC
Dmau          GAGAACTTCGAGAAGATGGCCAAAAAGGCGCATGGCCTACGGGTCCTGGGTTCGGCAGCC
Dsec          GAGAACTTCGAGAAGATGGCCAAAAAGGCGCATGGCCTACGGGTCCTGGGTTCGGCAGCC

Dsim          CTTAATATGTCGATGGTTGCTCTGGGAGCCGCTGACGCCAACTACGAATTTGGAATTCAC
Dmau          CTTAATATGTCGATGGTTGCTCTGGGAGCCGCTGACGCCAACTACGAATTTGGAATTCAC
Dsec          CTTAATATGTCGATGGTTGCTCTGGGAGCCGCTGATGCCAACTACGAATTTGGAATTCAC
```

```
Dsim          GCCTGGGATGTGTGTGCCGGCGACTTGATTGTCCGGGAGGCTGGTGGCGTAGTCATCGAT
Dmau          GCCTGGGATGTGTGTGCCGGCGACTTGATTGTCCGGGAGGCTGGTGGCGTGGTCATCGAT
Dsec          GCCTGGGATGTGTGTGCCGGCGACTTGATTGTCCGGGAGGCTGGTGGCGTGGTCATCGAT

Dsim          CCTGCTGGCGGCGAATTCGACATCATGTCTCGAAGGGTCCTGGCGGCAGCCACACCAGAG
Dmau          CCTGCTGGCGGCGAATTCGACATCATGTCTCGTAGGGTCCNGGCGGCN------------
Dsec          CCTGCTGGCGGCGAATTCGACATCATGTCTCGAAGGGTCCTGGCGGCAGCCACACCAGAG
```

CG14891

```
Dsim          ATGTTCCGCTTGAAGGGAAATTTTGGAAATTTTTACTTTAAAAATGGTTTGGTGTACACC
Dmau          ------------------------------------------------------------
Dsec          ATGTTCCGCTTGGGGGAAAATTTTGGAAATTTTTACTTTGAAAATGGTTTGGCGTACACC

Dsim          AAAGATCGCAAGGTTGTGCGACTGGTGACGATCTCTGCGAACTGGCACTTTTCGAAGAGC
Dmau          ----------AAGGTTGTGCGACTGGTGACGATCTCTGCGAACTGGCACTTTTCGAAGAGC
Dsec          GAAGATCGCAAGGTTGTGCGACTGGTGACGATCTCTGCGAAGTGGCACTTTTCGAAGAGC

Dsim          CAACTCTGGAAGCACTTCTCGAGTTTTGGAACTGTGGAGGATCTCCAATGGGAAAA-GGA
Dmau          CAACTCTGGAAGCACTTCTCGAGTTTTGGAACTGTGGAGGATCTCCAATGGAAAAA-GAA
Dsec          CAACTCTGGAAGCACTTCTCGAGTTTTGGAACTGTGGAGGATCTCCAA TGA AAAAAAGAA
                                                              ^

Dsim          TAAGAGAGTGGGATCGGTTCTTTTTCAAGAGGCTTCCCAAGCGGCAAAGGTTTTGGTGTT
Dmau          TAAGAGAGTGGGATCGGTTCTTTTTCAAGAGGCTTCCCAAGCGGCAAGGGGTTTGGTGTT
Dsec          TAA GAGAGTGGGATCGGTTCTTTTTCAAGAGGCTTCCCAAGCGGCAAGGGTTTTGGTGTT

Dsim          GACTAAACACCATTTGTATGGCCATGTTCTTTATTTGCAGCCCAGCACCTCCAGGCGCGA
Dmau          GGCTAAACACTATTTGTATGGCCATGTTCTATATTTGCAGCCCAGAACCTCCGGGCACGA
Dsec          GACTAAACACTATTTGTATGGCCATGTTCTTTATTTGCAGCCCAGCACCTCCAGGCGCGA

Dsim          ACCGCCGGTGAAGGAATCAGAAACTATTTCTGCCTACGATATACCCGTTGTCGATGACTT
Dmau          ACCGCCGGTGAAGGAATCAGGAACTCTTTCTGCCTACGATATACCCGTTGTCGATGACGT
Dsec          ACCGCCGGTGAAGGAATCAGAAACTATTTCTGCCTACGATACACCCGTTGTCGATGACGT

Dsim          TTGGTATAAAGTGCTCGAATATCTTCCACTAAATGCCCGTCTCAACTTTGCCGCCAGTTG
Dmau          TTGGTATAAAGTGCTCGAATATCTTCCACTAAATTCCCGTCTCAACTTTTCCGCCAGTTG
Dsec          TTGGTATAAAGTGCTCGAATATCTTCCACTAAAATCCCGTCTCAACTTTGCCGCCAGTTG

Dsim          TAAAAGATTCAAGACGATCTACGAATTGGAGTCGCGTCGTAACAATCGTGTTCTTAATAT
Dmau          TAAAAGATTCAAGACGATCTACGAATTGGAGTCGCATCGTAACAATCGTGTTCTTAATAT
Dsec          TAAAAGATTCAAGACAATCTACGAATTGGAGTCGCATCGTAACAATCGTGTTCTTAATAT

Dsim          GAAGGATGTTTGCACACTGGACGACTTTGGCATTAAAATATTGATGCGGCTATCAGGAAA
Dmau          GAAT---ATTTGCACACTGGACGACTTTGGCATTAAAATATTGATGCGGCTATCAGGAAA
Dsec          GAGGGATGTTTGCACACTGGACGACT-------------------GCGGCTATCCGGAAA

Dsim          ACACATTCATTGTGTAAAAGGTGGCCCGCTTCATTGGACGCTTATGTTGGAGTTCGTGCA
Dmau          ACACATTCATTGTGTAAAAGGTGGCCCGCTTCATCGGCCGCTTATGTCGGAGTTCGTGCA
Dsec          ACACATTCATTGTGTAAAAGGTGGCCCGCTTCATCGGCCGCTTATGTCGGAGTTCGTGCA

Dsim          GCTATTGGGTGTAAGCTGTCCAAATCTAGCAGAGCTAAGTTTCTACAAAA--------TT
Dmau          GCTATTGGGTGTAAGCTGTCCAAATCTAGCAGAGCTAAGTTTCTACAATA--------CT
Dsec          GCTATTGGGTGTAAGCTGTCCAAATCTAGCAGAGCTAAGTTTCTACAATAGAGCTAAGTT

Dsim          TCAGTCAGCCTAGACCACATGACTCACCTGTTCGATGGTGCCAATGGCTTGAATAATATC
Dmau          TCAGTCAAC---------ATGACTCACCTGTTCGATGGTGCCAATGGCTTGAATAATATC
Dsec          TCAGTCAGCCTAGACCACATGACTCACCTGTTTGATGGTGCCAATGGCTTAAATAATATC

Dsim          ACCACCATATCCTTGAGGTGTTGTGACTTGGCAGATCCTCAAATTTACTGCTTGCAGATG
Dmau          ACCACCATATCCTTGAGGTGTTGTGACTTGGCAGATACTCAAATTTACTGCTTGCAGATG
Dsec          ACCACCATATCCTTGAGGTGTTGTGACTTGGCAGATACTCACATTTACTGCTTGCAGATG
```

```
Dsim          CTATCTAAACTAAAGAGTCTGGACATCGCACAGAACCATTTCATTAGGGGCGAAAGTTTA
Dmau          CTATCTAAACTAAAGAGTCTGGACATCGCACAGAACCATTTCATTAGGGGCGAAAGTTTA
Dsec          CTATCTAAACTAAAGAGTCTGGACATCGTACAGAACTATTTCATTAAGGGCGAAAGTTTA

Dsim          AACTCTCTGCCAATTTCCTTGGAGATTTTAAATGTTTCAAAATGCGACAGACTGCGGCCC
Dmau          AACTCTCTGCCAATTTCCTTGGAGATTTTAAATGTTTCAAAATGCGATAGTCTGTTGCCC
Dsec          AAATCTCTGCCAATTTCCTTGGAGATTTTAAATGTTTCAAAATGCGACAGACTGCTGCCC

Dsim          AAGAATCTTATCAATCTTGCGTCCCTGACGCATCTCCGCGAACTGCGCTGCTCTGGCATT
Dmau          AAGCATCTTATCAATCTTGCGTCCCTGTCGCATCTCCGCGAACAG---------------
Dsec          AAGCATCTTATCAATCTTGCGTCCCTGTCGCATCTCCGCGAACTGCGCTGATCTGGCACT

Dsim          TCCAAGCTTACGAAAAATGAGCTGTTCAAACGGTTCGCACATTACTGTCCAATGCTCGAG
Dmau          ------------------------TACAAACGGTTCGCACATTACAGTCCAATGCTCGAG
Dsec          TCCAAGTGTGCGAAAAATGAGCTGTACAAACGGTTCGCACATTACTGTCCAATGCTCGAG

Dsim          GTTCTCGAGGTTACCGACATTATGAAGAAGATACAGCTGGGCGGTCTGTCTCGTCTCCAC
Dmau          GTT---------ACCGACACTATGAAGAAGATACAGCTGGGCGGTCTGTCTCGTCTCCAC
Dsec          GTTCTCGAGGTTACCGACATTATGCAGAAGATACAGCTGGGCGGTCTGTCTCGTCTCCAC


Dsim          ACCTTGGTCATTCAGTCTTCCGAAGGGTCTGGCGACCATATGAATAACTTGATGCTTTCG
Dmau          ACCTTGGTCATTCAGTCTTCCGAAGGGTTTGGCGACCATATGAATAACTTGCTGCTTACG
Dsec          ACCTTGGTCATTCCGTCTGCCGAAGGGTTTGGCGACCATATGAATAACTTGCTGCTTACG

Dsim          TCGATCGCGGAATCGTATTCGCTGCGCCGTCTGGAGATTATAGATTCTTTTGAACGTTTT
Dmau          TCGATCGCGGAATCGTATTCGCTGCGCCATCTGGAGATTATAGATTCTTTTGAACGTTTT
Dsec          TCGATCGCGGAATCGTATTCGCTGCGCCATCTGGAGATTATAGATTCTT-----------

Dsim          TTCACTATTTCCTTCGATCTGAGTATTTTATCCCCGCTTAAAGAACTGCGGACCCTAATA
Dmau          TTAGCTATTTCCTTCGATCTGAGTATTTTATCCCCGCCTAAAGAACTGCGGACCCTAATA
Dsec          ----------------------------AACTTAACTTAAAGAACTGCGGACCCTAATA

Dsim          TTACATAATCTGAACTTTACACCGGAACACCTAATGGGATTGCAAAAACTCCCTGCCTTG
Dmau          TTACATAATCTGAACTTTACACCGGTACACCTAATGGGATTGCAAAAACTCACTGCCTTG
Dsec          TTACATAATCTAAACTTTACACCGGTACACCTAATGGGATTGCAAAAATTCACTGCCTTG

Dsim          GAGTTTCTGGACCTGAGTGGCTCGCCCGATCTATCCAATGAGGACGTTGCAAAGTTGACG
Dmau          GAGTTTCTGGACCTGACTGGCTNGCCCGATCTATCCAATGAGGACGTTGCGAAGTTGACG
Dsec          GAGTTTCTGGACCTGACTGGCTCGCGCGATCTATCCAATGAGAACGTTGCGAAGTTGACG

Dsim          AAACCGCTGGGCAGACTGCGCCGACTAACGGTTGAGCGTTGTCCTTTTATCTCACGACAA
Dmau          AAGCCGCTGGGCAGACTG-------------------------------------------
Dsec          AAACCGCTTGGCAGACT-----------------------------------CACGACAA
```

CG9063

```
Dsim           TGGACAGTGTTGCAGCTGCCGCTCAACTACGCGGCGACAAACTGGCCAATCCGGTATGCT
Dsec           TGGACAGTGTTGCAGCTGCCGCTCAACTACGCGGCGACAAACTGGCCAATCCGGTATGCT
DsimL          --GACAGTGTTGCAGCTGCCGCTCAACTACGCGGCGACAAACTGGCCAATCCGGTATGCT

Dsim           GCTATTGATCCGGATGGACTCCACTTGGCGGTGGCTGGTCGCACTGGGCTGGCGCACTAT
Dsec           GCTATTGATCCGGATGGACTCCACTTGGCGGTGGCTGGTCGCACTGGGCTGGCGCACTAT
DsimL          GCTATTGATCCGGATGGACTCCACTTGGCGGTGGCTGGTCGCACTGGGCTGGCGCACTAT

Dsim           TCCCTAGTGACC-GGCGCTGGAAGCTTTT-GGCAA<u>**TGA**</u>GTCGCAGGAGAAGGACTTTGTT
Dsec           TCCCTAGTGACCCGGCGCTGGAAGCTTTTTGGCAATGAGTCGCAGGAGAAGGACTTCGTT
DsimL          TCCCTAGTGACCCGGCGCTGGAAGCTTTTTGGCAATGAGTCGCAGGAGAAGGACT-----
                                           ^
```

CG1041

```
Dmau          ------------------------------------------------------------
Dsec          ---------------ATGTACAAATATCAAAGCCCAACGTCTGAGCGAATCCTCAAGAAA
Dsim          ATGTTAATTAACCGGAGCAAAACAAGAGTGAGCCCAACGTCTGAGCGAATCCTCGAGAAA

Dmau          ------------------------------------------------------------
Dsec          CCCGACGCCGAGATGAAGTTTCGTGGCAATGGAAAGCTTTTGTGGAATTTGACCAAGAAC
Dsim          CCCGACGCCGAGATGAAGTTCCGTGGCAATGGAAAGCTTTTGTGGAATTTGACCAAGAAC

Dmau          ------------------------------------------------------------
Dsec          TCCTTGGCCCAACAGTCTCCAAATGGGATCGCCAAGAAAGTGCTGCCCGCCAGCAGCTAC
Dsim          TCCTTGGCCCAGCAGTCTCCAAATGGAATCGCCAAGAAAGTGCTGCCCGCCAGCAGCTAC

Dmau          ------------------------------------CTGCTGAAGTACCACGTCCTG
Dsec          AGCACCGTCCAGAAGACCATTCCCTTGGAGCAGCCGAATCTGCTGAAGTACCACGTCCTG
Dsim          AGCACCGTCCAGAAGACCATTCCCTTGGAGCAGCCGAATCTGCTGAAGTACCACGTCCTG

Dmau          CCGCTGGAGGAAACGCTAAACCGCTTCATGACCACGGTGGAACCTCTGCTGACGCCGGAG
Dsec          CCGCTGGAGGAAACGCTGAACCGCTTCATGACCACGGTGGAACCGCTGCTGACGCCGGAG
Dsim          CCGCTGGAGGAAACGCTGAACCGCTTCATGACCACGGTGGAACCTCTGCTGACGCCGGAG

Dmau          GAGTTTCAACAGCAAAAGGGAATCACCTCCGAGTTTTTGAAGAAGCAGGGACGCGAACTG
Dsec          GAGTTTCAACAGCAAAAGGGAATCACCTCCGAGTTTTTGAAGAAGCAGGGACGCGAACTG
Dsim          GAGTTTCAACAACAAAAGGGAATCACCTCCGAGTTTTTGAAGAAGCAGGGACGCGAACTG

Dmau          CAGCTGCTCCTGGAAGAAACCGGCAGCAAGGAGAAGAATTGGCTGGCCCACCGCTGGCTG
Dsec          CAGCTGCTCCTGGAAGAAACCGGCAGCAAGGAGAAGAATTGGCTGGCCCACCGCTGGCTG
Dsim          CAGCTGCTCCTGGAAGAAACCGGCAGCAAGGAGAAGAATTGGCTGGCCCACCGCTGGCTG

Dmau          AAGGCTGCCTATTTGACCTATCGAGACCCAGTCACCGTGTTCGTGAGTCCTGGCATGACC
Dsec          AAGGCTGCCTATTTGACCTATCGAGACCCAGTCACCGTGTTCGTGAGTCCCGGCATGACC
Dsim          AAGGCTGCCTATTTGACCTACCGAGACCCAGTCACCGTGTTCGTGAGTCCCGGCATGACC

Dmau          TTCCCCAAGCAAAACTTCAGGGACTCACGCGCTTTCGTGGACTATACCGCCAGGGTTATC
Dsec          TTCCCCAAGCAAAACTTCAGGGACTCACGGGCTTTCGTGGACTATACCGCGAGGGTTATC
Dsim          TTTCCCAAGCAAAACTTCAGGGACTCACGCGCTTTCGTGGACTATACCGCCAGGGTTATA

Dmau          TATGGCCTGGGCGANNTCAACGACATGGTGCACGCCAACCAAATTCCGATCGTTAAAATG
Dsec          TATGGACTGGGCGAATTTAACGACATGGTGCACGCCAACCAAATTCCGATCGTTAAAATG
Dsim          TATGGCCTGGGCGAATTCAACGACATGGTGCACGCCAACCAAATTCCGATCGTTAAAATG

Dmau          GGCAAGAACGAGCGGGACAACAGCCAGTTTGGCAANGTATTCGGCACATGTCGGATT-CC
Dsec          GGCAAGAACGAGCTGGACAACAGCCAGTTTGGCAAGGTATTCGGCACATGTCGGATTTCT
Dsim          GGCAAGAACGAGCTGGACAACAGCCAGTTTGGCAAGGTATTCGGCACATGTCGGATT-CC
                                                                       ^
Dmau          CAGACGGGACACCGACGAGATCGTATACAATCCCGACTCCGATTATGTGGTGGTGATCTA
Dsec          CAGACGGGGCACCGACGAGATCGTATACAATCCTGACTCCGATTATTTGGTGGTGATCTA
Dsim          CAGACGGGGCACCGACGAGATCGTATACAATCCCGACTCCGATTATGTGGTGGTGATCTA

Dmau          CAAGAATCACTTCTACCAACTGAAGATATACAGTAAGGAGGGAAAGCTCATTGCTGCTCC
Dsec          CAAGAATCACTTCTACCAACTGAAGATATACAGTAAGGAGGGAAAGCTCATTGCTGCTCC
Dsim          TAAGAATCACTTCTACCAACTGAAGATATACA----------------------------

Dmau          ATGTCTAGCTGCTCAACTCGAGAATATCTTGTTGAAGGAAACGCAAGTGGGAGTACCTTA
Dsec          ATGTCTAGCTGCTCAACTGGAGAATATCTTGTTGAAGGAAACGCAAGTGGGAGTACCCTA
Dsim          --------CTGCTCAACTGGAGAATATCTTGTTGAATGAAACGCAAGTGGGAGTACCCTA

Dmau          TGGTATTCTGACCACCGACTCCAGGGACANTTGGGCCCAACCCTACGAGTATCTGGCTGA
Dsec          TGGTATTCTGACCACCGACTCCAGGGACAATTGGGCCGAAGCCTACGAATATCTGGCTGA
Dsim          TGGTATTCTGACCACCGACTCCAGGGACAATTGGGCCGAAGCCTACGAATATCTGGCTGA

Dmau          GACTCCTGGCAACCGGGATGCCCTCAAGACCATACAGAGTGCTCTGTTCACCGTCTCACT
Dsec          AACTCCTGCCAACCGGGATGCCCTCAAGACCATACAGAGTGCTTTGTTCACCGTCTCACT
```

```
Dsim          GACTCCTGGTAACCGGGATGCCCTCAAGACCATTCAGAGTGCTCTGTTCACCGTCTCACT

Dmau          CGATGAGGGTACTAGCCTAAAGGACGGCGAAGAGACTGACGAACTTATTCTATCGCTGAT
Dsec          CGATGAGGGTACTAGCCTAAAGGACGGCGAAGAGACTGACGAACTTATTCTATCGCTGAT
Dsim          CGATGAGGGTACTAGCCTAAAGGAAGGCGAAGAGACCGACGAGATTATTCTATCGCTGAT

Dmau          CCATGGCAGTGGCAGCAAGAGGAACAGCGGCAACCGCTGGATGGACAAGACTATTCAGCT
Dsec          CCATGGCAGTGGCAGCAAGAGGAACAGCGGCAACCGTTGGATGGACAAGACTATTCAGCT
Dsim          CCATGGCAGCGGCAGCAAGAGGAACAGTGGCAACCGCTGGATGGACAAGACTATTCAGCT

Dmau          GGTGGTAAACCCCAATGGAAACGTCGGATTCACCTATGAGCACTCGCCGGCTGAGGGCCA
Dsec          GGTGGTTAACCCCAATGGAAACGTCGGATTCACCTATGAGCACTCGCCGGCTGAGGGCCA
Dsim          GGTGGTAAACCCCAATGGAAACGTCGGGTTCACCTATGAGCACTCGCCGGCTGAGGGCCA

Dmau          GCCCATTGCGATGATGATGGACTACGTGGTGCAAAAGATGTANGAAGACCCTAGCTTCGG
Dsec          GCCCATTGCGATGATGATGGACTACGTGGTGCAAAAGATGAAGGAAGACCCTAGCTTCGG
Dsim          GCCCATTGCGATGATGATGGACTACGTGGTGCAAAAGATGAAGGAAGACCCTAGCTTCGG

Dmau          GCAAAC-GGCTCACAGAACTT-GCTCCCGCACAGAAAATTCAGTTNNTCTCGAGCA-TAA
Dsec          GCAAACTGGCTCACAGGACTTTGCTCCCGCACAGAAAATACAGTTCTCTTCGAGTAATAA
Dsim          GCAAACTGGCTCACAGGACTTTGCTCCCGCACAGAAAATTCAGTTCTCTTCGAGCAATAA
                                                                        ^

Dmau          AAGTCTAGAGAAATCTTTAAACGTGTCCCNGGCAANCGTGGACAAACTTGCCGATGCTCT
Dsec          AAGTCTAGAGAAATCTTTAAACGTCGCACAGGCAAACGTGGACAAACTTGCCGATGCTCT
Dsim          AAGTCTAGAGAAATCTTTAAACGTCGCCCAGGCAAACGTGGACAAACTTGCCGATGCTCT

Dmau          CCAAATGAAGGTTCTGAAATTCACCGGCTTCGGAAAGGATTTCATAAAGAAACAGCGTCT
Dsec          CCAAATGAAGGTTCTGAAATTCAACGGCTTCGGAAAGGATTTCATAAAGAAACAGCGTCT
Dsim          CCAAATGAAGGTTCTGAAATTCACCGGCTTCGGAAAGGATTTCATAAAGAAACAGCGTCT


Dmau          GGGTCCGGACAGCTTTGTTCAGATGGCTCTTCAGCTCGCCTTCTACAAAATGCACTCGGA
Dsec          GGGTCCGGACAGCTTTGTTCAGATGGCGCTGCAGCTCGCCTTCTACAAAATGCACTCGGA
Dsim          GGGTCCGGACAGCTTTGTTCAGATGGCGCTGCAGCTCGCCTTCTACAAAATGCACTCGGA

Dmau          ACCGCCGGCGCAATATGAGTCGGCTCATCTGCGCATATTCGACGGTGGACGAACCGAAAC
Dsec          ACCGCCTGCGCAATATGAGTCGGCTCATCTGCGCATATTCGACGGCGGACGAACCGAAAC
Dsim          ACCGCCTGCGCAATATGAGTCGGCTCACCTGCGCATATTCGACGGCGGACGAACCGAAAC

Dmau          CATACGCTCTTGCTCCAACGAATCCCTGGCCTTTTCCCGCGCTATGCAGGACCCAAATGC
Dsec          CATACGCTCTTGCTCCAACGAATCCGTGGCCTTTTCCCGCGCTATGCAGGACCCAAATGC
Dsim          CATACGCTCTTGCTCCAACGAATCCCTGGCCTTTTCCCGCGCTATGCAGGACCCAAATGC

Dmau          TACCGATCAGGAACGCGCCGCTAAGCTTCGTGAGTGCAGTAG-GTCTCATCGAACA----
Dsec          TACCGATCAGGAACGCGCCACTAAGCTTCGTGAG-GCAGTAGTGTCTCATCAGACATATG
Dsim          TACCGACCAGGAACGCGCCGCTAAGCTTCGTGAA-GCAGTAGTGTCCCATCAGACATATG


CG7362
Dsim           ATGCTACCTACGGGAAACAGATTTCCAAAGAATAATGGGAAAATAATGCCTCTAATTATC
DsimL          ------------------------------------------------------------
Dmau           ------------------------------------------------------------
Dsec           ATGCTACCTACGGGAAACAGATTTCCAAAGAATAATGGGAAAATAATGCCTCTAATTATC

Dsim          ATTAATAAATCGAGAGACAAGTCAACAAACCCGGCCGTCGAGTCAACAACTTGGTTGCGC
DsimL         ------AAATCGAGAGACAAGTCAACAAACCCGGCCGTCGAGTCAACAACTTGGTTGCGC
Dmau          ------AAATCGAGAGACAAGTCAACAAACCCGGCCGTCGAGTCAACAACTTGGTNGCGC
Dsec          ATTAATAAATCGAGAGACAAGTCAACAAACCCGGCCGTCGAGTCAACAACTTCGTTGCGC

Dsim          TTCAAGGCAGCCAAAGGAAAACGTTTCCTTTTGGCCAATGGATTGGATGGCGGCCAAACA
DsimL         TTCAAGCAGCCCAAAGGAAAACGTTTCCTTTTGGCCAATGCATTAAATGGCGGCCAAACA
Dmau          TTCAANGCAGCCAAAGGAAAACGTTTCCTTTNGGCCAATGGATNGGATGACNGCCAAACA
```

```
Dsec          TTCAAGGCAGCCAAAGGAAAACGTTTCCTTTTGGCCAATGGATTGGATAGCGGGAAAACA

Dsim          ATGCGTCCAGTTTGGACAAACATTTGCGGCAGCAAACCCATTTCATGCCCATTCCAACTT
DsimL         ATGCGTCCAGTTTGGACAAACATTTGCGGCAGCAAACCCATTTCATGCCCATTCCAACTT
Dmau          ATGCGTNCAGTTTGGACAAACATTTGCGGCAGCAAACCCATTTCATGCCCATTCCAACTT
Dsec          ATGCGTCCAGTTTGGACAAACATTTGCGGCAGCAAACCCATTTCATGCCCATTCCAACTT

Dsim          CCCACGTTGCACGGCAGCCTCAAAATTCAGATTTTTTTTAAATTTATCATATATTTGGTC
DsimL         CCCACGTTGCACGGCAGCCTCAAAATTCAGATATTTATTAAATTTATCATATATTTGGTC
Dmau          CCCACGTTGCACGGCAGCCTCAAAATTCAGATTTTTATTAAATTTATCATATATTTGGTC
Dsec          CCCACGTTGCACGGCAGCCTCAAAATTCAGATTTTTATTAAATTTATCATATATTTGGTC

Dsim          CGCATTACGCAAAATTTTTTTCAAAATTCGAAAATACGCCGGAAATCTTTGTCAGAGGGC
DsimL         TGCATTACGCAAAATTTTTT-CAAAATTCGAAAATACGCCGGAAATCTTTGTCAGAGGGC
Dmau          TGCATTACGCAAAATTTTTT-CAAAATTCGAAAATACGCCGGAAATCTTTGTCAGAGGGC
Dsec          CGCATTACGCAAAATTTTTTTCAAAATTCGAAAATACGCCGGAAATCTTTGTCAGAGGGC
                                    ^

Dsim          AAGCCTCTGCGTTACGTCACCATGTTCCAATTTTCCATTGAGCCTTAGATTTCGCTGCTA
DsimL         AAGCCTCTGCGTTGCGTCACCATGTCCCAATTTTCCATCAAGTTTTAGATTTCGCTGCTA
Dmau          AAGCCTCAGCGTTGCGTCACCATGTCCCAATTTTCCATCAAGCCTTAAATTTCGCTGCTA
Dsec          AAGCCTCTGCGTTGCGTCACCAT----------------GCCTTAGATGTCGCTGCTA

Dsim          TGCCGACAAATGCAAACGGAAAGTGCAACCCAAGAAGTTCCGGCTTGACCGTGGACCCTA
DsimL         TGCTGACAAATGCAAACGGAAAGTGCAACCCAAGAAGTTCCGGCTTGACCGTGGACCCTA
Dmau          TGCCGACAAATGCAAACGGAAAGTGCAACCCAAGAAGTTCCGGCTAGACCGTGGACCCTA
Dsec          TGCCGACAAATGCAAACGGAAAGTGCAACCGAAGAAGTTCCGGCTAGACCGTGGACCCTA

Dsim          CCTTTCCCAGCTGGACTACCAATCTCGTCTGCAGTTCCAAGCACCAGCTCTGAGGTTACC
DsimL         CCTTTCCCAGCTGGACTACCAATCTCGTCTGCAGTTCCAAGCACCAGCTCTGAGGTTACC
Dmau          CCTTTCCCAGCTGGACTACCAATCTCGTCTGCAGTTCCAAGCACCAGCTCTGAGGTTACC
Dsec          CCTTTCCCAACTGGACTACCAATCTCGTCTGCAGTTCCAAGCACCAGCTCTGAGGTTACC

Dsim          CCTCACCAGCATTATATGCACCATTGGACCCTCATCCAGCCAGCCTGAAGTTCTCCTTAA
DsimL         CCTCACCAGCATTATATGCACCGTTGGACCCTCATCCAGCCAGCCCGAAGTTCTCCTTAA
Dmau          CCTCACCAGCATTATATGCACCATTGGACCCTCATCCAGCCAGCCCAAAGTTCTCCTCAA
Dsec          CCTCACCAGCATTATATGCACCGTTGGACCCTCATCCAGCCAGCCCGAAGTTCTCCTTAA

Dsim          TCTCATTCATGCTGGGATGAAGGTGGTCCGATTGGACTTCTCCCACGGCACCCACGATTG
DsimL         TCTCATTCATGCTGGGATGAAGGTGGTCCGATTGGACTTCTCCCACGGAACCCACGAATG
Dmau          TCTCATTCATGCTGGGATGAAGGTGGTCCGATGGGACTTCTCCCACGGCACCCACGAATG
Dsec          TCTCATTCATGCTGGGATGAAGGTGGTCCGATTGGACTTCTCCCACGGCACCCACGAATG

Dsim          CCATTGCCAGGCAATACAGGCGGCACGTAAAGCCATCGCCATGTATGTGGAGGAAACGGG
DsimL         CCATTGCCAGGCAATCCAGGCGGCACGTAAAGCCATCGCCATGTATGTGGAGGAAACGGG
Dmau          CCATTGCCAGGCAATCCAGGCGGCACGTAAAGCCATCGCCATGTATGTGGAGGANACGGG
Dsec          CCATTGCCAGGCAATCCAGGCGGCACGTAAAGCCATCGCCATGTATGTGGAGGAAACGGG

Dsim          TCTTTCCAGGAGCTTGGCCATTGCACTGGACACCAAGGGTCCGGCAATCAATCCACAGGG
DsimL         TCTTCTCAGATGCTTGGCCATTGCACTGGACACCAAGGGTCCGGCAATCAATCCACCGGG
Dmau          TCTACCCAGAAGCTTGGCCATTGCCCTGGACACCAAGGGTGCGGCAGTCAATCCACAGGG
Dsec          TCTTCCCAGATGCTTGGCCATTGCCCTGGACACCAAGGGTCCAGAAATCAATCCACAGGG

Dsim          TGTAGCTGTTGATTTAAACGCCATAACCGAGCAAGACAAACTGGATCTCAAGTTTGGGGC
DsimL         TGTAGCTATTGATTTAAACACCATAACCGAGCAAGATAAACTGGATCTCAAGTTTGGGGC
Dmau          TGCAGCTGTTGATTTCAACGCCATAACCGAGCAAGACAAACTGGATCTCAAGTGTGGGGC
Dsec          TGTACCTGTTGATTTAAACGCCATAACCGAGCAAGACAAACTGGATCTCAAGTTTGGGGC

Dsim          GGATCAGAAGGTGGACATGGTATTCGCGTCGTTCATCCGCGATGCCAAAGCTTTGCAAGA
DsimL         GGATCAGAAGGTGGACATGATATTCGCGTCGTTCATCCGCGATGCCAAAGCTTTGAAAGA
Dmau          GGATCAGAAGGTGGACATGATCTTCGCGTCGTTCATCCGCGATGCCAAAGCTTNGCAAGA
Dsec          GGATCAGAAGGTGGACATGATCTTCGCGTCGTTCATCCGCGATGCTAAAGCTTTGCAAGA
```

```
Dsim          AATTCGCCAGGCACTGGGTCCATCAAGTGAGCACATAAAGATCATTTCCAAGATCGAAAG
DsimL         AATTCGCCAGGCACTGGGTCCATCAAGTGAGCACATAAAGATCATTTCCAAGATCGAAAG
Dmau          AATTCGCCAGGCACTGGGTCCATCAAGTGAGCACATAAAGATCATTTCCAAGATCGAAAG
Dsec          AATTCGCCAGGCACTGGGTCCATCAAGTGAGCACATAAAGATCATTTCCAAGATAGAAAG

Dsim          TCAACAGGCTCTGGCGAACATAGATGAGATAATCCGCGAATCCGATGGCATAATGGTGGC
DsimL         TCAACAGGCTCTGGCGAACATAGATGAGATAATCCGCGAATCCGATGGCATAATGGTGGC
Dmau          TCAACAGGCTCTGGCGAACATAGATGAGATAATCCGCGAATCCGATGGCATAATGGTGGC
Dsec          TCAACAGGCTCTGGCGAACATAGATGAGATAATCCGCGAATCCGATGGCATAATGGTGGC

Dsim          CCTTGGGAATATGGGCGACGAAATAGCACTGGAGGCTGTACCGCTGGCCCAGAAATCGAT
DsimL         CCTTGGGAATATGGGCAACGAAATAGCACTGGAGGCTGTACCGCTGGCCCAGAAATCGAT
Dmau          CCTTGGGAATATGGGCAACGAAATAGCACTGGAGGCTGTACCGCTGGCCCAGAAATCGAT
Dsec          CCTTGGGAATATGGGCAACGAAATAGCACTGGAGGCTGTACCGCTGGCCCAGAAATCGAT

Dsim          CGTGGCCAAGTGCAACAAAGTTGGAAAGCCTGTGATCTGTGCCAATCAAATGATGAATTC
DsimL         CGTGGCCAAGTGCAACAAAGTTGGAAAGCCTGTGATCTGTGCCAATCAAATGATGAATTC
Dmau          CGTGGCCAAGTGCAACAAAGTTGGAAAGCCTGTGATCTGTGCCAATCAAATGATGAATTC
Dsec          CGTGGCCAAGTGTAACAAAGTTGGAAGGCCTGTGATCTGTGCCAATCAAATGATGAATTC

Dsim          GATGATAACCAAGCCACGTCCCACTCGCGCCGAATCCTCTGATGTGGCAAACGCAATCTT
DsimL         TATGATAACCAAGCCACGTCCCACACGCGCCGAATCTTCTGATGTGGCAAACGCAATCTT
Dmau          GATGATAACCAAGCCACGTCCCACACGCGCGGAATCTTCTGATGTGGCAAACGCAATCTT
Dsec          GATGATAACCAAGCCACGTCCCACACGCGCCGAATCTTCTGATGTGGCAAACGCAATCTT

Dsim          GGATGGTTGTGATGCCCTTGTGTTGTCTGATGAAACGGCCAAGGGTAAGTACCCGGTGCA
DsimL         GGATGGTTGTGATGCCCTTGTGTTGTCAGATGAAACGGCCAAGGGTAAGTACCCGGTGCA
Dmau          GGATGGTTGTGATGCCCTTGTGTTGTCAGGTGAAACGGCCAAGGGTAAGTACCCGGTGCA
Dsec          GGATGGTTGTGATGCCCTTGTGTTGTCTGATGAAACGGCCAAGGGTAAGTACCCGGTGCA

Dsim          ATGTGTGCAGTGCATGGCCAGAATCTGCGCCAAGGTGGAGTCGGTTTTATGGTACGAGAG
DsimL         ATGTGTGCAGTGCATGGCCAGAATCTGCGCCAAAGTCGAGTCCGTTTTATGGT-------
Dmau          ATGTGTGCAGTGCATGGCCAGAATCTGCGCCAAGGTGGAGTCGGTTTTATGGTACGAGNG
Dsec          ATGTGTGCAGTGCATGGCCAGAATCTGCGCCAAGGTGGAGTCGGTTTTATGGTACGAGAG
```