REPRODUCTIVE ISOLATION IN *DROSOPHILA*: FROM BEHAVIOR TO GENOME

by

DAINA MA

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2010

ACKNOWLEDGEMENTS

My study, research and life at the University of Texas at Arlington were the greatest dream that I could ever have and realized it in a real life! This dissertation would not have been possible without the guidance and support from lots of people. First, I would like to express my special gratitude to my supervisor, Dr. Pawel Michalak, who helps me in the wonderful world of little fruit flies with his guidance, motivation, and patience all the time. Besides his instruction to my research itself, he also inspires me a lot with his positive attitude towards life. I sincerely hope he is going to have a better career and life in another university. He really deserves it. I would like to express my truly appreciation to all my thesis committee members: Dr. Jeffery Demuth, thank you for your willingness to supervise me together with Pawel. Dr. Esther Betrán, thank you for your patience and knowledgeability; whenever I have questions for my research or for my coursework, you are there to help. Dr. Andre Pires da Silva, thank you for your suggestions in research and in class, I love your teaching style very much. Dr. John Fondon, thank you to be agreeable to be my committee member in a second. Also I appreciate all your valuable time in the busy summer to review my dissertation and schedule a time for defense. I express my sincere gratitude for all your work.

I also thank my lab members Mersee Madison-Villar, Alie Koroma, Daiju Hoshino, John Malone, Ivan Lee, Ivy Zheng etc. for their collaboration in research, and for the happy time we spend together in the past five years. I am also grateful for the support from faculty, stuff and all graduate students in biology department.

Last but not the least, I thank my family for their unconditional love and enthusiastic support during my life. Thanks to dad and mom's spiritual support to let me pursue my dream here instead of keeping me at their elbow in China.

July 26, 2010

ABSTRACT

REPRODUCTIVE ISOLATION IN *DROSOPHILA*:

FROM BEHAVIOR TO GENOME

Daina Ma, PhD

The University of Texas at Arlington, 2010

Supervising Professors: Pawel Michalak and Jeffery Demuth

In this study, I investigate factors underlying reproductive isolation in *Drosophila*. I begin (Part I) by measuring behavioral isolation between Zimbabwe and cosmopolitan *D. melanogaster*, and asking the following questions: What is the pattern? What are the sensory cues leading to isolation? In Part II, I present results of a genomic screen for candidate genes underlying behavioral isolation between Zimbabwe and cosmopolitan *D. melanogaster*. Then I turn to the genomic causes of postzygotic reproductive isolation (Parts III and IV): What are the possible genes associated with hybrid sterility? What are the relationships between fertility/sterility and genotypes and allelic expression? What are the regulatory changes in hybrid sterile males compared to normal males? Finally (Part V), I broaden the analysis from a single gene to a microsatellite mapping of chromosome III, and ask the following question: How many loci are involved in hybrid sterility? What is the degree of microsatellite divergence between species?

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure Page

LIST OF TABLES

CHAPTER 1

INTRODUCTION TO REPRODUCTIVE ISOLATION: FROM BEHAVIOR TO GENES, AND FROM GENES TO HYBRID STERILITY (REVIEW)

According to Mayr's biological species concept (BSC), species are groups of interbreeding natural populations that are reproductively isolated from other such groups (Mayr 1942). Therefore, sexually reproductive isolation from one lineage to other lineages is the main feature characterizing distinct species. Characters that evolve within one population and perform regular functions may produce side-effects in the form of isolating barriers against other populations (pleiotropy). In some cases, these isolating mechanisms are direct and intuitive: geographic separation and adaptation to new conditions allow isolated populations (allopatry) to diverge and accumulate differences. In other cases, it is indirect: divergent selection may lead to different adaptations between taxa even within same habitat (parapatry and sympatry). Then the two central questions are: what reproductive barriers were involved in isolating certain populations; and what evolutionary forces produced the barriers?

There are two major forms of reproductive isolation: prezygotic and postzygotic isolation, which create barriers preventing fertilization and barriers preventing fertilized eggs from developing into perfectly healthy and fertile adults, respectively (Johnson, 2006). Prezygotic isolation has been considered as a more critical and effective barrier to reduce gene flow than postzygotic isolation (Jiggins *et al.* 2001; Kirkpatrick and Ravigne 2002). Given the fact that they act early in the life cycle and have sequentially effects, prezygotic barriers are surely the strongest current obstructions to gene flow. Forms of prezygotic isolation include spatial, temporal, mechanical, and behavioral barriers. The spatial and temporal barriers are largely due to environmental factors. Spatial isolation occurs when the organisms are facing different habitats or

breeding at different sites. In plants, the classic reciprocal transplant experiment of Clausen *et al.* (1940) revealed spatial isolation between allopatric species. In California, *Potentilla glandulosa (Rosaceae)* is subdivided into several subspecies in the San Joaquin Valley, from the coast to the sierra Nevada, because they vary substantially in quantitative characters (in height and flowering time etc.) along this altitudinal transect. Individuals originally from low elevations had poor survival at high elevations. For animals, Lynch (1978) provided an example of habitat isolation between the frogs *Rana blairi* and *R. pipiens*. Although their ranges overlap in Nebraska, they mate and breed in different cites—*R. blairi* breeds in turbid streams while *R. pipiens* stays in clear streams.

Temporal isolation occurs when individuals from different species breed at different times. Opler *et al.* (1975) provided an example of eight species of *Cordia* in Costa Risa that differ in flowering time by several weeks. Knowlton *et al.* (1997) showed that the spawning times between two sister species of corals *Montastraea annularis* and *M. franksi* are 1.5-3 hours apart, which sufficiently dilutes sperm from the earlier-spawning species and thus prevents fertilization of eggs from the later species.

Mechanical isolation often results from incompatibility between the organisms' reproductive structures failing to lead to successful copulation. Grant (1994) reviews modes of mechanical isolation in angiosperms, and gives an example of flowers *Salvia apiana* and *Salvia mellifera* that differ in size and conformation to adapt to different pollinators--*S. apiana* are pollinated by large carpenter bees (*Xylocopa*) and *S. mellifera* by relatively small bees (*Anthophora, Osmia*, etc.). Conversely, the floral differences set up a barrier for reciprocal pollination. Paulson's study (1974) on damselfly species showed that males have difficulties to clasp females from other species because of the unfitness of the male appendages to the heterospecific female thorax, thus preventing copulation.

Behavioral isolation occurs when species differences reduce the attraction between two sexes and therefore prevent mating between heterospecific individuals. Compared to the other three prezygotic barriers, behavioral isolation with its intricate subtleness is more difficult to

dissect and understand. Obviously, behavioral isolation is limited to animals, and it involves the interaction between males and females in different species for certain traits. Typically, one sex has a stronger preference for certain traits of the conspecific individual of the opposite sex than the heterospecific individual, and therefore would mate with conspecific individual more often. Because it is an early acting isolation barrier in the life cycle, as well as many observations for sexual dimorphisms that can by driven by sexual selection for adaptation (e.g., Hawaiian *Drosophila*), many biologists have considered behavioral isolation instrumental in speciation. Mayr (1963) claimed that "if we were to rank the various isolating mechanisms of animals according to their importance, we would have to place behavioral isolation far ahead of all others." Description of behavioral isolation is relatively easy, but to understand which traits are involved in the process is more difficult. Despite hundreds of behavioral studies on *Drosophila*, only several study cases identify traits involved in behavioral isolation (Greenspan 2000). Also, different species even within *Drosophila* group most likely rely on different traits to distinguish conspecific from heterospecific individuals. For example, *D. melanogaster* can mate in the dark whereas *D. simulans* are less likely to do so (Spieth 1950). Those traits and variability between taxa must be explored if we want to understand the evolution of behavioral isolation and female preference/male courting.

Compared with prezygotic isolation, postzygotic isolation occurs later in the life cycle. Although much work is needed to determine which are more important in the process of speciation --- prezygotic or postzygotic barriers, the fact that postzygotic isolation acts late does not mean it is insignificant. In fact, postzygotic isolation barriers are critically important factors for many reasons. In nature, the existence of hybrids for many sister species indicates that reproductive isolation is not complete. Hybrid sterility and inviability represent various levels of gene flow disruption between species. In practice, the most efficient approach to detect and measure postzygotic isolation strength is to compare fitness of hybrids and their closely-related parental species, when prezygotic isolation has not lead to complete speciation yet. Therefore, evolutionists prefer to study postzygotic isolation using hybrid schemes for various experiments.

Postzygotic isolating barriers are classified as two distinct forms: extrinsic and intrinsic isolation (Coyne and Orr 2004; Table 7.1). Extrinsic postzygotic isolation occurs when hybrids have intermediate phenotypes, although not suffering inherent developmental defect, experience lower fitness than the two parents in the present environments, due to the divergent natural selection (Schluter, 2000) or as a by-product of selection. For example, Wang *et al.* (1997) found that two subspecies of the big sagebrush *Artemisia tridentia* (in basin and mountain) perform best in their respective environment, whereas hybrids show much lower fitness in each parental habitat and most fit in the intermediate environment.

Intrinsic postzygotic isolation is considered to reflect inherent developmental problems that render unfit hybrid partially or completely inviable or sterile (Coyne 1992). A well known example, first studied by Sturtevant (1920), is that when *D. melanogaster* females are crossed with *D. simulans* males, only hybrid daughters survive, hybrid sons die in the early development; the reciprocal cross produces only hybrid sons but females die as embryos. Another example is the one between the *Drosophila* sibling species used in this study, *D. simulans* and *D. mauritiana,* when crossed produce both sexes viable in the hybrid progeny, with fertile females and sterile males. Hybrid dysfunctions are caused by many factors, and the most accepted view is that chromosomal differences and genic incompatibilities resulting from the fact that two distinct genomes are merged into one individual play a major role. Chromosomal differences could act in different ways: different ploidy levels; same chromosome sets but miss or gain one or more chromosomes; and rearrangements of partial chromosome. All of these could lead to mismatch of sister chromosomes and therefore mess up whole sets of genes resulting in hybrid difficulties and reproductive isolation. Theoretical models of genic incompatibility demonstrate that a single gene incompatibility between two species can be sufficient to produce reproductive isolation (Coyne and Orr 2004). The Dobzhansky-Muller genetic incompatibility model is based on a simple idea that is widely used to explain intrinsic postzygotic isolation and refers to incompatibility between loci. If only one locus of the two changes, the population is perfectly viable and fertile; but a cross of two populations with both substituted loci present would lead to

deleterious effects resulting in hybrid inviability or sterility. Since the model provides a mathematically tractable theory of genic incompatibilities, it has been extensively studied and well received (Dobzhansky 1936, Muller 1942).

Many studies focus on intrinsic postzygotic isolation, most of which involve *Drosophila* because of their relatively short generation times and well-characterized genetic markers, and particularly species of the *melanogaster* group (see Appendix A: phylogenetic tree of *Drosophila*). A large portion of those studies has been inspired by a single problem—Haldane's rule. Haldane's rule states that "when in the offspring of two different animal races one sex is absent, rare, or sterile, that sex is the heterozygous [heterogametic, i.e., XY or ZW] sex"(Haldane 1922). The pattern is well obeyed in most of the taxa surveyed—*Drosophila*, mammals, Lepidoptera, birds, and others (Coyne 1992a, Table 8.1 The frequency of Haldane's rule). A number of theories have been offered to explain Haldane's rule, with the following four to represent the most popular ones – the dominance theory (Mulller 1940), the faster-male evolution theory (Wu and Davis 1993), the large-X theory (Charlesworth *et al.* 1987), and the newly proposed meiotic drive theory (Jaenike 2001)

For dominance theory, a so-called Dobzhansky-Muller model describes a situation in which two or more genes interact to cause hybrid unfitness. If X-linked and autosomal genes interact, hybrid males (XY) are all affected regardless of whether these genes are dominant or recessive because of the heterogametic genotype; but hybrid females (XX) will be affected only if the X-linked genes are dominant (Orr and Turelli 2001).

The faster-male evolution theory assumes that hybrid male sterilities are more common than female sterilities because of stronger sexual selection on or more sensitive expression of male-expressed genes. Recent genetic analyses have found a very high rate of evolution of hybrid male sterility relative to female sterility in support of Wu and Davis' faster male evolution model. In True *et al.*'s paper (1996), only 5.4% introgression lines were female sterile, compared with 50% for male sterility.

The large-X effect theory is proposed given the fact that genes affecting postzygotic

isolation often map to the X chromosome regardless of their dominance (Guenet *et al.* 1990); therefore heterogametic hybrids suffer more than homogametic hybrids. Charlesworth *et al.* (1987) showed that X-linked gene mutations occur more often than those on the autosomes.

Meiotic drive may take on many forms but when occurring in the sex chromosome, it distorts Mendelian ratio away from 1:1 sex ratio at the cost of one sex (Sandler and Novitski 1957; Frank 1991). The gene *Overdrive (GA19777)* causes hybrid male sterility in F_1 hybrids between the Bogota and U.S.A. subspecies of *Drosophila pseudoobscura,* and also produces a female-biased sex ratio when hybrid males with $GA19777^{BOG}$ transgene crosses with Bogota *white* females (Phadnis and Orr 2009).

A consensus has been reached that dominance and faster-male evolution are primary causative factors underlying Haldane's rule, but large-X effects and meiotic drive also play a role in hybrid unfitness (Coyne and Orr 2004). In contrast to theory of speciation genetics, progress in identifying actual genes responsible for reproductive isolation was slow and unsatisfactory, a situation perhaps not unexpected given that genetic crosses between different species are almost by definition difficult or impossible (Appendix B existing genetic analyses of reproductive isolation between *Drosophila* species). As a result, only a handful genes have been thus so far documented for hybrid sterility and inviability: *Tu* (macromelanophore), located on X chromosome, which is a dominant tumour gene for melanoma in *Xiphophorus* fish hybrids (Wittbrodt *et al.* 1989); *Odysseus* (*Ods*), a male sterility gene on the *Drosophila* X chromosome (Ting *et al.* 1998); *Hybrid male rescue* (*Hmr)* (also X chromosome), a dosage-dependent gene that affects hybrid viability in *Drosophila* (Barbash *et al.* 2000); *Nup96*, a hybrid inviability gene in hybrid males located on the 3R chromosome in *Drosophila* (Presgraves *et al.* 2003); *JYAlpha,* located on the fourth chromosome in *D. melanogaster* (but third chromosome in *D. simulans*), a hybrid male sterility gene via gene transposition in *Drosophila* (Masly *et al.* 2006); *Prdm9*, a hybrid sterility gene (chromosome 17) responsible for spermatogenic failure in the mouse (Mihola *et al.* 2009); *Overdrive* (*GA19777*), a gene (X chromosome) causes both male sterility and segregation distortion in *Drosophila* (Phadnis and Orr 2009); *NB-LRR*, an inviability gene (chromosome 3)

causing hybrid necrosis due to improper autoimmune-like responses in *Arabidopsis* (Bomblies *et al.* 2007); *AEP2* (chromosome 13), a sterility gene that causes a sporulation defect by the inability to regulate the translation *OLI1* mRNA in yeast (Lee *et al.* 2008)

Recent developments in comparative genetics and genomics have provided new tools for explorations into the genetic background of reproductive isolation. For example, gene regulation changes and epistatic interactions (such as Dobzhansky-Muller incompatibilities) were long postulated to play a critical role in speciation and species divergence (Dobzhansky 1937, Muller 1940, King and Wilson 1975), but their systematic examination at the genome-wide level had to await the renaissance of reverse genetics and the emergence of microarray technology in particular (Reiland and Noor 2002; Michalak and Noor 2003; Ranz *et al.* 2004; Barbash and Lorigan 2006; Hegarty *et al.* 2006; Josefsson *et al.* 2006; Lai *et al.* 2006; Ranz and Machado 2006; Ortíz-Barrientos *et al.* 2007; Michalak *et al.* 2007). This new transcriptome-oriented approach starts at the level of phenotype and compares transcriptomes of parental species with transcriptomes of their dysfunctional hybrids and leads to the identification of the candidate loci that contribute to reproductive isolation or/and divergence between parental species.

In this study, I investigate factors underlying reproductive isolation in *Drosophila*. I begin (Part I) by measuring behavioral isolation between Zimbabwe and cosmopolitan *D. melanogaster*, and asking the following questions: What is the mating pattern? What are the sensory cues leading to isolation? In Part II, I present results of a genomic screen for candidate genes underlying behavioral isolation between Zimbabwe and cosmopolitan *D. melanogaster*. Then I turn to the genomic causes of postzygotic reproductive isolation (Parts III and IV): What are the possible genes associated with hybrid sterility? What are the relationships between fertility/sterility and genotypes and allelic expression? What are the regulatory changes in hybrid sterile males compared to normal males? Finally (Part V), I broaden the analysis from a single gene to a microsatellite mapping of chromosome III, and ask the following question: How many loci and which chromosome segments are involved in hybrid sterility? What is the degree of microsatellite divergence between species?

PART I

CHAPTER 2

BEHAVIORAL ISOLATION BETWEEN ZIMBABWE AND COSMOPOLITAN *DROSOPHILA MELANOGASTER*

2.1 Introduction

Genetically inherited changes in fitness following reproductive isolation are thought to underlie evolution. Reproductive isolation can stem from physical isolation preventing gene pool mixing, for example when populations in separate locations or adopting activity at different times of day in the same location fail to interbreed, or alternatively through behavioral isolation, where populations fail to interbreed because of innate mating preferences (Coyne & Orr 2004). Understanding how sexual isolation evolves requires that we capture the process in *status nascendi*; before it has reached completion.

African populations of *Drosophila melanogaster* from Zimbabwe (Z) preferentially mate with flies from their own populations when given choice (Wu *et al.* 1995; Hollocher *et al.* 1997a,b). Sexual discrimination in other *D. melanogaster* populations or between strains derived from different continents is much weaker or non-existent (Wu *et al.* 1995; Henderson & Lambert 1982; Capy *et al.* 2000, Korol *et al.* 2000, Yukilevich & True 2008a,b), hence this non-discriminative behavioral morph has been dubbed Cosmopolitan (M). Further, Z males readily mate with M females, indicating that only Z females discriminate between these two types of males. This mating pattern provides a rare example of incipient speciation through behavioral isolation in an easily tractable genetic model.

There has been much interest and controversy surrounding the genetic bases of behavioral isolation between Z and M populations, with a particular focus on the role of the candidate pheromone synthesis locus *desat*-2, identified through a polymorphism in female cuticular hydrocarbons (Adams, Celniker *et al.* 2000). It has been suggested that *desat*-2 variants

cause sexual isolation and adaptation to climate, with Z alleles better adapted to tropical conditions and M alleles to more temperate ones, respectively (Greenberg *et al.* 2003; Michalak e*t al.* 2007; Yukilevich & True 2008b; but see Coyne & Elwyn 2006). However, little attention has been addressed to analysis of the role of basic senses in the Z-M mating preference, despite the fact that such knowledge might also be highly informative in disentangling the genetic bases of the phenotypes. For example, if one demonstrates that chemosensory signals play no role in producing behavioral isolation in this system, efforts to validate candidate genes could be streamlined by avoiding these genes.

We hypothesize that the Z-M discrimination pattern reflects an evolutionary divergence in one or more sensory systems. To dissect the relative contribution of sensory factors to behavioral isolation between Zimbabwe flies and cosmopolitan flies, we used multiple-choice tests in which Zimbabwe females are given a choice between monotypic (Z) and heterotypic (M) males under conditions controlling olfactory, visual, and auditory cues known to contribute to sexual isolation behavior in *D. melanogaster* (Markow & O'Grady 2005).

In *D. melanogaster*, courtship behavior involves a series of interactions between males and females in response to reception and processing of various sensory signals (e.g., Hall 1994; Villella & Hall 2008). Once an adult male detects a potential mate, he immediately runs after (or toward) the female. This is likely mediated by visual and olfactory cues (Koganezawa *et al.* 2009; Krstic *et al.* 2009). When the courting pair gets closer, physical contact occurs. The male touches the female's body with his legs or antennae to identify chemical cues that are species and sex dependent. The female may also touch the male the same way to validate an appropriate courting male. Later, the male vibrates one of his wings to produce a species-specific courtship "song" (Clyne & Miesenbock 2008; Ejima & Griffith 2008). Following wing vibration, the male runs after the female and tries to grab the female's body with his forelegs (there are sex combs on them to secure the holding) and curls the tip of his abdomen (generative organ) to attempt copulation. If the female accepts the male, after repeating this process several times, they start to mate and it will last for about 10 to 30 min. If the female rejects a male, she will try to escape from the male's

grabbing and move away from him.

During the courting process, vision is one of the most important senses for the courtship behavior. *norpA* mutants, that are completely blind due to lack of phopholipase C activity in their photoreceptors (Bloomquist *et al.* 1988) have dramatic delays in courting (Hall 1994). However, blind flies, or wild type flies in dark still have the ability to mate to appropriate partners without vision. This implicates that mating can take place without vision and that other prominent cues (such as chemosensation and wing vibration) allow them to mate (Sturtevant 1915; Bastock and Manning 1955; Spieth 1974; Greenspan and Ferveur 2000). Here, we analyze the relative contributions of visual, chemosensory and auditory input to courtship.

2.2 Methods

2.2.1 Treatments

Four strains of *Drosophila melanogaster* were used including Zimbabwe zim30, zim49 and zim53 (kindly provided by Jerry Coyne), and the cosmopolitan strain EC175 (Ecuador, courtesy of Esther Betrán). 7 day-old virgin individuals were collected prior to mating and kept separately in dark conditions. Males were marked by painting a small dot on the thorax with color markers. Colors were alternated between experiments to ensure no mating preference for marker color or odor.

Four groups of mating experiments were set up with different treatments. The first experiment was a control group with normal flies without any body parts altered. A group of 30 Z females, 30 Z males, and 30 M males each were placed together in a translucent mating chamber. When a pair of flies successfully copulated, they were taken out and the type of male (Z or M) was determined. After half of the females mated or until there were no more flies mated in two-hour period, the experiment was terminated.

To control for olfactory cues, 30 females with both antennae and maxillary palps surgically removed were mated with 30 Z males and 30 M males. Olfactory signal detection in *Drosophila* is mediated by sensory neurons located in hair-like sensilla distributed over the surface of the third antennal segment and the maxillary palps. A large family of about 60 odorant

receptors, 1300 olfactory receptor neurons (ORNs), and 50 glomeruli have been identified to express topographically in either the antenna or the palp (Hallem and Carlson 2004; Couto, Alenius *et al.* 2005). We removed the whole antenna as well as the palps in this experiment and herein refer to these flies as “antennae-less”. To eliminate visual cues, the multi-choice mating experiment with Z females and Z and M males was conducted in a dark room. To control for wing-based auditory cues, both Z and M males with wings surgically removed were mated with normal Z females. Each experiment was repeated multiple times and the data was subjected to statistical analysis (Table 2.1).

Table 2.1 Number of experiments performed for each treatment of Z-M behavior.

	olfaction	vision	audition
zim30	8	26	8
zim49	22	11	39
zim53	23	19	25

2.2.2 Statistical analyses

Numbers of monotypic mates and heterotypic mates from each mating experiment were recorded. Ratios of mated males over total females in each mating trial were used in order to make the numbers for each treatment more comparable. Ratios of monotypic mated flies in experimental group (Ez-z), heterotypic mated flies in experimental group ($E_{Z\text{-}M}$), monotypic mated flies in control group ($C_{Z\text{-}Z}$), and heterotypic mated flies in control group ($C_{Z\text{-}M}$) were derived and used to estimate the relative effects (distance) on each treatment as $d=(E_{Z\text{-}Z} - C_{Z\text{-}Z}) - (E_{Z\text{-}M} - C_{Z\text{-}M})$. In this equation, subtraction of ($E_{Z\text{-}Z} - C_{Z\text{-}Z}$) shows the pure effect of treatment (e.g., antennae removal) for monotypic mating preference, and ($E_{Z\text{-}M} - C_{Z\text{-}M}$) calculates the pure treatment effect for heterotypic mating preference. The difference between these two terms indicates a deviation from original mating preference exclusively due to treatment effect, and the bigger the absolute d value is, the greater treatment effect on the process of assortative mating. A $d=0$ means no difference between monotypic mating and heterotypic mating after treatment; $d<0$ indicates a

decrease (and d>0 an increase) of assortative mating due to treatment. After the transformation, three groups of distance (d) values among three strains and three senses were compared by a two-way ANOVA, followed by Tukey's post hoc test, and *p* values less than 0.05 were considered significant. All statistics and graphs were produced using SYSTAT software.

2.3 Results and discussion

Similar to previous reports, we found that the Zimbabwe female flies showed a preference for mating with Zimbabwe males in our standard assay conditions (Figure 2.1) (Wu *et al.* 1995; Hollocher *et al.* 1997a,b; Alipaz *et al.* 2001; Ting *et al.* 2001). Similarly, we observed slightly variant degrees of assortative mating among Z strains. Population zim49 showed the greatest assortative mating, and zim30 the lowest (Figure 2.1), but all were above those of the M strain (Figure 2.1).

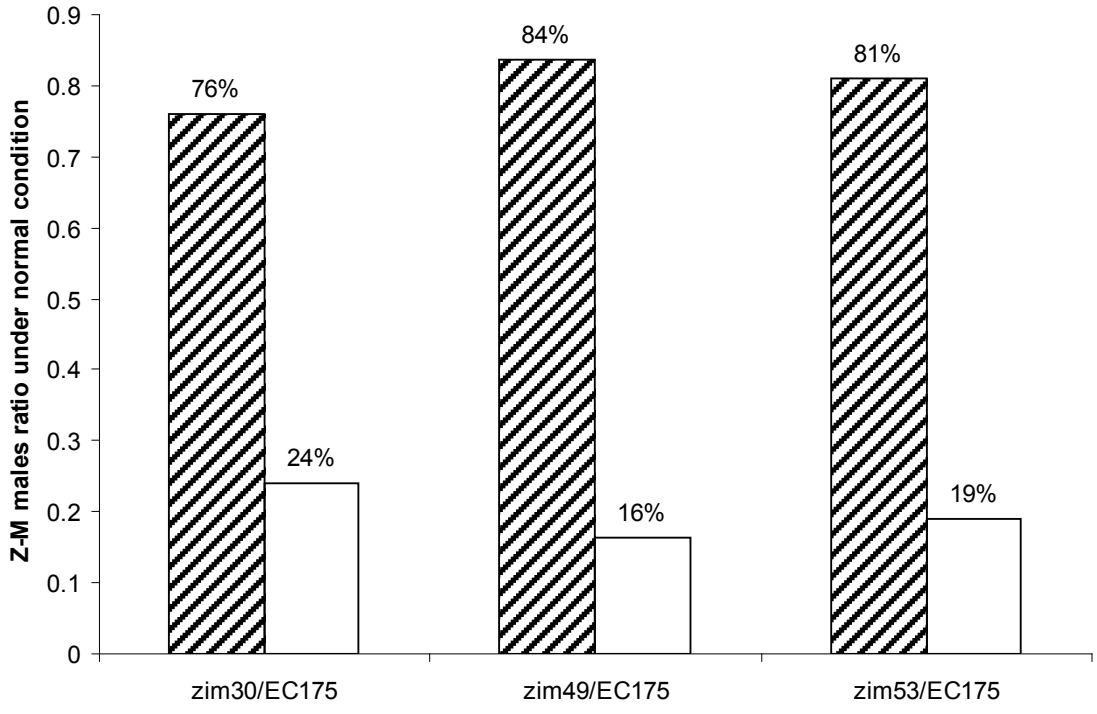


Figure 2.1 Count percentage of assortative mating for all three Z strains under normal conditions (Shaded box indicates monotypic mating (Z-Z), and blank box indicates heterotypic mating (Z-M)

Two-way ANOVA was performed and showed a significant effect ($F_{2,171}$=12.188, P<0.0001) for the three factors measured, and a significant effect ($F_{2,171}$=13.014, P<0.0001) of different strains and a significant factors × strains interaction effect ($F_{4,171}$= 4.414, P = 0.002) (Table 2.2). Post-hoc analysis revealed that auditory signals were significantly different from olfaction and vision effects (P < 0.001) but there was less difference between olfaction and vision (P=0.626) (Table 2.3). This result suggests that assortative mating within Z is largely suppressed after the wings were removed (Table 2.5). Post-hoc analysis also revealed that zim49 was significantly (P < 0.0001) different from the other two strains regarding the three factors measured(Table 2.4).

Table 2.2 ANOVA table of Z-M behavior for three factors among strains

Analysis of Variance

Source	Type III SS	df	Mean Squares	F-ratio	p-value
SPP$	0.634	2	0.317	13.014	0.000
FACTOR$	0.594	2	0.297	12.188	0.000
SPP$*FACTOR$	0.430	4	0.108	4.414	0.002
Error	4.165	171	0.024		

Table 2.3 Post-hoc test between each factor

Tukey's Honestly-Significant-Difference Test

FACTOR$(i)	FACTOR$(j)	Difference	p-value	95.0% Confidence Interval	
				Lower	Upper
olfaction	sound	0.119	0.001	0.053	0.186
olfaction	vision	-0.031	0.626	-0.101	0.040
sound	vision	-0.150	0.000	-0.215	-0.085

Table 2.4 Post-hoc test between each strain

Tukey's Honestly-Significant-Difference Test					
SPP$(i)	**SPP$(j)**	**Difference**	**p-value**	**95.0% Confidence Interval**	
				Lower	**Upper**
zim30	zim49	0.136	0.000	0.064	0.208
zim30	zim53	0.002	0.997	-0.070	0.075
zim49	zim53	-0.134	0.000	-0.196	-0.072

Table 2.5 Comparison of Z-Z/Z-M under normal condition and with treatments

	females	total females used	number of Z-Z/Z-M	preference on Z-Z(%)
normal conditions	zim30	135	41/13	75.9
	zim49	151	82/16	83.7
	zim53	219	64/15	81.0
olfaction	zim30	152	31/0	100.0
	zim49	349	31/3	91.2
	zim53	382	33/3	91.7
vision	zim30	223	18/3	85.7
	zim49	151	54/16	77.1
	zim53	269	41/15	73.2
audition	zim30	60	2/6	25.0
	zim49	404	32/5	86.5
	zim53	211	17/7	70.8

*(Preference on Z-Z (%) was calculated as # of females mated with monotypic males/ total # of females mated *100)*

To show the distribution of effects for each trait in three strains, students t-tests were performed.

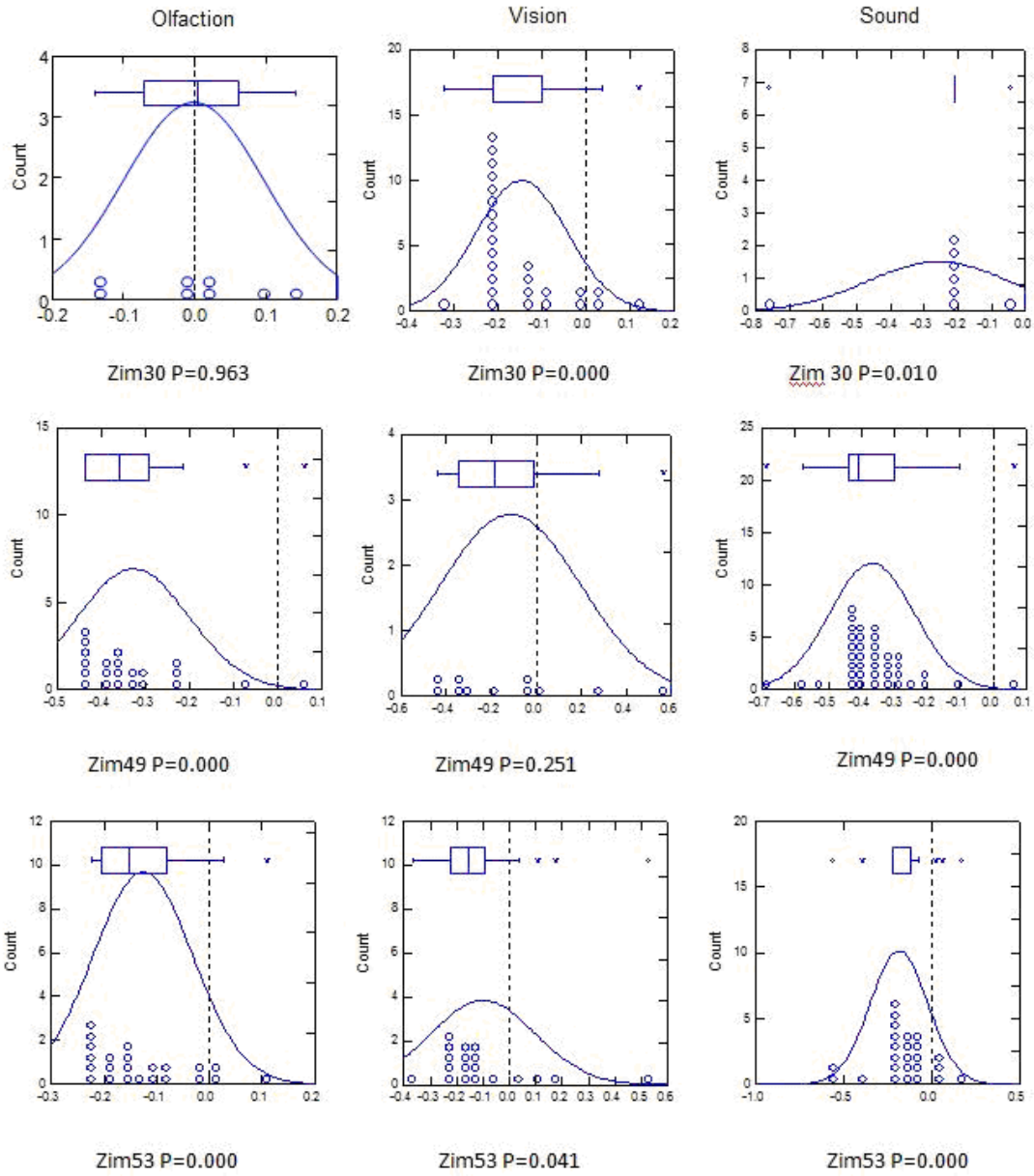


Figure 2.2 Distribution of d for all traits in three strains

(Ho: Mean = 0.000 against Alternative = 'not equal'
Each circle indicates one trial. The error bar represent the range of distribution with several dots indicate outliners.)

Table 2.6 Mean d value for each factor among three Zimbabwe strains (mean value calculated from tests above)

	olfactory	vision	**sound**
zim30	-0.002	-0.146	**-0.258**
zim49	-0.328	-0.116	**-0.368**
zim53	-0.126	-0.1	**-0.186**

All t-test results show negative values for the three factors in our study (Figure 2.2, Table 2.6). This occurred because all three cues are involved in assortative matings between Zimbabwe and cosmopolitan populations but the strength of each factor is different. Sound appears to contribute most to sexual isolation between Z-M. The vibration frequency of wings has been known to diverge between *D. melanogaster* populations (Iliadi *et al.* 2009). In the tropical environment, vibration changes may be more easily recognized by females than chemosensory or visual cues. Auditory isolation has been well documented in crickets (Hoy & Paul 1973, Shaw 1996), frogs (e.g., Ryan & Rand 1993), and birds (e.g., Irwin *et al.* 2001). The importance of auditory cues identified here does not necessarily mean that these are the most important factors in the initial sexual isolation process. The local environment may change through time, and therefore the strength of each trait may also change as a byproduct of adaptation to the environment at that particular time period. Also, reinforcement may play a role if sound helps to attract their monotypic females and thus avoid postzygotic reproductive costs.

The strengths of chemosensory and visual traits are variable among Z strains. Vision is the next-greatest cue of assortative mating in zim30 followed by olfaction. In zim49, olfaction is the second important cue followed by vision. In zim53, olfaction and vision are almost equal in importance. A similar polymorphism with respect to sexual preference for auditory factors was observed in *D. melanogaster* by Yukilevich and True (2008b). They found that females originating from the US preferred to mate with males that had higher wing vibration frequency. Females from Bahamas had weaker mating preferences for male wing vibrations, but actively rejected males presumably based on other sensory cues. This considerable variation of sensory characteristics lends support to the idea that the genetic basis in each case of incipient sexual isolation could

easily be different (Kilias *et al.* 1980; Dodd 1989). However, Ernst Mayr's statement that "if we were to rank the various isolating mechanisms of animals according to their importance, we would have to place behavioral isolation far ahead of all others" appears to hold true for the Zimbabwe female preference for Zimbabwe males and appears to define a potential species transition (Mayr 1963).

PART II

CHAPTER 3

GENOMIC SCREEN FOR CANDIDATE GENES CAUSING BEHAVIORAL ISOLATION BETWEEN ZIMBABWE AND COSMOPOLITAN *D. MELANOGASTER*

3.1 Introduction

Rapid differentiation and speciation are often associated with striking behavioral and morphological changes and at the same time limited mutations in functional genes. To explain this pattern, changes in regulatory regions have been proposed as a major source of biological differences between species, potentially also contributing to speciation (King and Wilson 1975). Identifying changes in gene regulation during the earliest stages of speciation therefore becomes an important direction for understanding how new species form.

Much evidence suggests that behavioral isolation is the first step in the formation of new species. First, behavioral isolation is consistently higher than postzygotic isolation among sympatric but not among allopatric species (Coyne and Orr 1989, 1997; Sasa *et al.* 1998). Second, reproductive isolation breaks down in the absence of behavioral isolation (Seehausen *et al.* 1997) and finally, comparative studies in insects and birds imply a positive correlation between indices of sexual selection and species richness (reviewed in Coyne and Orr 2004). Because of the importance of behavioral isolation to the formation of new species, it is then natural to explore variation in gene expression among populations that show signs of behavioral isolation.

In order to achieve this goal, it is important to distinguish expression variation related to genetic variation within a species (i.e. polymorphism), genetic variation between species (i.e. endogenous factors) and variation in expression that may occur due to the complex behaviors that isolate species (exogenous factors like sexual stimuli). A mechanistic link between altered gene expression and behavioral changes has been well established in a wide range of taxa at all three levels (Bucan and Abel 2002; Rankin 2002; Robinson *et al.* 2005). DNA microarray studies

have established that natural genetic variation produces significant differences in gene expression within a species at the polymorphism level (Cowles *et al.* 2002; Oleksiak *et al.* 2002; Rifkin *et al.* 2003; Cheung *et al.* 2003; Nuzhdin *et al.* 2004; Ranz and Machado 2006; Whitehead and Crawford 2006). At the endogenous level, novel social behavior in closely related whiptail lizards is associated with modified gene expression of a steroid hormone receptor in the brain (Woolley *et al.* 2004). Genome-wide expression profiling of relative transcript abundance in the brain indicated differences between domestic dogs and wolves (Saetre *et al.* 2004), as well as between humans and chimpanzees (Enard *et al.* 2002; Cáceres *et al.* 2003; Gu and Gu 2003; Preuss *et al.* 2004; Uddin *et al.* 2004). At the exogenous level, various labor occupations in the honeybee, such as nursing and foraging, are associated with massive changes in brain transcript levels (Whitfield *et al.* 2003a, 2003b; Grozinger *et al.* 2003; Goodman 2004; Cash *et al.* 2005). Transcription activation is crucial in mediating various forms of mate recognition as exemplified by changes in gene expression and neural circuits in the forebrain of zebra finches presented with conspecific and heterospecific birdsongs (Mello *et al.* 1992; Jarvis *et al.* 2002; Clayton 2004). Sexual experience was observed to dramatically affect expression profiles in the brain of female hamsters as well (Bradley *et al.* 2005). A microarray comparison between mated and nonmated females and between courted (exposed to males, but nonmated) females and unexposed females in *D. melanogaster* detected dozens of differentially expressed genes (Lawniczak and Begun 2004).

Substantial variation in gene expression among natural populations, between biological species, and the clear link between sexual behavior and transcriptional alterations, allow us to hypothesize that gene expression changes may also underlie behavioral isolation. Two morphs of *Drosophila melanogaster* (Cosmopolitan and Zimbabwe) offer a model system to address variation in gene expression as it relates to behavioral isolation. *Drosophila melanogaster* Cosmopolitan and Zimbabwe morphs provide an example of incipient speciation (Wu *et al.* 1995; Hollocher *et al.* 1997a,b; Ting *et al.* 2001) because while there is little deviation from random mating between individuals from various Cosmopolitan populations (Henderson and Lambert

1982, but see Capy *et al.* 2000 and Korol *et al.* 2000), females from some Zimbabwean populations (Sengwa Wildlife and Harare) preferentially mate with their own males and discriminate against Cosmopolitan males. Subsequent experiments with Zimbabwean flies found considerable polymorphism in mating behavior, ranging from pure M-type ("Cosmopolitan"-like) to pure Z-type ("Zimbabwe") and mapped the trait to several genomic regions with the largest impact on the third chromosome (Hollocher *et al.* 1997a,b; Ting *et al.* 2001). Therefore, we hypothesized that fixed differences in gene expression as well as population specific patterns of gene expression due to external stimuli (i.e. exposure to mates) cause behavioral differences that are evolutionarily important.

We used DNA microarrays to identify genes differentially expressed between Cosmopolitan and Zimbabwe populations of *Drosophila melanogaster* and simultaneously, differentially expressed due to mating behavior. This transcriptomics-based approach provides a powerful entrée towards discovery and functional characterization of genes responsible for incipient stages of speciation. Thus far, there have been no such genes identified despite the fact that estimates of their numbers were obtained from genetic mapping and there have been a number of genes identified that are responsible for hybrid defects (reviewed in Orr *et al.* 2004 and Michalak and Noor 2006). Comparing expression profiles from the heads of mated and nonmated females, we identified a number of genes likely associated with mating (and postmating) effects.

3.2 Materials and Methods

3.2.1 Fly stocks

Stocks were reared in uncrowded cultures at 24°C with a 12-h light-dark cycle on Carolina 4-24® formula. Three stocks of *Drosophila melanogaster* Cosmopolitan were used as biological replicates: E10019 from Fukui, Japan (Ehime Drosophila Stocks), EC175 from Ecuador (courtesy of Esther Betrán), and 14021-0231.23 from Crete (Tucson Stock Center, TSC). Four stocks of *Drosophila melanogaster* Zimbabwe were used (courtesy of Jerry Coyne and Tami Panhuis): Z29, Z30, Z49, and Z53.

3.2.2 Experimental mating

Although the pattern of asymmetric isolation between M and Z types has been found highly repeatable by independent researchers (e.g., Wu *et al.* 1995; Hollocher *et al.* 1997a-b; Alipaz *et al.* 2001; Ting, Takahashi, and Wu 2001; Fang *et al.* 2003; Panhuis *et al.* 2003; Takahashi and Ting 2004), we conducted 43 female multiple choice mating experiments to confirm that it persists in our laboratory. During each mating experiment, n (>20) 7-day old females from the same stock were given a choice between n conspecific and n heterotypic males. Z and M males were marked by a dot on their scutella with red and blue colored sharpie markers. Colors were alternated between experiments but they were found to have no effect on mating frequencies. Replicates were aggregated together within each female type and organized into contingency tables.

3.2.3 RNA collection

For RNA collection, virgin females were collected, separated, and aged for 7 days. On day 7, half of them were transferred to bottles with males for mating and the other half used as nonmated controls. Copulating pairs were aspirated into individual vials and then males were removed after 30 min and simultaneously frozen with females along with corresponding nonmated control females for comparison. This scheme ensured that female age and time of exposure to treatments were identical in both groups. RNA was extracted from heads only, after 20 females per sample were collected. At least five micrograms of total RNA were extracted using RNApure and following the manufacture's protocol (GenHunter Corp., Nashville, TN). After purification with GeneChip Sample Cleanup Modules (QIAGEN), concentrations were measured with the NanoDrop® ND-1000 UV-Vis and checked for integrity via denaturing gel electrophoresis following the protocol of Liang and Pardee (1997).

3.2.4 Affymetrix Drosophila *GeneChip® profiles*

A total of 13 Affymetrix Drosophila GeneChip® arrays (9 for Z and 4 for M females) were used. Double-stranded cDNA was synthesized with a T7-(dT)24 primer; cRNA was synthesized and biotin-labeled in an in vitro transcription reaction using the ENZO BioArray HighYield RNA

Transcript Labeling Kit. The target cRNA was hybridized to Affymetrix GeneChip® Drosophila Genome Arrays that allow assays of the relative abundance of >13,500 mRNA transcripts. Probes for a particular transcript are present in 28 cells on the array chip. Half of those cells possess Perfect Match (PM) sequences that are complementary to the reference transcript. The other half possesses Mismatch (MM) sequences that are complementary to the reference transcript except for a homomeric base mismatch at the 13^{th} position. The number of instances in which the PM hybridization signal is larger than the MM signal is computed along with the average of the logarithm of the PM:MM ratio (after background subtraction) for each probe set using Affymetrix Microarray Suite (MAS) 5.0 software. The software Detection algorithm calculates the Discrimination Score from each probe pair (PM vs. MM), assesses probe saturation, calculates a Detection P value from a one-sided Wilcoxon Signed Rank test, and assigns a Present, Marginal, or Absent Call in relation to a threshold value Tau=0.015.

3.2.5 Gene expression analysis

Prior to the statistical analysis, MAS 5.0 data were scaled to 10,000 and $\log_2$-transformed. Additionally, Affymetrix CEL files were reanalyzed with DNA-Chip Analyzer (dChip; Li and Wong 2003). dChip is a software package that implements a statistical model-based expression analysis to account for probe variability and enables both low-level and high-level analyses (see Li and Wong 2003 for details). Expression values were generated using Li-Wong transformation and normalization (Li and Wong 2001a, 2001b). dChip normalizes all arrays to a common baseline array having the median overall brightness (Li and Wong 2003). Normalization is based on an invariable set of probes identified through iterative procedures. The CEL images were visually checked for local contamination and data were additionally filtered to include only those 3237 transcripts that had more than 50% present calls among all samples.

ANOVA contrasts implemented in dChip were used to find the intersection of two candidate gene lists, one differentiating all Z and M and the other differentiating mated and nonmated Z females. Candidate transcripts from the intersection represent mating-related genes that significantly differ between the Z and M morphs. They may include genes involved in sexual

discrimination but transcripts underlying postmating effects will likely be included as well. We focused on contrasting mated and nonmated Z females, as we reasoned that comparing females from the discriminating group would be more informative than those from the nondiscriminating group (M). Although expression profiles of M females may also be changed due to their mating status, this change will not be associated with mating discrimination against foreign males (in contrast to Z females), and as such will fail to contribute to phenotypic (behavioral) differences between Z and M groups. Therefore, Z females were mostly compared with nonmated M females to assure a cost-effective approach. We did not address the problem of gene expression polymorphism within M or Z groups. Instead, we aimed to detect inter-group differentiation between M and Z. For that reason, we used geographically distant M populations to provide a wide coverage of genetic differentiation and thus assure a statistically conservative approach. The moderated t-statistic based on an empirical Bayes method and implemented in the limma package (Irizarry 2005) was used to estimate the false discovery rate (FDR) and FDR-adjusted P-values. Additionally, the SAM thresholding procedure (Tusher *et al.* 2001) was used to generate the plots for FDR (Figure 3.1).

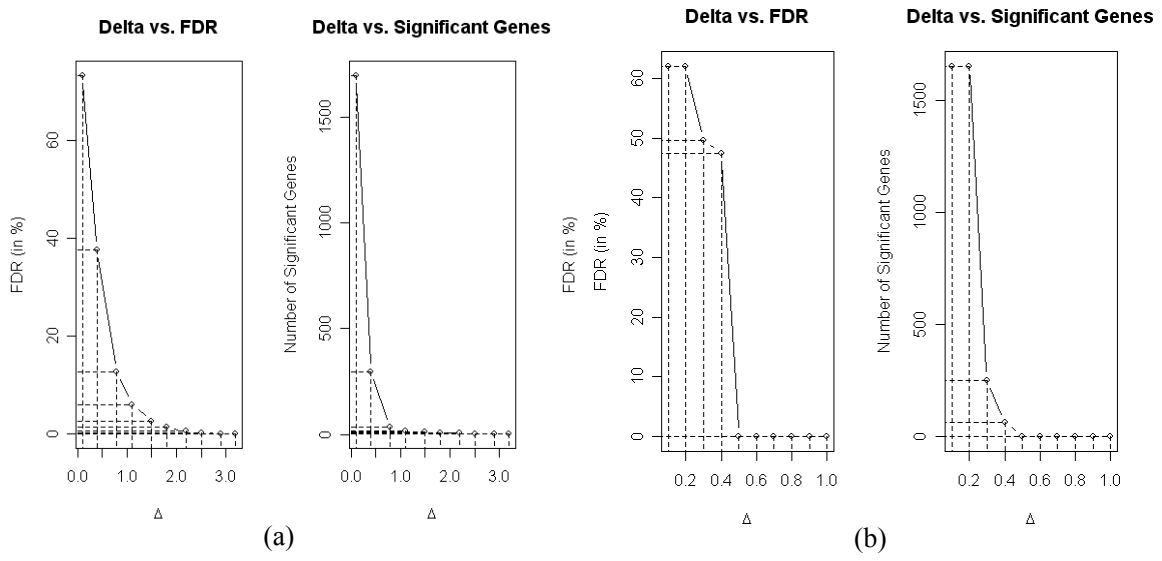


Figure 3.1 SAM plot of the false discovery rate (FDR) against the fixed rejection region (delta) and the number of significant genes for two contrasts: (a) Zimbabwe versus Cosmopolitan and (b) mated versus nonmated Zimbabwe.

Hierarchical clustering was used to infer patterns of coregulation among the candidate genes. The distance between two genes is defined as 1 - r where r is the Pearson correlation coefficient between the standardized expression values (mean 0 and standard deviation 1) of the two genes across the samples used (Li and Wong 2003). Standardization and clustering methods followed Golub *et al.* (1999) and Eisen *et al.* (1998). To detect overrepresented functional groups, information from the Gene Ontology (GO) database (Gene-Ontology-Consortium 2001) was integrated with the expression data using the EASE application (Hosack *et al.* 2003), and probability values were assigned to each observed GO representation relative to GO fractions for all Affymetrix probe sets.

3.2.6 Quantitative RT-PCR

Based on their statistical significance and potentially relevant functional classification, four candidate genes (Table 3.2) were selected for quantitative real-time fluorescent polymerase chain reaction (QRT-PCR) analysis to confirm the microarray results with independent RNA sampling and methodology. For QRT-PCR, 100 ng of total RNA prepared as described above were reverse transcribed and amplified in two-step reactions using Promega Access RT-PCR system (cat. # A1250) and Promega protocols. QRT-PCR reactions contained 5 µL AMV/ Tfl5 Reaction Buffer, 1 µL 25mM $MgSO_4$, 1.5 µL 0.0001% SYBR Green, 1 µL 20 µM downstream and upstream primers each, 0.5 µL dNTP mix (10 mM each), 0.5 µL TflDNA Polymerase, 0.5 µL AMV Reverse Transcriptase, and complemented to a total of 25 µL with DEPC-treated water. All PCRs were run on the same BIO-RAD iCycler with MyiQ^{TM} Optical Module, and included an initial 45 min at 48ºC for reverse transcription, 2 min at 95ºC, 35 cycles of 95ºC for 30 s, 55ºC for 1 min and 72ºC for 2 min. CT-values of the target gene were corrected by β-Actin values as a normalizer. To test for contamination with genomic DNA, negative controls for each sample contained RNA and all other components, excluding the AMV Reverse Transcriptase. Three-way ANOVA of threshold cycle (C_T) values was used to analyze differential expression with Group (M and Z), mating status, and PCR reaction as fixed factors.

Table 3.2 Differences in threshold cycle (C_T) values from quantitative RT-PCR for four candidate genes. Lower C_T-values represent higher expression levels. Results of 3-way ANOVA are shown, with the PCR reaction effect and interactions (not significant) not shown. The Z-M difference is subtraction of mean C_T-value for Cosmopolitan from the mean C_T for Zimbabwe. The nv-v is between mean C_T-values for mated (nv) and nonmated (v) Zimbabwe females.

Gene	df	Z-M difference	F	P-value	nv-v difference	F	P-value
Odorant Receptor 63a	1, 17	0.135	50.33	0.0001	0.08	12.85	0.002
desaturase2	1, 15	0.1	9.64	0.007	-0.002	0	P>0.05
CG10120	1,8	-0.072	5.51	0.047	-0.014	0.21	P>0.05
CG1812	1,16	-0.009	0.07	P>0.05	0.048	1.81	P>0.05

3.3 Results

Intergroup mate choice tests confirmed the expected pattern of asymmetric behavioral isolation between Z and M (Figure 3.2). All 19 replicates of tests with Z females exposed to both types of males resulted in preference for conspecific males (range 60-100%), compared to 24 replicates of tests with M females out of which 10 resulted in preference for heterotypic males (<84%), 6 for conspecific males (<71%), and 8 no preference (50% of each type).

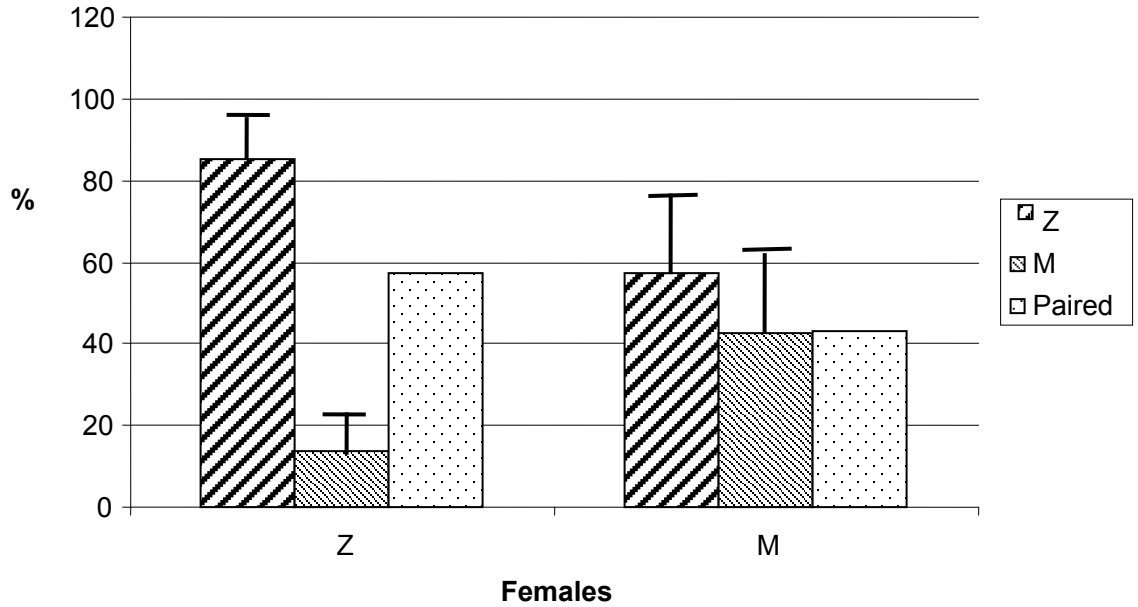
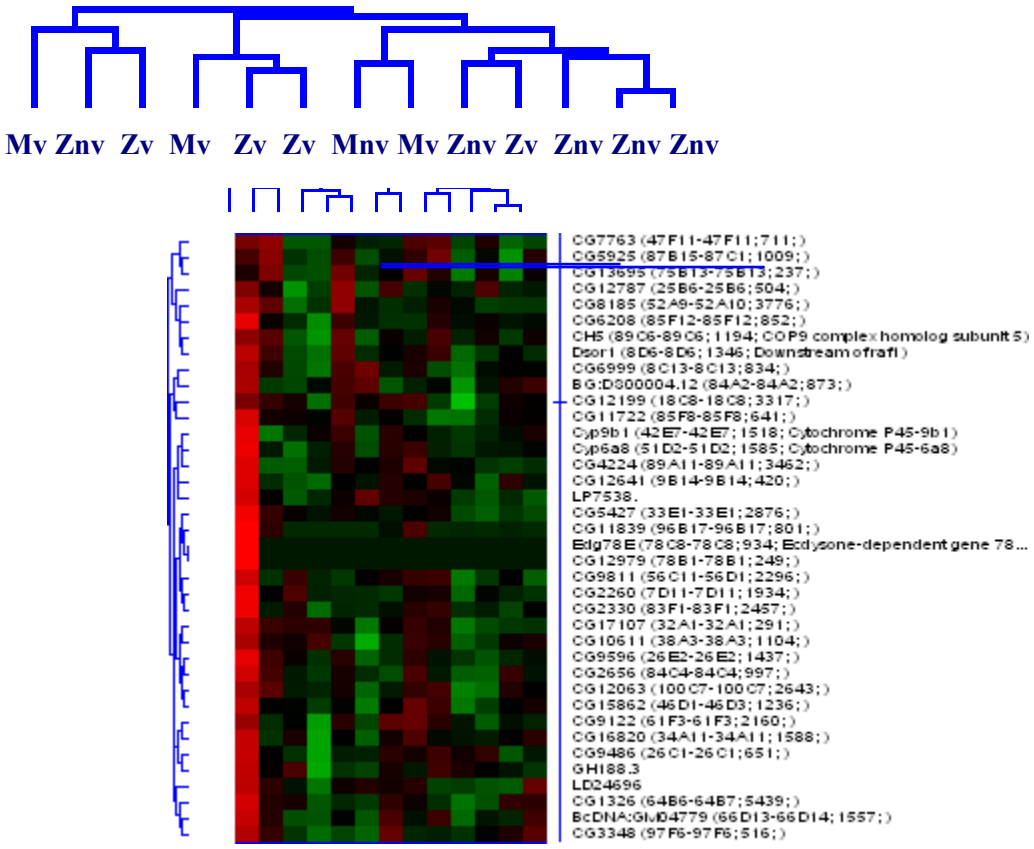


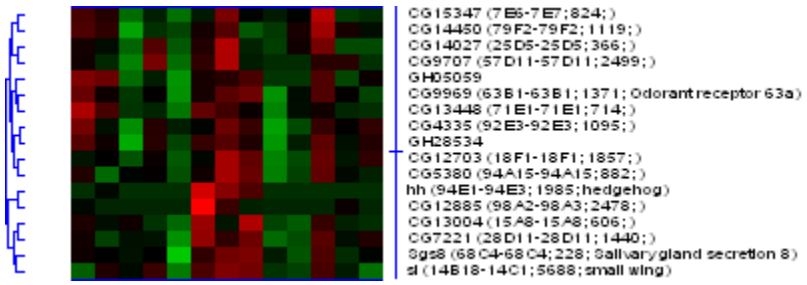
Figure 3.2 Average mating frequencies from the multiple choice mating tests. Third bar (Paired) represents percent of females paired across all replicates within each group.

We obtained four nonmated and five mated Z transcriptome profiles compared with three

nonmated and one mated M profiles. Knudsen (2002) showed that four replicates are sufficient to keep the rate of false positives low. Replicates were biologically diversified to represent variation within group (M stocks originated from different continents). We also reasoned that pooling of mRNA from multiple females (see Materials and Methods) effectively increases the level of biological replication and reduces variability between arrays (see also Kendziorski et al. 2005). We found that 1065 genes were differentially expressed between Z and M, 609 genes were differentially expressed between mated and nonmated Z, and only 45 genes were simultaneously different ($P < 0.05$) in both contrasts (Table 3.1). The intention behind this two-way contrast was to identify Z-specific transcript changes associated with mating effects, thus manifest in mated Z females but not nonmated Z and any M. We reasoned that if expression of a gene responsible for sexual discrimination in Z females is modulated through mating, nonmated Z (rather than mated Z) females will be more M-like. Thus of particular interest are those 17 genes, such as Odorant receptor 63a, whose expression difference between mated and nonmated Z was in the same direction, respectively, as between Z and M (Figure 3.3).



(a)



(b)

Figure 3.3 Hierarchical clustering of genes that were differentially expressed between Zimbabwe (Z) and Cosmopolitan (M) and simultaneously between mated (nv) and nonmated (v) Zimbabwe females: (A) Odorant receptor 63a and (B) desaturase2 (CG5925). Clustering at the top of the figure groups treatments (morphs and mating status), whereas clustering on the right panel of the figure groups genes with similar expression levels, ranging from overexpression (red color) to underexpression (green color).

We assessed whether probe sets with significantly altered transcript abundance due to our experimental conditions were randomly distributed across the five major chromosome arms.

We used a chi square goodness-of-fit test to check for significant deviation of observed from expected numbers of affected loci on each chromosome. By conventional criteria ($P < 0.05$), this deviation could not be considered significant ($\chi^2 = 9.387$, df = 4, $P = 0.0521$; after including the fourth chromosome the significance decreases even further, $\chi^2 = 10.181$, df = 4, $P = 0.070$). Three biological processes (physiological, cellular, and development), four molecular functions (catalytic activity, binding, transporter activity, and transcription regulator activity), and one cellular component (cell) were present among GO groups represented by the 45 genes, but none of them were statistically overrepresented according to EASE ($P < 0.05$).

Using QRT-PCR, we tested expression changes in four candidate genes: *Odorant Receptor 63a*, *desaturase2*, *CG10120*, *CG1812* (Table 3.2). We confirmed that *Odorant Receptor 63a* was downregulated in Z females relative to M and at the same time suppressed in mated Z females compared with nonmated Z females. For *desaturase2* and *CG10120*, we validated the difference between Cosmopolitan and Zimbabwe to be consistent with microarray results (Figure 3.4) but there was no significant difference between mated and nonmated Z females. Neither contrast was statistically significant for *CG1812* (Table 3.2).

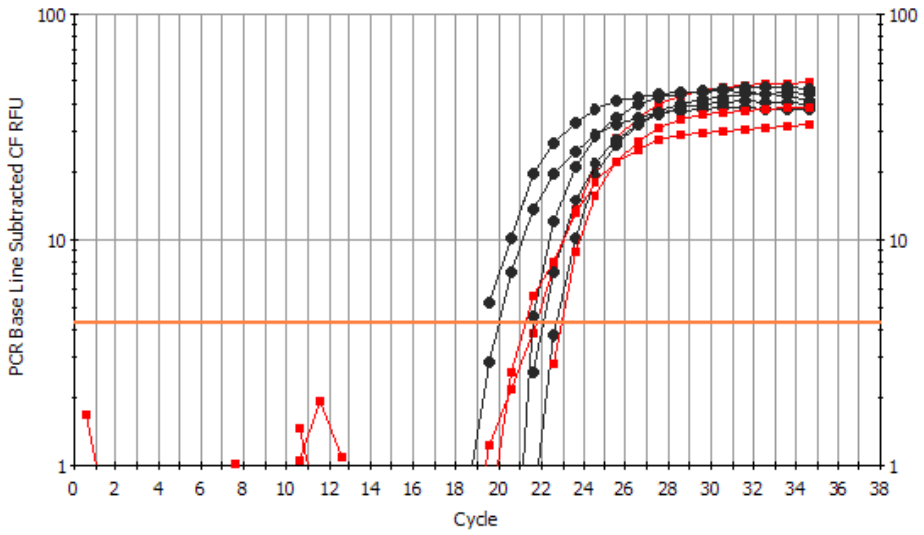


Figure 3.4 An example of quantitative real-time PCR amplification of desaturase 2. Black curves with circles represent Zimbabwe samples and red lines with squares correspond to Cosmopolitan samples.

Table 3.1 Genes differentially expressed between Zimbabwe (Z) and Cosmopolitan (M) and simultaneously affected by the mating experience (v – virgin Z, nv – nonvirgin Z). Genes chosen for the QRT-PCR analysis are denoted by a dagger (†) and P-values <0.05 after adjustment for FDR are denoted an asterisk (*).

Locus	Activity description	Z/M fold change	P-value (Z – M contrast)	nv/v fold change	P-value (nv – v contrast)
desaturase2†	fatty acid desaturase	1.87 3 Z > M	0.009*	1.655 nv < v	0.006
CG11357	enzyme	1.07 2 Z > M	0.004	1.053 nv < v	0.049
CG12400	NADH dehydrogenase	1.05 2 Z < M	0.024	1.034 nv < v	0.026
CG11836	endopeptidase	1.64 5 Z < M	0.001	1.140 nv > v	0.016
CG15547	enzyme	1.12 2 Z < M	0.013*	1.125 nv > v	0.004*
CG10120†	deoxyribonuclease	1.12 6 Z < M	0.003*	1.179 nv > v	0.0004
CG4199	enzyme	1.02 6 Z < M	0.036*	1.032 nv > v	0.0007*
CAH1	carbonate dehydratase	1.20 3 Z > M	0.034	1.219 nv < v	0.002
CG3027	beta-ureidopropionase	1.02 6 Z < M	0.009*	1.020 nv < v	0.020
Aats-gln	Glutaminyl-tRNA synthetase	1.27 1 Z > M	0.044	1.353 nv < v	0.009
SamDC	adenosylmethionine decarboxylase	2.04 0 Z < M	0.001	1.465 nv > v	0.025
Rpn11	proteasome regulatory particle	1.48 9 Z < M	0.004	1.307 nv < v	0.017
Arf72A	ARF small GTPase	1.01 6 Z < M	0.029	1.021 nv > v	0.023
Elk	potassium channel	11.05 Z > M	0.0004	1.815 nv < v	0.042
Pi3K68D	Phosphotidylinositol 3 kinase	1.26 8 Z > M	0.023	1.292 nv > v	0.035

Table 3.1 Continued

CG15088	neurotransmitter transporter	1.036 Z < M	0.006	1.036 nv < v	0.034
CG7050	Neurexin, cell adhesion	1.858 Z < M	0.004*	1.662 nv < v	0.028
Cyp4p2	cytochrome P45	1.026 Z < M	0.039*	1.029 nv > v	0.044*
CHORD	Zink ion binding	1.245 Z < M	0.023	1.392 nv < v	0.028
CG1812[†]	transcription factor	1.296 Z < M	0.000001*	1.124 nv > v	0.008*
CG14730	DNA binding	2.034 Z < M	0.019	4.675 nv < v	0.030
Bowl	RNA polymerase II transcription factor	1.077 Z < M	0.003*	1.076 nv > v	0.039
Trithorax	DNA binding	1.089 Z < M	0.011*	1.098 nv > v	0.048*
prospero	RNA polymerase II transcription factor	1.042 Z < M	0.023	1.049 nv > v	0.012*
CG14962	nucleic acid binding	1.033 Z < M	0.010	1.036 nv > v	0.020
Gcn2	elongation-factor-2 kinase	1.391 Z < M	0.004	1.390 nv > v	0.043
CG14641	RNA binding	1.281 Z < M	0.008	1.231 nv > v	0.018
CG15636	DNA binding	1.029 Z > M	0.010	1.036 nv > v	0.036
Schnurri	RNA polymerase II transcription factor	1.633 Z > M	0.030	1.433 nv > v	0.023
Odorant receptor 63a[†]	olfactory receptor	1.856 Z < M	0.009*	2.420 nv < v	0.026
CG3814	N-methyl-D-aspartate receptor-associated protein	1.449 Z < M	0.001	1.268 nv > v	0.046

Table 3.1 Continued

CG17988	---	1.036 Z < M	0.009	1.033 nv > v	0.009
CG13339	---	1.040 Z > M	0.039	1.049 nv < v	0.015
CG10127	---	1.03 Z > M	0.011	1.034 nv < v	0.023
CG4398	---	1.198 Z > M	0.0003*	1.056 nv < v	0.014
CG13301	---	1.058 Z < M	0.046	1.083 nv > v	0.022
CG8844	---	1.027 Z < M	0.009*	1.033 nv < v	0.035
CG7530	---	1.027 Z < M	0.019	1.028 nv > v	0.020*
Msta	---	1.350 Z > M	0.004	1.207 nv < v	0.020
CG3173	---	1.017 Z > M	0.002	1.014 nv > v	0.045
CG14899	---	1.401 Z > M	0.017	1.373 nv > v	0.008*
CG15929	---	1.098 Z > M	0.017*	1.093 nv < v	0.040
CG12765	---	1.037 Z < M	0.039*	1.024 nv < v	0.037
CG14701	---	1.032 Z > M	0.013	1.027 nv > v	0.041
CG17496	---	1.043 Z > M	0.049	1.071 nv > v	0.032

3.4 Discussion

Behavioral isolation in *Drosophila* is exerted through courtship behavior, consisting of sequential actions that exchange auditory, visual, and chemosensory signals between males and females (e.g., Greenspan and Ferveur 2000; Markow and O'Grady 2005). Given the complexity of mating behavior, it should not be surprising that candidate genes represent a variety of biological functions. One of the candidate genes, *Odorant receptor 63a*, plays a critical role in the olfactory system and mate recognition in *Drosophila*. Odorant receptors constitute the molecular basis for the detection of volatile odorous molecules and the perception of smell (Clyne *et al.* 1999; Vosshall *et al.* 1999, Vosshall *et al.* 2000). The members of the odorant receptor gene family (~60) in the *Drosophila* genome are considerably divergent, with an average amino acid identity of ~20%, and each of them encodes a putative seven-transmembrane domain protein of about 380 amino acids (Vosshall *et al.* 2000). Odorant receptors are expressed in small subsets of olfactory receptor neurons in the olfactory sensory organs of adult *Drosophila*, mostly the antenna and the maxillary palp (Clyne *et al.* 1999). Despite the extreme confinement and patchiness of expression, we were able to validate the microarray expression pattern of *Odorant receptor 63a* with quantitative RT-PCR (Table 3.2). This gene was downregulated in Z females relative to M and at the same time suppressed in mated Z females compared with nonmated Z females (Figure 3.3a).

Another receptor, *Nmda1* (N-methyl-D-aspartate receptor-associated protein), represents a class of glutamate receptors that are of central importance in synaptic plasticity. NMDA receptors allow Ca^{2+} influx and are thought to trigger Ca^{2+} dependent postsynaptic processes involved in long term potentiation and depression (Sucher *et al.* 1996). This gene has been reported among transcripts typical for neuroblast lineage development (Brody *et al.* 2002). Receptor activity was also changed at *neurexin* putatively involved in cell adhesion (Arbeitman *et al.* 2004). Two other genes with changes in expression, *elk* and *CG15088*, are involved in potassium transport. In mammals, sexual interactions have been known to cause extensive changes in channels, signal transduction and neurotransmission (Bradley *et al.* 2005).

Notably, we found that transcription of *desaturase2* (*CG5925*), a gene with stearoyl-CoA

9-desaturase activity involved in cuticle hydrocarbon biosynthesis, was also altered in our experimental conditions. Its regulation pattern clustered together with *CG7763* (putative sugar binding activity), *geko* (*CG13695*), *hoe1* (*CG12787*, producing a membrane component with transporter activity), CG6208 (soluble NSF attachment protein activity involved in vesicle-mediated transport), *CH5* (NEDD8 activating enzyme involved in axonogenesis), and *Dsor1* involved in signal transduction (Figure 3.3b). *Desaturase2* has been suggested to be involved in differential adaptation to climate as well as behavioral isolation between Z and M lines (Fang *et al.* 2002; Greenberg *et al.* 2003; Greenberg *et al.* 2006, but see Coyne and Elwyn 2006a,b). Moreover, functional variation at the *desaturase2* locus has a distinct geographic pattern: females from African (including Z) and Caribbean populations produce 5,9-heptacosadiene (5,9-HD) as the predominant cuticle hydrocarbon. This product seems to be absent from most Cosmopolitan females as a result of a 16-bp deletion at the 5' end of the gene, largely contributing to the polymorphism of cuticle hydrocarbons (Dallerac *et al.* 2000; Takahashi *et al.* 2001). The loss-of-function allele is widely distributed over the world and the nucleotide diversity at the locus as well as experimental data suggest that this spread may have been driven by positive selection for increased cold resistance (Takahashi *et al.* 2001; Greenberg *et al.* 2003; Greenberg *et al.* 2006, but see Coyne and Elwyn 2006a,b). Although downregulated relative to Z females, *desaturase2* transcripts were still present in M females in contrast to the report by Dallerac *et al.* (2000). Using RT-PCR, we confirmed a difference between Cosmopolitan and Zimbabwe consistent with the results of microarray analysis (Tables 1 and 2, Figure 4) but the difference between mated and nonmated females was not significant.

In addition to *desaturase2*, transcription levels of 10 other enzymes were related to divergence between the M and Z morphs, including *CAH1* with carbonic anhydrase activity and *Pi3K68D* with phosphatidylinositol-4-phosphate 3-kinase activity. Seven transcription factors or other DNA binding gene products were altered in the experimental conditions. These include *trithorax* which is required to maintain the proper spatial pattern of expression for multiple homeotic genes of the Bithorax and Antennapedia complexes (Kuzin *et al.* 1994). *Trithorax* has also been reported to interact genetically with *brm*, *ph-p*, *Pc*, *mod(mdg4)*, *Asx* and 32 other genes

(FlyBase Report). However, none of these occurred among the candidate genes from Table 1 or at the cluster of coregulated transcripts (not shown). One of the two most closely coregulated genes with *trithorax* was *CG6701* that has two conserved domains related to DNA and RNA helicases and helicase subunits (DNA replication, recombination, and repair; Celniker *et al.* 2002). The other tightly coregulated gene (*CG3558*) has not been functionally characterized. In adults, *trithorax* is known to affect the development of sex combs, which are male specific chaetae located on the prothoracic tarsal segment of the prothoracic leg, in addition to influencing development of at least eight other tissues (FlyBase).

In more than half the cases (29 out of 45 candidate genes), including *desaturase2*, expression of mated Z females moved in the direction of M females. If expression changes were to be associated with mating discrimination in Z females, one would expect to record the opposite pattern, with differences between Z and M females enhanced after Z females are mated. The simplest explanation of this inconsistency is that although these 29 genes are simultaneously differentiating Z/M and are affected by mating, they are not directly involved in sexual discrimination. An alternative explanation may be that observed expression changes are a part of a more complex regulation circuitry in which certain modules must be repressed in order to activate other modules in response to a common regulatory switch.

At least three loci on the third chromosome have been mapped for female discrimination between Zimbabwe and Cosmopolitan populations (Ting *et al.* 2001) and this raises a question about the genomic distribution of expression changes we observed. Sexual conflict provides a plausible mechanism leading to a nonrandom distribution of genes with sex-biased patterns of expression. Specifically, the evolution of sex-biased gene expression may alleviate sexual antagonism which has been shown to be extensive in adult *Drosophila* (Chippindale *et al.* 2001). Transcription level is an excellent measure of sex-biased function and global expression profiles do provide evidence of extensive sex-biased expression in *Drosophila* and *C. elegans* (Andrews *et al.* 2000; Reinke *et al.* 2000; Jiang *et al.* 2001; Jin *et al.* 2001; Swanson *et al.* 2001; Arbeitman *et al.* 2002; Kelly *et al.* 2002; Meiklejohn *et al.* 2003; Parisi *et al.* 2003; Ranz *et al.* 2003, Michalak and Noor 2003). As the X chromosome is depauperate for male-specific genes probably due to

demasculinization or feminization effects of selection (Parisi *et al.* 2003; Ranz *et al.* 2003; Oliver and Parisi 2004; Mackay *et al.* 2005), it is thus plausible that X chromosome-linked transcripts have a disproportionate contribution to sexual isolation. However, we observed no chromosomal overrepresentation of analyzed loci across the genome.

Our results also suggest substantial epistasis (many interacting genes involved) and pleiotropy (same genes, such as *trithorax*, affect multiple traits) underlying differences between populations of *D. melanogaster*. Further studies are evidently required to determine whether any candidate genes reported here are directly responsible for behavioral isolation between Cosmopolitan and Zimbabwe groups. The genetic architecture underlying behavioral isolation may be extraordinarily complex, as exemplified by mapping genes responsible for reinforcement in *Drosophila pseudoobscura* and *D. persimilis*; Ortiz-Barrientos *et al.* (2004) found that the genetic architecture of basal female mating discrimination (between allopatric populations) is different from that of reinforced mating discrimination (between sympatric populations).

In sum, we demonstrated significant differentiation of gene expression in two major morphs of *Drosophila melanogaster*, Cosmopolitan and Zimbabwe. As some of those differences were additionally modified by female mating status, we believe that the list of identified candidate genes may be very useful in further studies of sexual isolation between these two *Drosophila melanogaster* morphs. Prior to our study, only one gene, *desaturase2*, has been invoked to play a role in sexual isolation between Cosmopolitan and Zimbabwe flies. Remarkably, this gene was among 44 other candidate genes detected here with microarrays. However, there are several limitations in the presented analysis, such as relatively high FDR resulting from the moderate and unbalanced sampling and rather low fold changes. Also due to the experimental setup, expression changes in the candidate genes likely represent postmating effects, such as insemination, in addition to mating discrimination.

PART III

CHAPTER 4

GENE *ACYLPHOSPHATASE* (*ACYP*) IS ASSOCIATED WITH HYBRID MALE STERILITY

4.1 Introduction

According to the biological species concept (BSC), species is defined as a population of individuals reproductively isolated from individuals of other populations through a combination of extrinsic and intrinsic isolating mechanisms (Mayr, 1942). While extrinsic mechanisms, such as geographical barriers were relatively well understood, the nature and the origin of intrinsic isolating mechanisms have been poorly characterized. Most studies focused on genetics of hybrid dysfunctions, such as sterility and inviability, as critical but not exclusive components of intrinsic postzygotic isolation leading to speciation (Coyne and Orr, 2004). Despite this focus, progress in identifying genes responsible for reproductive isolation was slow and disappointing, a situation perhaps not unexpected given that genetic crosses between different species are almost by definition difficult or impossible. As a result, only nine hybrid inviability/sterility genes have been thus far documented in animals, five of them in *Drosophila* (Wittbrodt *et al.* ., 1989; Ting *et al.* ., 1998; Barbash *et al.* ., 2003; Presgraves *et al.* ., 2003; Masly *et al.* ., 2006; Bomblies *et al.* 2007; Lee *et al.* 2008; Mihola *et al.* 2009; Phadnis and Orr 2009).

Recent developments in comparative genetics and genomics have provided new tools for explorations into the genetic background of reproductive isolation. For example, gene regulation changes and epistatic interactions (such as Dobzhansky-Muller incompatibilities) were long postulated to play a critical role in speciation and species divergence (Dobzhansky, 1937; Muller, 1940; King and Wilson, 1975), but their systematic examination at the genome-wide level had to await the renaissance of reverse genetics and the emergence of microarray technology in particular (Michalak and Noor, 2003; Ranz *et al.* ., 2004; Hegarty *et al.* ., 2006; Josefsson *et al.* .,

2006; Lai *et al.* ., 2006; Ranz and Machado, 2006; Ortíz-Barrientos *et al.* ., 2007; Barbash and Lorigan, 2007; Michalak *et al.* ., 2007). This new transcriptome-oriented approach starts at the level of phenotype and compares transcriptomes of parental species with transcriptomes of their dysfunctional hybrids and leads to the identification of the candidate loci that contribute to reproductive isolation or/and divergence between parental species.

Transcriptome profiles of *Drosophila* hybrids relative to the parental species have consistently shown substantial gene misexpressions associated with hybrid sterility (Michalak and Noor, 2003; Ranz *et al.* ., 2004; Haerty and Singh, 2006; Moehring *et al.* ., 2006). In a follow-up study, Michalak and Noor (2004) used quantitative RT-PCR analysis and a backcross breeding scheme to confirm misexpression of the top five candidate genes from their earlier microarray study in *Drosophila simulans*, *D. mauritiana* and their sterile hybrid males (Michalak and Noor, 2003). In F1 generation from a cross between *D. simulans*, *D. mauritiana*, males are sterile and females are fertile but male fertility can be restored in a fraction of backcross hybrids. Michalak and Noor (2004) backcrossed hybrid females to *D. simulans* for five generations and found that four out of the five genes previously known to be misexpressed in F1 males relative to the parental males were also misexpressed in sterile compared with fertile backcross hybrid males. Additionally, they found that male sterility in the hybrid backcross males was tightly associated with the genotype at *Odysseus* locus: all fertile males were homozygous and all sterile males were heterozygous and thus had the foreign (*D. mauritiana*) *Odysseus* allele.

These observations raise new questions about a potential association between genotype, hybrid sterility and misexpression of the candidate genes identified by Michalak and Noor (2004). Do sterile hybrid backcross males tend to be heterozygous for the foreign allele? Do fertile males, on the other hand, tend to be homozygous? Is misexpression in sterile males explained by downregulation of the foreign allele only or both alleles? Here we focus on one of the four candidate genes, *Acyp* (CG16870), and the association between its genotype, allele-specific expression, and hybrid male hybrid sterility.

Acyp encodes acylphosphatase, a small cytosolic enzyme that catalyzes the hydrolysis of

acylphosphates, compounds containing a carboxylphosphate bond (Stefani *et al.* ., 1997). Although its role and function in vivo are as yet not completely clear, it has been shown that acylphosphatase is involved in membrane pumps and thus plays role in controlling the ion transport across biological membranes (Stefani *et al.* ., 1981; Nediani *et al.* ., 1991; Nassi *et al.* ., 1994). The pathway leading from *Acyp* to male sterility has yet to be determined but there is some evidence that malfunction of the ion pumps, such as the sarcolemmal calcium pumps, causes defects in sperm motility and subsequent male infertility in mice (Withers *et al*., 2006).

4.2 Materials and Methods

4.2.1 Fly Stocks

D. simulans were taken from the Florida City (FC) line, an isofemale line collected in 1985 in Florida City. *D. mauritiana* were taken from the SYN stock, a combination of six isofemale lines originally collected on Mauritius in 1981. Both stocks obtained (courtesy of Mohamed Noor) were reared at 24ºC under a 12-hour light-dark cycle on corn food.

4.2.2 Fertility Assay

Two reciprocal types of backcross progeny were produced and examined: 1) BC1-S from backcrossing of F_1 hybrid females (*D. simulans* females x *D. mauritiana* males) to *D. simulans* males, and 2) BC1-M from backcrossing F_1 hybrid females (*D. mauritiana* females x *D. simulans* males) to *D. mauritiana* males. Offspring were collected shortly after eclosion and males and females were separately stocked in dark for seven days for the following fertility assay (see Michalak and Noor 2004 for details). Backcross males were individually mated with virgin backcross females of their own type and based on the absence/presence of sperm in the female tracts they were assigned as either sterile (S: no full-length sperm observed in female's reproductive tract) or fertile (F: full-length, motile sperm observed).

4.2.3 Genotyping and expression analyses

Within an hour since the fertility assay, RNA was extracted from decapitated males using the RiboPure kit from Ambion. Heads were used for DNA extraction and genotyping. PCR and pyrosequencing primers (available upon request) were designed to target two species-specific

SNPs in the following coding sequence of *D. mauritiana*/*D. simulans Acyp*: GCGATCGCTATGGCTAC/TGCC/GAACTTTCATATCAAGCCCGAT (Figure 4.1). A PSQ^{TM}96HS pyrosequencing system was outsourced at EpigenDx to genotype the backcross males and quantify allele-specific expression in heterozygotes based on, respectively, genomic DNA and RNA samples. RNA was reverse-transcribed using Promega Access RT-PCR system in a one-step quantitative RT-PCR containing 5 µL AMV/ Tfl5 Reaction Buffer, 1 µL 25mM $MgSO_4$, 1.5 µL 0.0001% SYBR Green, 1 µL 20 µM downstream and upstream primers (one of which was biotinylated), 0.5 µL dNTP mix (10 mM each), 0.5 µL TflDNA Polymerase, 0.5 µL AMV Reverse Transcriptase, and complemented to a total of 25 µL with DEPC-treated water. Amplifications were visualized using BIO-RAD iCycler with MyiQ^{TM} Optical Module. Cycling included an initial 45 min at 48ºC for reverse transcription (skipped for genomic DNA), 2 min at 95ºC, 35 cycles of 95ºC for 30 s, 55ºC for 1 min and 72ºC for 2 min.

As a parental species reference for allele-specific expression in hybrids, 5 RNA combinations of pure-species samples were prepared in 1:0, 3:1, 1:1, 1:3, and 0:1 *D. simulans*:*D. mauritiana* ratios. RNA concentrations were quantified with Nanodrop spectrophotometer and samples were processed identically to the backcross RNA extractions. A total of six biological replicates of each dilution ratio, each with two additional technical replicates, were pyrosequenced along with the heterozygous samples.

In addition to pyrosequencing, RNA samples were subjected to TaqMan quantitative RT-PCR. A total of 50 ng/sample of RNA (5 µL) were reverse-transcribed and PCR-amplified in a one-step reaction containing ABI Multiscribe (1.25 µL), TaqMan Universal Master Mix (12.5 µL), RNase inhibitor (0.5 µL), and a mix of primers and FAM-labeled MGB probe (8.25 µL). PCR cycling conditions consisted of initial 30 min at 48ºC, followed by 10 min at 95ºC, and 40 cycles of 95ºC for 15 s and 60ºC for 1 min. Ribosomal protein L32 (ABI TaqMan Gene Expression Assay) was used as an endogenous normalizer (ABI TaqMan Gene Expression Assay) and the *Acyp*/*RpL32* differences of the threshold cycle (C_T) values were compared between phenotypes (fertile and sterile) and between genotypes (homozygotes and heterozygotes).

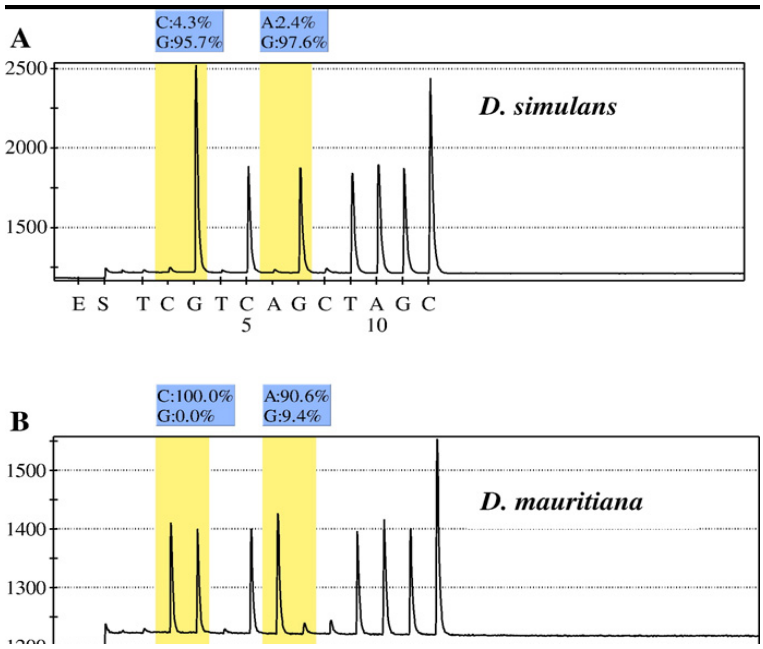


Figure 4.1 Pyrograms of parental Drosophila simulans (A) and D. mauritiana (B) showing species-specific single nucleotide polymorphism (shaded) in Acyp.

4.3 Results and Discussion

4.3.1 Genotype–phenotype associations

Backcrossing F_1 females from *D. simulans* x *D. mauritiana* crosses to either parental species produces sterile males along with fertile males, thus enabling a direct comparison of genotypes in these two phenotypic classes. A total of 35 fertile and 41 sterile males were genotyped. There was a clear association between the genotype and the phenotype: homozygous males tended to be fertile and heterozygous males tended to be sterile (Table 4.1) in both the BC1-S cross (Fisher exact test P = 0.017) and the BC1-M cross (P = 0.024). Note that homozygotes carry two endogenous alleles whereas heterozygotes carry one endogenous and one exogenous (foreign) allele. The association was stronger for homozygosity/fertility, especially in the BC1-M cross where all 13 fertile males were homozygotes. The relationship between heterozygosity and sterility was less consistent, as 71% of sterile BC-S males were heterozygous but only 35% sterile BC1-M males were heterozygous. A possible explanation of this difference is that the category of sterile males is not completely homogenous and may potentially include

misassigned fertile males. Detection of fertility based on the sperm presence is more obvious and the probability of misassignment is therefore less likely. Even if real, however, this classification bias makes our analysis statistically more conservative and does not change our conclusions.

Table 4.1 2 × 2 contingency tables for measuring association between the genotype (S — D. simulans, M — D. mauritiana) and the fertility/sterility phenotype for two reciprocal crosses (BC1-S and BC1-M)

Zygosity	Fertile	Sterile
BC1-S (i.e. [S♀ × M♂]F_1♀ × S♂)		
SS	15	7
SM	7	17
BC1-M (i.e. [M♀ × S♂]F_1♀ × M♂)		
MM	13	11
MS	0	6

4.3.2 Expression analyses

TaqMan assays demonstrated almost 7-fold underexpression of *Acyp* in sterile males relative to fertile males in both BC1-S (t-test = -2.178, df = 12, P = 0.05, unequal variances) and BC1-M (t = -2.263, df = 10, P = 0.047) (Figure 4.2A). A question arises whether this difference in expression could be explained by an incompatibility caused by the foreign allele. To address this, mean expression values of heterozygotes and homozygotes were compared. Heterozygotes exhibited ~7-fold underexpression relative to homozygotes (Figure 4.2B) in both BC1-S (t = 2.266, df = 17, P = 0.037) and BC1-M (t = 4.001, df = 9, P = 0.003). We then narrowed down the analysis to heterozygotes only and asked if sterile heterozygotes expressed *Acyp* differently compared to fertile heterozygotes. The former still tended to show underexpression in both backcrosses, and although it was more than a 3-fold change, this difference was not statistically significant when measured as C_T-values (t-test, P > 0.05).

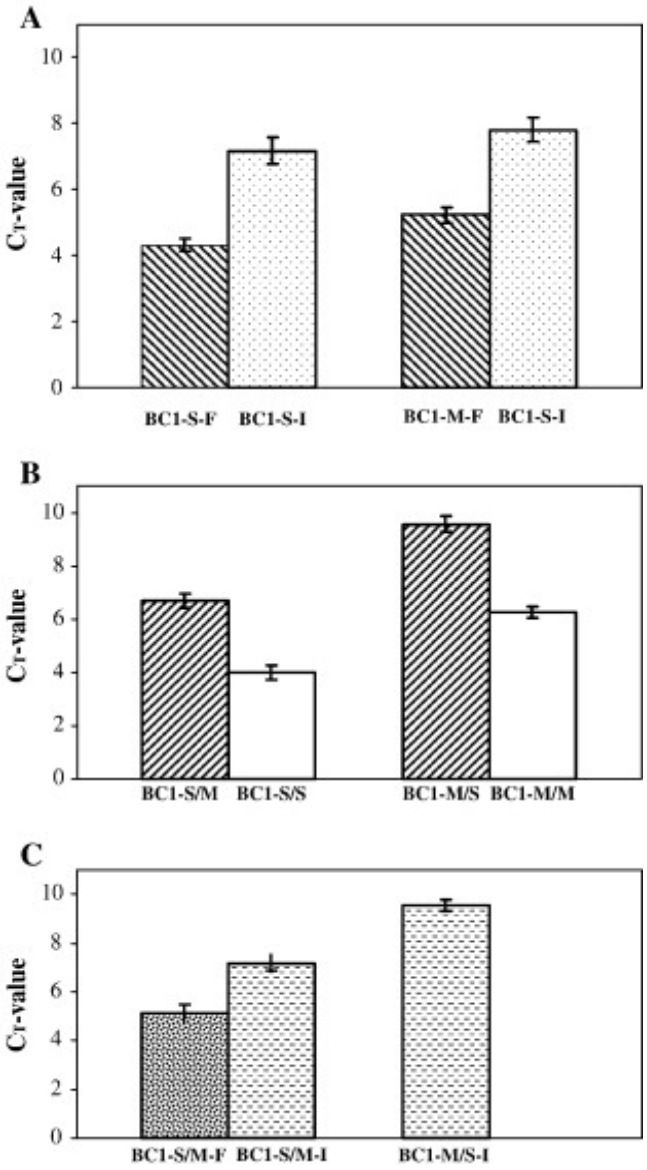


Figure 4.2 Mean expression (±SE) measured as TaqMan CT-values in two reciprocal backcrosses (BC1-S and BC1-M), sterile (I) vs. fertile (F) males (A), heterozygotes (BC1-S/M and BC1-M/S) vs. homozygotes (BC1-S/S and BC1-M/M) (B), and sterile vs.fertile heterozygotes (C).

4.3.3 Allele-specific expression

It is tempting to assume that misexpression of the foreign allele underlies observed expression differences. However, the analysis of allele-specific expression in heterozygotes reveals a less intuitive pattern (Figure 4.3). Regardless of the backcross direction and the fertility/sterility pattern, the *D. simulans* allele is consistently overrepresented (52.18-55.22%), with the overexpression statistically significant in BC1-S Sterile (t-paired test, P = 0.002), BC1-S

Fertile (P=0.003) but not BC1-M Sterile (P = 0.6; no BC1-M Fertile were found). Combined with a high overrepresentation of this allele in the parental 1:1 RNA mix, this suggests that the ratio of these two alleles is determined by a strong *cis*-type regulation.

Allele-specific expression of non-imprinted genes is a relatively poorly characterized phenomenon (Cheung and Spielman, 2002; Knight, 2004) but has been documented in yeast, flies, mice, and humans (Cowles *et al.* ., 2002; Yan *et al.* ., 2002; Lo *et al.* ., 2003; Wittkopp *et al.* ., 2004; Ronald, 2005). In plants, naturally occurring allele-specific expression has been observed in maize and cotton (Guo *et al.* ., 2003; Guo *et al.* ., 2004; Adams and Wendel, 2005) and suggested to be an important genetic component for phenotypic variation in general (Doebley and Lukens, 1998; Buckler and Thornsberry, 2002). Allele-specific expression patterns in hybrids relative to the parental species can be used to infer the relative importance of *cis*- and *trans*-regulatory variation at the genome-wide level (Cowles *et al.* ., 2002; Wittkopp *et al.* ., 2004; Landry *et al.* ., 2005). Experiments on yeast showed that 100-200 *trans*-acting loci controlled the variation in gene expression of up to 1716 genes (Brem *et al.* ., 2002; Yvert *et al.* ., 2003); in addition to transcription factors, these loci represented multiple functional classes. Several other studies implicated a preponderance of *trans*-acting effects in humans (Morley *et al.* ., 2004; Monks *et al.* ., 2004), *Drosophila* (Gibson *et al.* ., 2004), *Caenorhabditis elegans* (Denver *et al.* ., 2005), and *Eucalyptus* (Kirst *et al.* ., 2005). However, other studies on flies, human, mice, and maize provided conflicting results suggesting predominance of *cis*-acting factors (Wittkopp *et al.* ., 2004; Yan *et al.* ., 2002; Schadt *et al.* ., 2003; Cowles, 2002). Noteworthy, regulatory elements may diverge genetically between species as a result of mutual compensation of *cis*- and *trans*-regulatory elements without changing the level of gene expression (Landry *et al*. 2005).

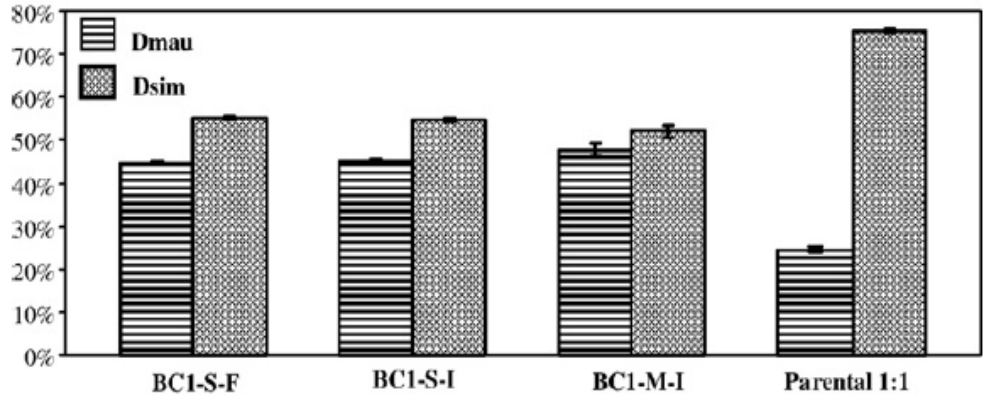


Figure 4.3 Allele-specific expression (mean percent±SE) in backcross males heterozygous for Acyp, dependent on the direction of backcrossing, (BC1-S and BC1-M), sterility (I), and fertility (F). A 1:1 mix of parental RNA is included for comparison.

4.4 Conclusions

We have found a clear link between the genotype of Acyp and the fertility/sterility pattern. While homozygous males tend to be fertile, heterozygous males that carry an exogenous allele from the foreign species tend to be sterile. Moreover, we observed that the locus was significantly downregulated in sterile and heterozygous males relative to, respectively, fertile and homozygous males. Interestingly, this underexpression appeared to be due to a roughly proportional downregulation of both alleles rather than the exogenous allele alone. The relative quantities of allele-specific transcripts in heterozygotes remained at a similar level, with one of the alleles (D. simulansoriginated) consistently overrepresented regardless of the sterility/fertility pattern and the direction of backcrossing. The role of acylphosphatase in hybrid sterility needs yet to be explored using targeted mutagenesis, transgenic manipulations and RNAi. Nevertheless, our analysis has already shown that Acyp provides an excellent candidate for a gene involved in Dobzhansky–Muller genetic incompatibilities in Drosophila hybrids.

PART IV

CHAPTER 5

GENE *CG5762* IS ASSOCIATED WITH HYBRID STERILITY

5.1 Introduction

Reproductive isolation is a required step to speciation in sexual organisms, which could be achieved by two modes-- prezygotic isolation and postzygotic isolation. Understanding the genetic basis of the two modes is essential to determine the evolutionary forces making two populations diverge into distinct species. Dobzhansky and Muller (Dobzhansky 1936, Muller 1942) first proposed the idea of genetic incompatibilities that genes may not function properly when in conflict with other alleles from different species. Those interacting genes play a crucial role in isolation, but identification of the genes has not been very successful so far. Only five hybrid sterility genes have been isolated- Odysseus (*Ods*) in *Drosophila* (Ting *et al.* 199 8), *JYAlpha* in *Drosophila* (Masly *et al.* 2006), AEP2 in yeast (Lee *et al.* 2008), Overdrive in *Drosophila* (Phadnis and Orr 2009), and Prdm9 in mouse (Mihola *et al.* 2009). One difficult issue for identification is that hybrids are often unavailable for many species. *Drosophila* is a good animal model to overcome this obstacle, since *D. melanogaster* and its sibling species (*D. simulans* complex) have diverged for the last 3 million years (Hey and Kliman 1993) butthey can still mate and produce partially viable or fertile hybrids.

Michalak and Noor (2003, 2004) used microarrays and quantitative RT-PCR to survey genome-wide gene misexpression in sterile and fertile hybrid males from crosses between *D. simulans* and *D. mauritiana*. We further analyzed one of the candidate genes, *Acyp*, by quantitative RT-PCR, pyrosequencing, and TaqMan qRT-PCR assays (Michalak and Ma 2008). The results showed a strong association of *Acyp* homozygosity/heterozygosity with the

fertility/sterility pattern. Moreover, *Acyp* expression was downregulated in sterile and heterozygous males compared to fertile and homozygous males, respectively. Here, I expand this methodology to *CG5762*, another candidate gene for hybrid sterility (Michalak and Noor 2003, 2004).

Unlike *Acyp*, *CG5762* is located on Chromosome 3R, and its function is essentially unknown except that it is expressed in testes (Andrews *et al.* ., 2000). Noor (2005) analyzed sequence divergence between *D. pseudoobscura*, *D. miranda* and *D. melanogaster* and found an accelerated amino acid substitution rate in *CG5762* in the *D. pseudoobscura* species group (but not *D. simulans*; see Michalak & Noor, 2004), suggestive of positive selection acting upon this gene. Given its expression pattern as well as sequence divergence signatures, *CG5762* meets criteria of a candidate gene that may have played role in the evolution of hybrid male sterility.

5.2 Materials and Methods

5.2.1 Fly Stocks

D. simulans were taken from the Florida City (FC) line, an isofemale line collected in 1985 in Florida City. *D. mauritiana* were taken from the SYN stock, a combination of six isofemale lines originally collected on Mauritius in 1981. Two reciprocal types of backcross progeny were produced: 1, backcross of F1 hybrid females (*D .simulans* females×*D. mauritiana* males) to *D. simulans* males (assigned as BC1-S); and 2, backcross of F1 hybrid females (*D. mauritiana* females×*D.simulans* males) to *D. mauritiana* males (assigned as BC1-M). All flies were reared at 24 °C under a 12-hour light–dark cycle on cornfood.

5.2.2 Fertility assay

Backcross males were individually mated with virgin backcross females of their own type and based on the absence/presence of sperm in the female tracts they were assigned as either sterile, or infertile (S: no full-length sperm observed in female's reproductive tract) or fertile (F: full-length, motile sperm observed). A total of 84 males (40 Fertile, 44 Sterile) were collected. Although this assay has been previously demonstrated to reflect true functional sterility as tested in experimental crosses (Michalak and Noor, 2004), I have conducted an additional pilot

experiment to confirm this correlation. A total of 43 sterile males that classified as “S” were mated with three *D. simulans* females each for four days. None of the 43 males was fertile.

5.2.3 Genotyping and expression analyses

Within an hour since the fertility assay, RNA was extracted from males using the RiboPure kit from Ambion. Heads were used for DNA extraction and genotyping. PCR and pyrosequencing primers were designed to target two species-specific SNPs in the following coding equence of *D. simulans*/ *D. mauritiana CG5762*: GCG/TCGGATCGAATGCGCCTCCACGCGTCACATCGCCA (Figure 5.1).The following primers were used: 5′-TGGAGCGTCGCTTGTTCT -3′(PCR, biotinated), 5′-TGATGGTGAACCGCTGCAT -3′ (PCR), and 5′-ATCGAGCTCCTCCGT -3′ (sequencing).

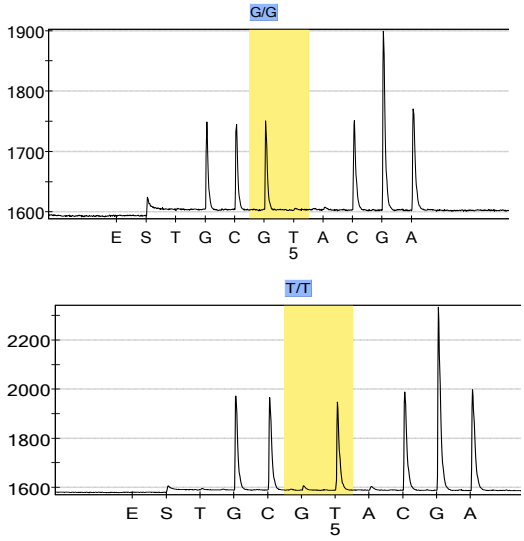


Figure 5.1 Pyrograms of parental Drosophila mauritiana (left) and D. simulans (right) showing species-specific single nucleotide polymorphism (shaded) in CG5762

RNA was reverse-transcribed using Promega Access RT-PCR system in a one-step quantitative RT-PCR containing 5 µL AMV/Tfl5 Reaction Buffer, 1 µL25 mM MgSO4, 1.5 µL 0.0001% SYBR Green, 1 µL 20 µM forward and reverse primers 0.5 µL dNTPmix (10 mM each), 0.5 µL TflDNA Polymerase, 0.5 µL AMV Reverse Transcriptase, to a total volume of 25 µL. Amplifications were visualized using BIO-RAD iCycler with MyiQ™ Optical Module. Cycling

included an initial 45 min at48°C for reverse transcription (skipped for genomic DNA), 2 min at95 °C, 35 cycles of 95 °C for 30 s, 55 °C for 1 min and 72 °C for 2 min.

As a parental species reference for allele-specific expression in hybrids, 6 single-fly RNA samples of *D. simulans* and 6 single-fly RNA samples of *D. mauritiana* were derived, quantified with Nanodrop spectrophotometer and combined into 6 independent *D. simulans*:*D. mauritiana* mixes in 1:1concentration proportions. Another 4 pairs of independent flies were mixed in 4:0, 3:1,1:3, 0:4 concentration proportions (1:1 ratio from above were also used). All these samples were PCR-amplified and pyrosequenced along with the heterozygous samples. This procedure provided a correction for technical errors for both PCR and pyrosequening process. After PCR, a PSQ™96HS pyrosequencing system was outsourced at EpigenDx to genotype the backcross males for genomic DNA samples and quantify allele-specific expression in heterozygotes.

In addition to pyrosequencing, RNA samples were subjected to TaqMan quantitative RT-PCR. A total of 50 ng/sample of RNA (5 µL) were reverse-transcribed and PCR-amplified in a one-step reaction containing ABI Multiscribe (1.25 µL), TaqMan Universal Master Mix (12.5 µL), RNase inhibitor (0.5 µL), and a mix of primers and FAM labeled MGB probe (8.25 µL). PCR cycling conditions consisted of initial 30 min at 48 °C, followed by 10 min at 95 °C, and 40 cycles of 95 °C for 15 s and 60 °C for 1 min. Ribosomal protein L32 (ABI TaqManGene Expression Assay) was used as an endogenous normalizer (ABITaqMan Gene Expression Assay) and the *CG5762* /RpL32 differences of thethreshold cycle (CT) values were compared between phenotypes(fertile and sterile) and between genotypes (homozygotes and heterozygotes).

5.2.4 *CG5762* sequence comparison and protein structure prediction

Orthology information on *CG5762* has been taken from UCSC genome browser. I also blasted the protein sequence in BLASTP to confirm the annotation. I further aligned nucleotide and protein sequence of *CG5762* in closely related species (*D. simulans*, *D. mauritiana*, and *D.melanogaster*) using software Mega4. Ka/Ks ratios were calculated using DnaSP4.0. The conserved regions of *CG5762* in these three species were searched in Pfam to survey for functional ORFs match. The secondary structure was predicted using the Phyre server from

website http://www.sbg.bio.ic.ac.uk (Kelley and Sternberg 2009).

5.3 Results

5.3.1 Genotype-Phenotype association

Two distinct phenotypic classes of backcross males were record—fertile and sterile, a total number of 40 fertile and 44 sterile males were genotyped for gene *CG5762*. The overall data for both direction clearly showed that homozygous males tended to be fertile (28 fertile vs 18 sterile) and heterozygous males tended to be sterile (26 sterile vs. 12 fertile) (Fisher exact test P=0.009) (Table 5.1)

Table 5.1 2 × 2 contingency tables for measuring association between the genotype (S — D.simulans, M — D.mauritiana) and the fertility/sterility phenotype for two reciprocal crosses (BC1-S and BC1-M)

Zygosity	Fertile	Sterile
AA(SS or MM)	28	18
Aa (SM)	12	26
Fisher-exact		P =0.009075

(AA: homozygous alleles both from D. simulans (SS) or from D. mauritiana (MM)
Aa: heterozygous alleles one from D .simulans (S) and one from D. mauritiana (M))

5.3.2 Expression analysis

TaqMan assays demonstrated almost 3-fold underexpression of *CG5762* in sterile males relative to fertile males (P=0.08, two-tailed t test with unequal variances, df=27) (Figure 5.2)

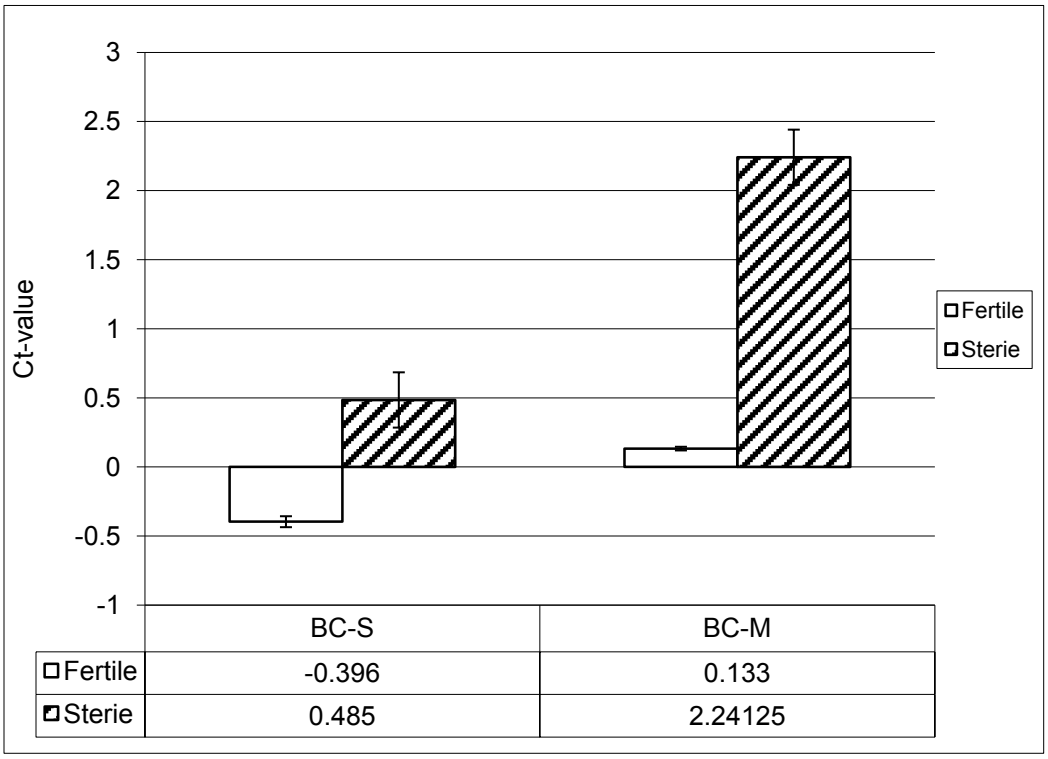


Figure 5.2 Mean expression (±SE) measured as TaqMan CT-values in two reciprocal backcrosses (BC1-S and BC1-M), sterile (shaded bar) vs. fertile (white bar) males

5.3.3 Allele-specific expression

A series of parental cDNA mix (concentration ratio of D.mau:D.sim 4:0; 3:1; 2:2; 1:3; 0:4) were amplified and pyrosequenced as well as all the backcross samples. The expression data of parental mix was used to build in a linear regression. All the backcross males' expressions raw data were corrected by the regression. This step was applied to account for possible technical errors in the procedure of qRT-PCR and pyrosequencing.

The results showed that regardless of the backcross direction and the fertility/sterility pattern, the *D. mauritiana* allele is consistently overrepresented (71–75%), and this suggests that the ratio of these two alleles is determined by a strong trans-type regulation (Figure 5.3).

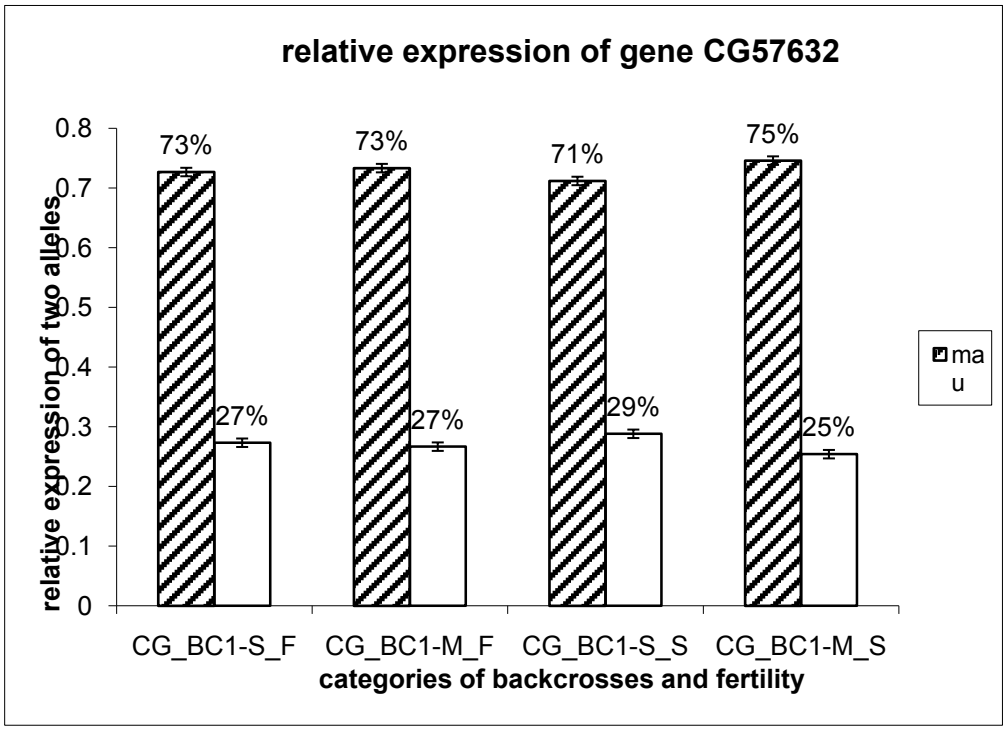


Figure 5.3 Allelic expression of CG5762. Four categories used for the comparison are: backcross to D. simulans, fertile males (BC1-S_F); backcross to D. mauritiana, fertile males (BC1-M_F); backcross to D. simulans, sterile males (BC1-S_S); backcross to D. mauritiana, sterile males (BC1-M_S).

5.3.4 Comparative sequence analysis

Using UCSC genome browser and BLASTP, *D. simulans CG5762* sequence was compared with published sequences from other *Drosophila* species. Orthologous genes were analyzed using the best BLASTP hit (or reciprocal-best BLASTP hits), and filtering out non-syntenic hits. The results show that *CG5762* has no orthologs in human, mouse, rat, zebrafish, *C. elegans* or *S. cerevisiae*. Although the absence may reflect incomplete annotations in the other species, it seems reasonable to assume that *CG5762* is a lineage-specific gene present only in *Drosophila*.

A *CG5762* alignment of published genome sequences from 12 flies, a mosquito, a honeybee, and a beetle species clearly shows that it is present in *D. simulans*, *D. sechellia*, *D. yakuba* and *D. erecta*, which indicates that the gene emerged before the split of *melanogster* subgroup. Our sequence data of *CG5762* in *D. mauritiana* also show a similarity to that in *D.*

simulans. (Figures 5.4 and 5.5)

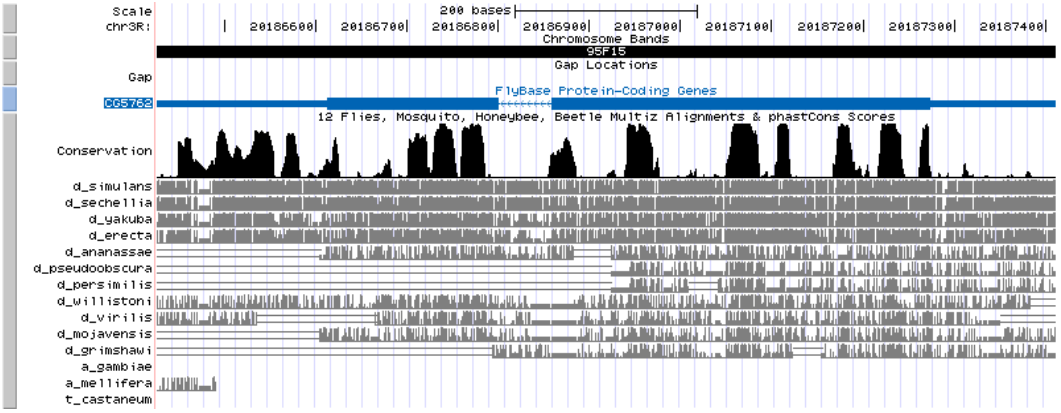


Figure 5.4 Alignment of CG5762 in closely-related species

To investigate the recent evolutionary history of this gene in *D. simulans* and *D. mauritiana*, we aligned their nucleotide and protein sequences. *CG5762* in both species contains two exons (1-417bp, 477-665bp) and one intron and encodes a 201 amino acid protein (Figure 5.6a). We identified five synonymous polymorphisms and three nonsynonymous polymorphisms. The estimated ratio of non-synonymous to synonymous substitutions (Ka/Ks) is thus 0.18. *CG5762* in *D. melanogaster* has one exon only (139-741bp) and encodes 200 amino acids. There are 25 synonymous polymorphisms and 13 nonsynonymous polymorphisms between *D. melanogaster* and *D. simulans*, resulting in a Ka/Ks ratio of 0.14. (Figure 5.6b) However, the sequence in *D. pseudoobscura* shows a dramatic difference compared to that of *D. simulans* with only 54 amino acid matched (Noor 2005). Noor (2005) further compared sequences of this gene from *D. miranda* and *D. pseudoobscura*. The calculated Ka/Ks ratio was 1.35, most likely a sign of positive selection ongoing between these two species.

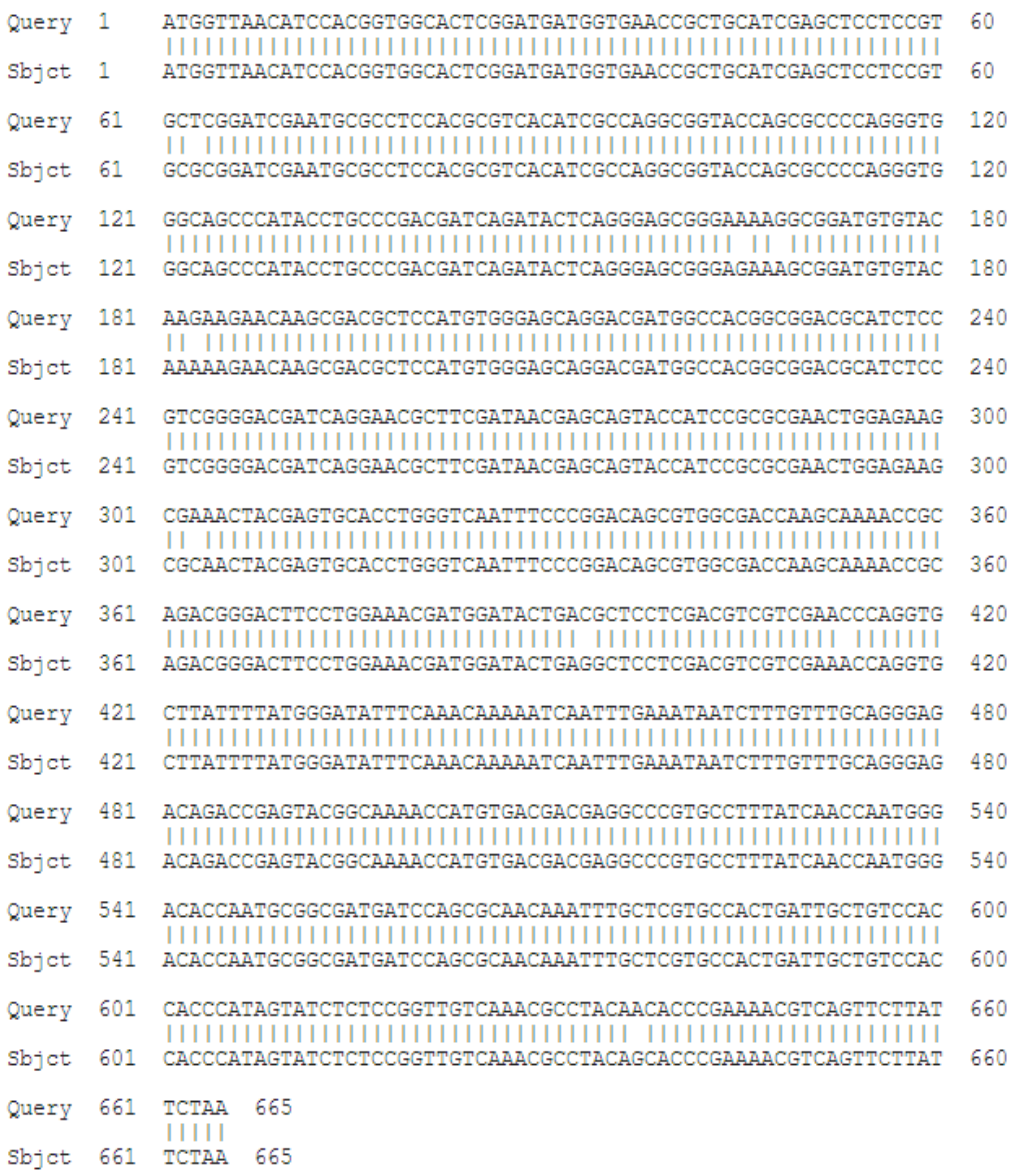


Figure 5.5 Alignment of CG5762 nucleotide sequence between D. simulans(query) and D. mauritiana (subject). (657/665 identities, no gaps)

```
Query  1    MVNIHGGTRMMVNRCIELLRARIECASTRHIARRYQRPRVGSPYLPDDQILREREKADVY  60
            MVNIHGGTRMMVNRCIELLRARIECASTRHIARRYQRPRVGSPYLPDDQILREREKADVY
Sbjct  1    MVNIHGGTRMMVNRCIELLRARIECASTRHIARRYQRPRVGSPYLPDDQILREREKADVY  60

Query  61   KKNKRRSMWEQDDGHGGRISVGDDQERFDNEQYHPRELEKRNYECTWVNFPDSVATKQNR  120
            KKNKRRSMWEQDDGHGGRISVGDDQERFDNEQYHPRELEKRNYECTWVNFPDSVATKQNR
Sbjct  61   KKNKRRSMWEQDDGHGGRISVGDDQERFDNEQYHPRELEKRNYECTWVNFPDSVATKQNR  120

Query  121  RRDFLETMDTDAPRRRRTQGDRPSTAKPCDDEARAFINQWDTNAAMIQRNKFARATDCCP  180
            RRDFLETMDT+APRRRR QGDRPSTAKPCDDEARAFINQWDTNAAMIQRNKFARATDCCP
Sbjct  121  RRDFLETMDTEAPRRRRNQGDRPSTAKPCDDEARAFINQWDTNAAMIQRNKFARATDCCP  180

Query  181  PPIVSLRLSNAYNTRKRQFLF  201
            PPIVSLRLSNAY+TRKRQFLF
Sbjct  181  PPIVSLRLSNAYSTRKRQFLF  201
```

(a)

```
Query  1    MVNIHGGTRMMVNRCIELLRARIECASTRHIARRYQRPRVGSPYLPDDQILREREKADVY  60
            MVNIHGGTRMMVNRCIELLRARIECASTRHIARR QRPRVGS YLPDDQILREREKADVY
Sbjct  1    MVNIHGGTRMMVNRCIELLRARIECASTRHIARRSQRPRVGSQYLPDDQILREREKADVY  60

Query  61   KKNKRRSMWEQDDGHGGRISVGDDQERFDNEQYHPRELEKRNYECTWVNFPDSVATKQNR  120
            KKNKRRSMWEQD+GHG RISVGDDQERFDNEQYHPRELEKR YECTWVNFPDSVATK+NR
Sbjct  61   KKNKRRSMWEQDEGHG-RISVGDDQERFDNEQYHPRELEKRKYECTWVNFPDSVATKRNR  119

Query  121  RRDFLETMDTDAPRRRRTQGDRPSTAKPCDDEARAFINQWDTNAAMIQRNKFARATDCCP  180
            RRDFLE+M+T+APRRRR Q +RPSTAKPCDDEARAFINQWDTNAAMIQRNKFARATD CP
Sbjct  120  RRDFLESMETEAPRRRRAQEERPSTAKPCDDEARAFINQWDTNAAMIQRNKFARATDSCP  179

Query  181  PPIVSLRLSNAYNTRKRQFLF  201
            PPIVSLRLSNAY+TRKRQFLF
Sbjct  180  PPIVSLRLSNAYSTRKRQFLF  200
```

(b)

Figure 5.6 Alignment of CG5762 protein, D. simulans (query) vs. D. mauritiana (subject) (198/201 identities, no gap) (a); D. simulans (query) vs. D. melanogaster (subject) (187/201 identities, 1 gap) (b).

An amino acid search in Pfam shows a match between region (86-135)of *CG5762* and DUF1431, a gene located on chromosome 2R, which belongs to a protein family of unknown function, found only in *Drosophila* species so far. This family contains several conserved cysteine residues (Figure 5.7).

Family	Description
DUF1431	Protein of unknown function (DUF1431)

```
#HMM       prlDlkyYkpsdkakRkYqrtWvecpklk.vkpkkvcllekakppeikrR
#MATCH      r+D++ Y+p + +kR+Y+ tWv+ p  +  k+ +   + ++   + +rR
#PP        59**********************9965431333332223333333344444
#SEQ       DRFDNEQYHPRELEKRNYECTWVNFPDSVaIKQNRRRDFLETMDTDAPRR
```

Figure 5.7 A comparison of CG5762 and DFU1431 amino acid sequences (HMM: consensus of the hidden Markov models of the family; #PP: Posterior probability, or the degree of confidence in each residue from 100% probability (*), 90% probability(9) to no probability (0); #SEQ: sequence of CG5762)

Interestingly, when we used orf Finder to identify the possible ORF based on the nucleotide sequences (422-607bp), we found a frameshift duplicate that functions as a Phage shock protein G which is upregulated in response to a number of stress conditions, including ethanol, expression of the filamentous phage secretin protein IV and other secretins, as well as heat shock.

5.3.5 Expression in different development stages

Embryonic expression pattern (from Berkeley *Drosophila* Genome Project) indicates that *CG5762* is expressed in the embryonic anal pad at stage 13-16. (Figure 5.8)

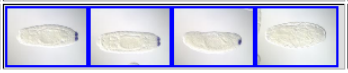
Gene	Stage	Image	Body Part
CG5762 [GH07296] **Map**: 95F15-95F15	stage4-6		no staining
	stage7-8		no staining
	stage9-10		no staining
	stage11-12		no staining
	stage13-16		embryonic anal pad

Figure 5.8 Embryonic expression of CG5762 (Image from BDGP gene expression report of CG5762)

Anal pads are additional structures in *Drosophila* larvae that regulate the ion concentration in the hemolymph by transporting ions from the environment into hemolymph. Larvae in a hypotonic solution would have enlarged anal pads (Keyser, Borge-Renberg *et al.*

2007). Since *CG5762* is only expressed in anal pads in early stage, it may have some functions related to ion control, similar to *Acyp*.

In *D .melanogaster* adults, *CG5762* are reported to express predominantly in testis (Andrews, Bouffard *et al.* 2000). RT-PCR expression data indicate a lower concentration of the transcript in hybrids than in pure species (Michalak and Noor 2003).

5.3.6 Protein structure prediction

The Phyre server shows results of structure prediction via three independent secondary structure prediction programs (Psi-Pred, SSPro and JNet) together with a consensus prediction. The query sequence is then compared with known protein structure taken from the Structural Classification of Proteins (SCOP) database and the Protein Data Bank (PDB).The output is in the form of a three state-- alpha helix (H in red), beta strand (E in blue) and coil (C in grey). The number in color below each position indicates the confidence of the prediction from 0 (low) to 9 (high). Based on the consistency between the three prediction methods and the confidence value, this result implies a quite accurate prediction of the secondary structure of *CG5762* (Figure 5.9).

We then focused on the best match alignments according to their secondary structure and amino acid sequences. The top two are c2rh1A (500aa, 7% identity) and d1iu4a (331aa,16% identity). C2rh1A is a membrane protein/hydrolase which belongs to a family of eukaryotic signal transduction proteins that communicate across the membrane (Cherezov, Rosenbaum *et al.* 2007). D1iu4a is a cysteine proteinase in the family of microbial transglutaminases.

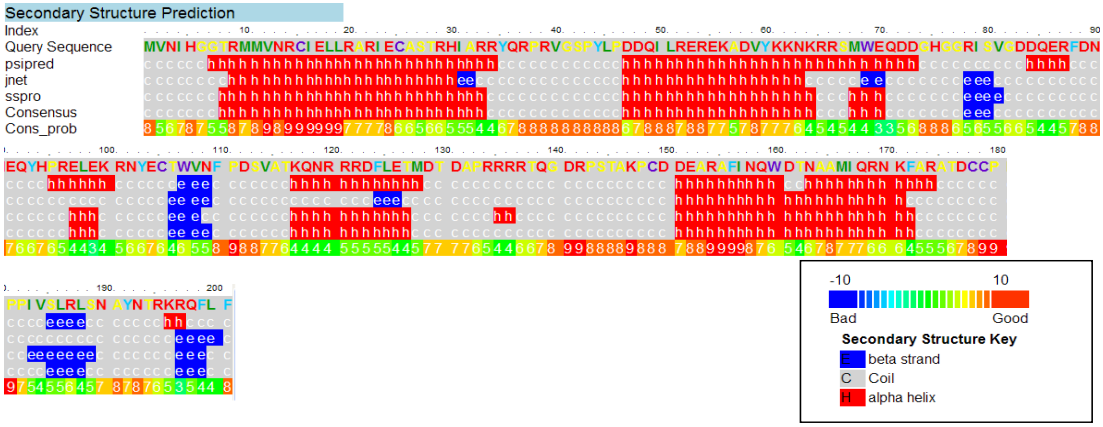


Figure 5.9 Detailed view of the secondary structure prediction of CG5762.

5.4 Discussion

In a similar way to *Acyp*, *CG5762* genotypes and expression patterns are associated with male fertility/sterility. Homozygous males tended to be fertile and heterozygous males tended to be sterile. There is a significantly less expression of the gene in sterile than fertile males.

Sequence comparison of *CG5762* between species indicates that it is a fast evolving gene present only in *Drosophila*. Given the tissue-specific expression pattern in both embryo and adults, we speculate that *CG5762* possibly functions in ion transportation through membranes which somehow translates into a role in hybrid male sterility. To cast more light on normal functions of this gene and its effects when disrupted or misexpressed in hybrids, we plan on using RNAi and P-element insertions. It is plausible that *Acyp* and *CG5762* are both involved in the same or similar pathways related to transport across membranes. Misexpression of these genes or dysfunctions of their original functions in hybrid background may contribute to hybrid male sterility. A similar phenomenon also exists in other genes—which perform certain metabolic functions in pure species but misexpression in hybrid contributes to sterility/inviability thus contributing to isolation between species. For example, the gene *Tu* (macromelanophore) is a dominant tumour gene for melanoma in *Xiphophorus* fish hybrids. The protein structure analysis reveals that it is a transmembrane protein acting as a receptor tyrosine kinase in pure species. The malignant phenotype is suppressed with tumour suppressor gene R, but in the absence of R in hybrid,this gene leads to melanoma (Wittbrodt *et al.* 1989). Introgressions containing *Odysseus* (Ods) region of *D. mauritiana* into *D. simulans* (or the opposite direction) result in completely sterile male offspring. The homologs of *OdsH* in mammals and nematodes play a function in neural tissues. Since *OdsH* is also expressed in the *Drosophila* testis, it is plausible that *OdsH* evolves a male germ line function in *Drosophila*, misexpression in hybrid results in complete sterility in males (Ting *et al.* 1998). In a recent study by Bayes and Malik(2009), they found that *D. mauritiana* or fertile introgressed *D. simulans* males without detectable OdsH protein, are still completely fertile, thus, hybrid male sterility did not require the gene's function itself, but rather

caused by a gain-of-function interaction between *OdsHmau* and heterochromatic Y chromosome of *D. simulans*.

However, it is also likely that other cis- or trans-located genetic factors produce spurious associations between candidate genes and hybrid defects. Therefore, it is imperative to determine a complete map of all genetic interactions involved in creating reproductive isolation.

PART V

CHAPTER 6

A SURVEY OF MICROSATELLITE DIVERGENCE BETWEEN *DROSOPHILA SIMULANS* AND *D. MAURITIANA* CHROMOSOME III IN AN APPROACH TO MAPPING OF HYBRID MALE STERILITY

6.1 Introduction

Microsatellites are short tandemly repeated sequences of 1-6 base pairs of DNA that are highly polymorphic in animals and plants. DNA slippage is presumably the primary mutation mode of microsatellites (Schlötterer 2000). Two possible mechanisms may account for the slippage: unequal crossover between misaligned repeats resulting in a deletion in one strand and an addition in the other (Smith 1974), and DNA polymerase slippage. In the process of replication, two stands transiently disassociate and then reanneal. If there are unpaired bases in the primer strand, continued elongation will result in an increase in length (Streisinger, Okada *et al.* 1966; Schlotterer and Tautz 1992). Because of their high levels of inter- and intra-specific polymorphism and broad genetic distributions, microsatellites are often used as informative genetic markers in studies of disease mutations (Batra, Charizanis *et al.* 2010), linkage mapping (Knapik, Goodman *et al.* 1998) and population genetic structures (Estoup, Jarne *et al.* 2002).

Microsatellites have become important markers for many applications in *Drosophila* as well. Goldstein and Clark surveyed 18 microsatellites in 15 isofemale lines collected throughout North America and found out 17 of these 18 markers are still highly polymorphic between populations (Goldstein and Clark 1995). However, the mutation rate in *Drosophila* is considerably lower than that of various mammals ($6.3X10^{-6}$ vs. 10^{-3}—10^{-5}) (Schug, Mackay *et al.* 1997). Schlötterer (1998) broadened the study of mutation rate in *Drosophila* with different microsatellites markers in 119 *D. melanogaster* lines and found that the mutation rate varied between loci, with one long microsatellite locus having the highest mutation rate (Schlotterer, Ritter et al. 1998). At the same

time, Schug *et al.* (1998) did a comprehensive search of *D. melanogaster* DNA sequences in GenBank and summarized the distribution and frequency of 1,298 microsatellites identified (Schug, Wetterstrand *et al.* 1998).

D. melanogaster is the closest species to *D. simulans* complex (*D. simulans*, *D. sechellia,* and *D. mauritiana)* from which it diverged about 3 million years ago (Hey and Kliman 1993). Microsatellite markers also provide useful insights into the evolution of *D. simulans* complex. For example, Colson and Goldstein (1999) selected 107 microsatellites loci and analyzed the genetic differences between *D. simulans* and *D. sechellia*, whose divergence time has been estimated as around 413,000 years ago (Kliman 2000). Their results showed that 71 out of the 107 loci derived from *D. melanogaster* sequences could be used to distinguish between *D. simulans* and *D. sechellia* with variable effectiveness. They speculated that *D. simulans* and *D. mauritiana* may have also accumulated a similar degree of genetic differentiation. However, according to Kliman (2000), *D. simulans* is a species more closely related to *D. mauritiana* (263,000 years of divergence time) than to *D. sechellia*. Evidently more data is needed to reconcile phylogeny of these *Drosophila* sibling species. In this study, I evaluate a total of 24 markers on the third chromosome derived from Colson and Goldstein study (2000) in *D. simulans* and *D. mauritiana* to select better loci in discrimination of those two closed species.

As reviewed in Part I, Haldane's rule inspired intensive theoretical and empirical studies on genetic basis of speciation. One explanation of Haldane's rule is that there is a large X effect on *Drosophila* male sterility due to consequences of gene dominance. Because *Drosophila* males are hemizygous (XY) and females are homozygous (XX), if some genes on the X cause hybrid sterility, the hemizygous sex will be affected more than homozygous sex regardless of dominance or recessiveness of the genes. However, the results are by no means unequivocal. Hollocher and Wu (1996) introgressed three regions from *D. sechellia* and *D. mauritiana* second chromosome into a mainly *D. simulans* background. They found that homozygous autosomal introgressions had a similar effect on hybrid sterility and inviability to hemizygous X-linked introgressions (Hollocher and Wu 1996). True *et al.* (1996) performed a genome-wide screen of 335

homozygous introgressions of *D. mauritiana* segments into a largely *D. simulans* genomic background. Their results showed that introgressions on the X cause hybrid male sterility more often than those on the autosomes (True, Weir *et al.* 1996). Masly and Presgraves (2007) performed a genome-wide introgression of *D. mauritiana* chromosomes into a mainly *D. sechellia* background. They found that more X-linked introgressions resulted in hybrid male sterility than autosomal introgressions (60%>18%) (Masly and Presgraves 2007). This does not necessarily indicate that there are more male sterility genes on the X chromosome. In fact, an examination of global gene expression patterns in *D. melanogaster* showed that genes with male-biased expression are dramatically underrepresented on the X chromosome (Parisi, Nuttall *et al.* 2003).

Hence, understandably, there has been much interest and focus on X chromosome. To compensate for the interest bias, here I focused on chromosome III to explore the patterns of *D. simulans*/*D. mauritiana* hybrid male sterility. One of loci associated with hybrid male sterility in the first generation of backcross between *D. simulans* and *D. mauritiana,* CG5762, is located on chromosome III (Michalak and Noor 2003, 2004). This analysis aims to map sterility gene(s) using microsatellite markers and recombinants from further backcross generations.

6.2 Materials and Methods

6.2.1 Fly stocks

Virgin *D. simulans* females were mated with *D. mauritiana* males to produce F_1 females that were in turn backcrossed to *D. simulans*. Backcrossing was continued for four generations, and only the last generation (hereafter designated BC4-S) was used for genetic analysis. The reciprocal backcross to *D. mauritiana* was also produced and the fourth generation (hereafter designated BC4-M) was used. All conditions (24ºC, 12-hour light-dark cycle, corn food) were the same for both backcrosses.

6.2.2 Fertility assay

After copulation of each BC4 female and male, female's reproductive tracks were checked under microscope for presence of sperm to determine male fertility (see Part III for details).

6.2.3 Microsatellite markers

After confirmation of each individual phenotype, DNA was extracted from the head of each single BC4 male fly using the standard protocol (Appendix C). A total of 24 microsatellite markers were tested. These 24 microsatellite loci were used to discriminate *D. simulans* and *D. sechellia* before (Colson, Macdonald *et al.* 1999). Primers were designed based on literature. followed by 35 cycles of 94°C denaturation for 1min, 55°C annealing for 1min, 72°C extension for 1min and a final extension at 72°C for 8 min. (Table 6.1)

Table 6.1 Microatellites loci studied

Locus name	Cytological location	Repeat (in D. mel)	Primer Forward	Primer Reverse	Reference
dmrhob	62A	(AC)10	tatactaaagtcacttaatgcgttaca	gtttgtgtgaatttcactaaattatta	Schug et al. 1997
ac004343	62A1-A2	(CT)11	acggtaattgcggatgagac	acgatggcaacaaggatctc	Colson & Goldstein 1999b
ac004658	63D2-E1	(AC)12	atttggtccacgagagattt	tgggaaaactgtgccacata	Colson & Goldstein 1999b
dmu36477	64A	(AT)14	cggcgagccaaactcttat	attatttgtggcaaaagcgg	Colson & Goldstein 1999b
drodsrc	64B	(AT)7	gacaacatttatcagctgctgc	agagttcatgtgtgttggttgg	Schug et al. 1998 (dmdsrc)
ac004352		(AG)14	tcctcggtgagaccgtaatc	gggcagaggaaaagcactca	Colson & Goldstein 1999b
dmu14395	65D1-D3	(CT)12	gggcagaggaaaagcactc	tcggtgagaccgtaatctgc	Colson & Goldstein 1999b
drolamb2a	67C	(ATT)5	cgtaggaaggaaagaaatcgg	aatttgcagttgataggcagc	Schug et al. 1998
drogtpaap		(CAG)5	ctgaaatacggcagcagac	taggcgttcatgttgctcgt	Colson & Goldstein 1999b
AC006414	89A1-A5	(GT)9	gaaagagctccaaggcaatcagg	tgtttcccaggacaggataagcg	Colson & Goldstein 1999b
dmz60mex	71C-D	(TTC)8	aaatctgttgctcatactgccc	aaccggcgaaatgttcag	Schug et al. 1998
dm22f11t	73A1-B7	(AT)22	ggatgctcggataccaaaaa	tcgcctgtgacttagattgc	Colson & Goldstein 1999b
dmcathpo	75D-E1	(ACC)6	ttcgacggatcagacttggtttttggc	gcgttcgcctttcttagtcaatttcgg	Schug et al. 1997
AC001665	84C1-C4	(AC)9	ttcgtttttgaattattgcgc	cctgttcgtccaactatgtgc	Schug et al. 1998 (dro17dc4z)
drohoxnk4	91E1-E3	(CAG)5	ctgaagttgaagtccgagcc	tacatgtgctgcatctgttgc	Colson & Goldstein 1999b
dronanos	91F13	(AT)18	cgcaagtattcatttcaacaca	tgctggcggttgtttcat	Colson & Goldstein 1999b
dmehab	90B1-B2	(AGCC)5	atttattagttttttgctaatacttgc	agagttcttgttgtatttatac	Colson & Goldstein 1999b
drotropi288F5		(AT)11	gtacatcccgaatcccacac	aatacactgaaactgttggggg	Colson & Goldstein 1999b
dmtrxiii 88B3	3–100.3	(AC)9	gaccgtttgtttgccttgat	tgcctgtacaagtctgaccg	Colson & Goldstein 1999b
dmprosper	86E	(AC)6 (AG)12	caataaccacacgcattcc	aaccacttcctgtttggcc	Colson & Goldstein 1999b
dmu25686	93F	(AT)5	cgataatttactctgtgctcc	cagctcacacaaaaggcaaa	Colson & Goldstein 1999b
dmtf125 95C6-C8	3–119.7		ctcgagcgggccatacaaga	tgattgaagaggccactcaa	Colson & Goldstein 1999b
dmsidna 96E1-E4			tgcacatactcctttcaattcg	aacaggactcggaacaatgc	Schug et al. 1998
DMU43090	99D6-D9	(CAG)8	tgcacccagcaataccagta	gctgttgtcgtggtctg	Colson & Goldstein 1999a

Modified from Colson and Goldstein (1999b) Table 1

Two multiplex PCRs (Table 6.2) were performed at the University of Arizona Genetics Core.□PCR reactions containing 0.625 U *Taq* DNA polymerase, 75mm Tris-HCl, 20mm (NH4)2SO4, 2.5mm MgCl2, 0.01% Tween 20 and 0.2 mm of each dNTP to a total volume of 12.5ul. One ul of DNA and 7 pmol of each primer labeled with three different fluorescent colors (Fam, Hex, Tamra) were added to the PCR mix. The PCR was performed using a hot-start procedure for 5 min at 94°C.

Table 6.2 PCR multiplexes

MultiPlex number	Bins	Locus	Size range in D.sim/D.sech	Fluorescent label
1	**50-100bp**	drotropi2	84	Fam
	101-150bp	dmprosper	110-114	Hex
		dmu36477	103-107	Fam
		drohoxnk4	136-139	Tamra
		dronanos	104-110	Tamra
		dmsidna	*126-128*	*Fam*
	151-200bp	dm22f11t	185-203	Tamra
		AC006414	192-198	Hex
		DMU1043090	180-186	Fam
	201-250bp	dm22f11t	185-203	Tamra
		ac004343	202-212	Fam
	251-300bp	dmtf125	289-297	Fam
		dmu14395	263-281	Tamra
2	**50-100bp**	*dmz60mex*	*85*	Fam
	101-150bp	ac004658	112-126	Fam
		dmcathpo	*108-117*	*Hex*
		AC001665	*134-140*	*Tamra*
		drolamb2a	*150-207*	*Fam*
	151-200bp	dmu25686	153-155	Hex
		dmrhob	*198-206*	*Tamra*
		drolamb2a	*150-207*	*Fam*
	201-250bp	drogpaap	220-226	Fam
		dmtrxiii3	211-214	Hex
		dmrhob	*198-206*	*Tamra*
		drolamb2a	*150-207*	*Fam*
	251-300bp	ac004352	267-283	Hex
	301-350bp	*drodsrc*	*312-317*	*Fam*
	351-400bp	dmehab	353-361	Tamra

After PCR, PCR products were resolved on ABI PRISM® 3730 DNA Analyzer with LIZ™ size standard. The obtained electrophoretograms were checked using GeneMarker1.85 software (SoftGenetics LLC).

6.2.4 Statistics

Species specific markers were identified as follows: All fertile BC4-S and fertile BC4-M samples were checked against each peak position. Fisher exact test was used to calculate the difference between BC4-S and BC4-M. P value less than 0.01 was considered as a significant difference.All markers selected as above were used to compare fertile BC4-S and sterile BC4-S. Fisher exact test was used to determine statistical significance.

6.3 Results

6.3.1 Markers used to distinguish *D. simulans and D. mauritiana*

16 fertile BC4-S and 7 fertile BC4-M were used in the comparison. Four major modes of discrimination between *D. simulans* and *D .mauritiana* were observed; each mode is elucidated with one example below.

6.3.1.1 Single peak in *D. simulans* vs. single peak in *D. mauritiana*

At position 172 (D.sim) or 170 (D.mau) in multiplex 1 (FAM-labeled), all *D. simulans* samples have a single peak at 172 bp (100% over all *D. simulans* samples), and all *D. mauritiana* have a single peak at 170 bp (100% over all *D. mauritiana* samples), and this result was statistically highly significant (Fisher exact test, $P<<0.01$). These peaks were out of the size range of any previously described markers, but closest to DMU1043090 (Colson and Goldstein 1999), a marker at location 99D6-D9 on chromosome 3R with eight CAG repeat (Figure 6.1).

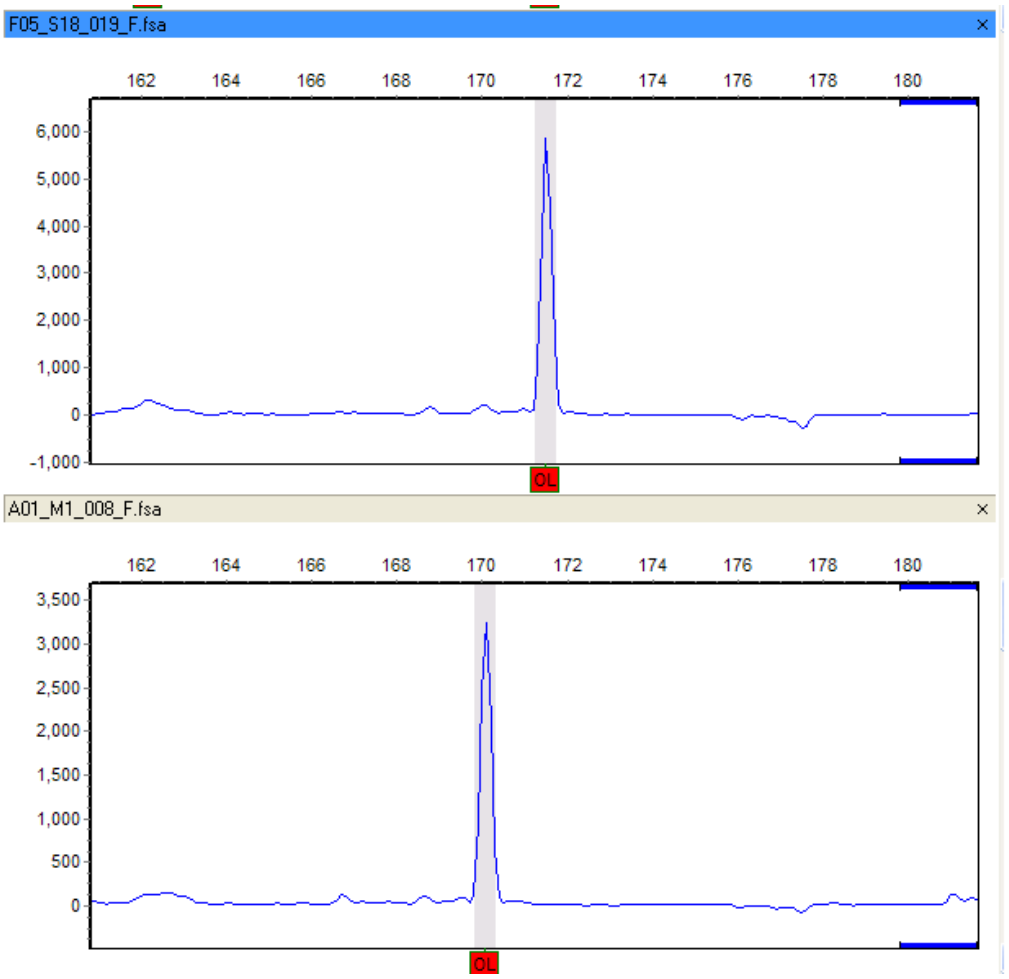


Figure 6.1 Comparison of allele size difference between D. simulans and D. mauritiana (FAM-labeled position 170 vs.172) (Figure on top is a representative of D. simulans peak, the bottom one is a representative of D. mauritiana peak)

6.3.1.2 Single peak in *D. simulans* vs. no peak in *D mauritiana*

At position 165 (D.sim) in multiplex 2 (FAM-labeled), all *D. simulans* individuals have a single peak at 165 bp (100%). Six *D. mauritiana* show no peaks at this position (86%); but one has a single peak at 165bp. This peak falls into the size range of drolamb2a (Colson and Goldstein 1999), a marker located on 67C with five ATT repeat unites (Fisher exact test of presence/ absence of this allele P=0.00009) (Figure 6.2).

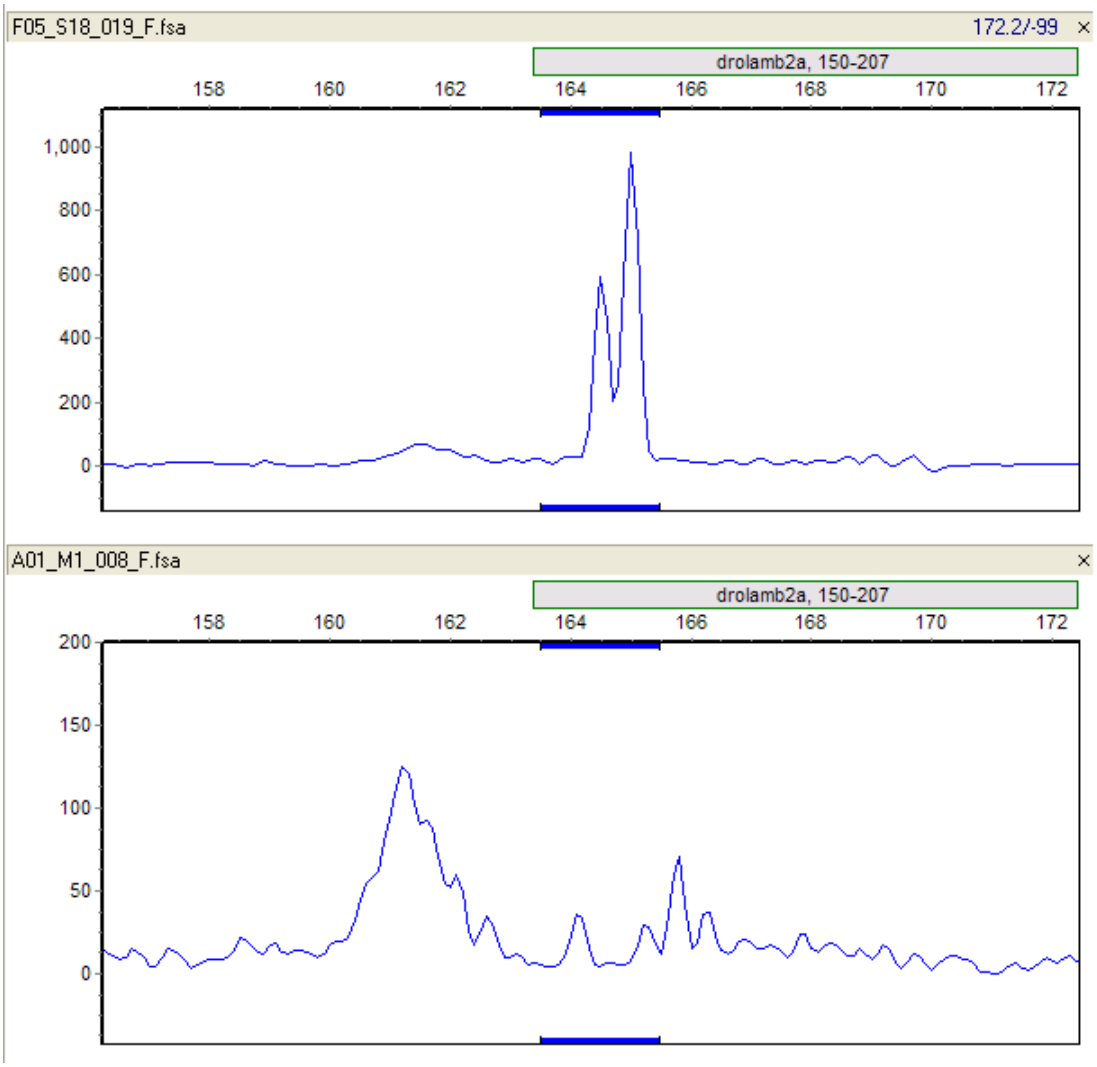


Figure 6.2 Comparison of allele size difference between D. simulans and D. mauritiana (FAM-labeled position 165)

6.3.1.3 No peaks in *D. simulans* vs. single peak in *D. mauritiana*

At position 211 (D.mau) in multiplex 1 (HEX-labeled), none but one *D. simulans has* a single peak at 211 bp (93.75% over all *D. simulans* samples). Five *D. mauritiana* individuals have a single peak at 211 bp (71% over all *D.mauritiana* samples); two have no peaks. This peak is closest to AC006414 (Colson and Goldstein 1999), a marker located on 89A1-A5 with nine GT repeats (Fisher exact test P=0.003) (Figure 6.3).

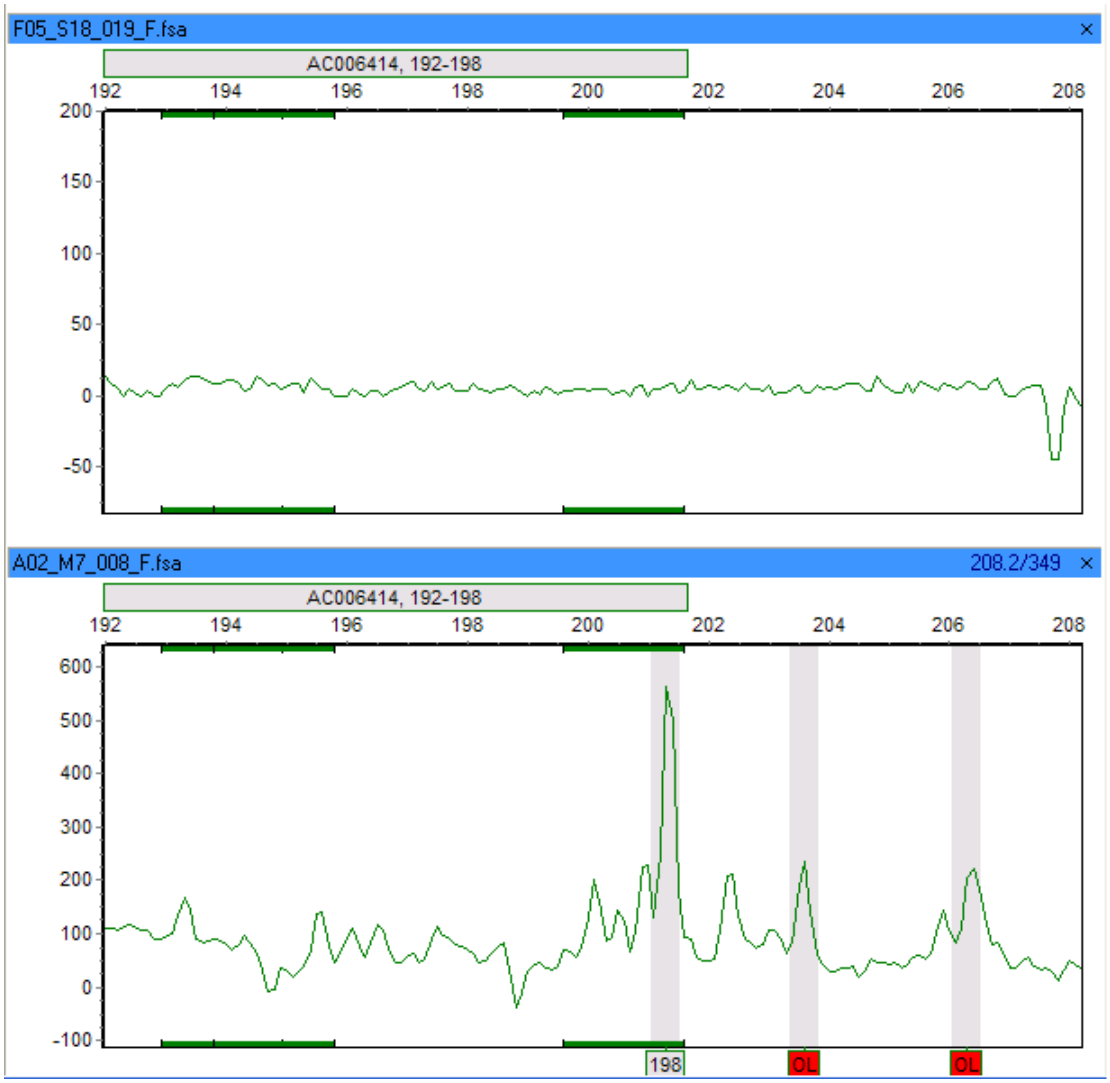


Figure 6.3 Comparison of allele size difference between *D. simulans and D. mauritiana* (HEX-labeled position 211)

6.3.1.4 Multiple peaks in *D. simulans* vs. single peak in *D mauritiana*

At position 207,209 (D.sim) or 213 (D. mau) in multiplex 1 (TAMRA-labeled), *D. simulans* show four different allelic positions. Two individuals show a single peak at 207 bp (12.5% over all *D. simulans* samples) (Figure 6.4a); Two show a single peak at 209bp (12.5%) (Figure 6. 4b); Ten show double peaks at 207 bp and 209 bp(62.5%) (Figure 6.4c); Two individuals show double peaks at 207bp and 213bp (12.5%) (Figure 6.4d). All *D. mauritiana* show a single peak at position 213 (100% over all *D. mauritiana* samples) (Figure 6.4e).

Comparisons at these positions show that pure *D. simulans* may have two types of

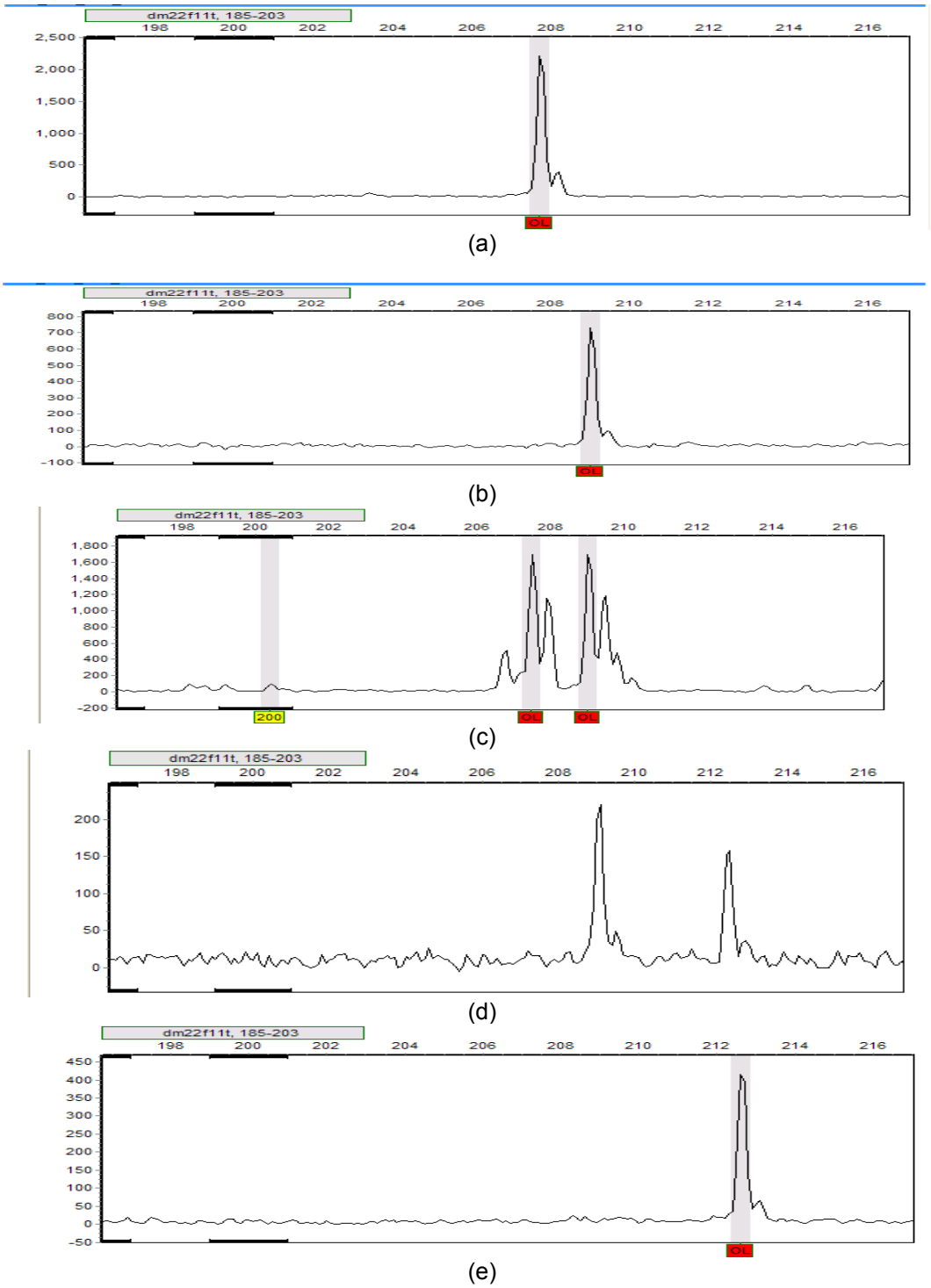


Figure 6.4 Comparison of allele size difference between D. simulans and D. mauritiana (TAMRA-labeled position 207,209 vs.213.Single peak at 207(a); single peak at 209(b); double peaks at 207 and 209(c); double peak at 209 and 213(d); single peak at 213(e))

microsatellite alleles - position 207 and 209. This marker in *D. mauritiana* is two di-nucleotide

repeats larger than in *D. simulans*. This peak is out of the size range of any markers, but closest to dm22f11t (Colson and Goldstein 1999), a marker located on 73A1-B7 with 22 AT repeats (Fisher exact test P=0.0001) (Figure 6.4).

6.3.1.5 Summary of species-specific markers

Eleven markers out of 24 microsatellites were found to distinguish *D. simulans* and *D. mauritiana* with all P value less than 0.01 (Table 6.3).Interestingly, most loci are less polymorphic than in D. melanogaster (the number of alleles in Table 6.3 are compared with that in Colson and Goldstein 1999). The size of microsatellites markers for several loci were also smaller than that in *D. melanogaster*, indicating less repeats in *D. simulans* for some loci. For example, the marker DMU1043090 ranged from 180-186bp in *D. melanogaster*, but 172 bp in *D. simulans*, 170 bp in *D. mauritiana*.

Table 6.3 Microsatellite loci to distinguish between D. simulans and D. mauritiana (P<=0.01)

Locus name	Cytological location	Genetic location	Repeat (in D. mel)	Number of size alleles	
				D.sim	D.mau
dmrhob	62A	3–0	(AC)10	1	1
ac004343	62A1-A2		(CT)11	2	2
DMU43090	99D6-D9	3–132.8	(CAG)8	2	2
drolamb2a	67C	3–37.9	(ATT)5	2	2
AC006414	89A1-A5	3–94.7	(GT)9	1	1
dm22f11t	73A1-B7	3–55.2	(AT)22	2	1
AC001665	84C1-C4	3–69.4	(AC)9	1	0
drohoxnk4	91E1-E3		(CAG)5	2	1
dmtrxiii	88B3	3–100.3	(AC)9	3	2
dmu25686	93F		(AT)5	0	1
dmtf125	95C6-C8	3–119.7	(CAG)6	0	2

6.3.2 Can species-specific microsatellite markers be used for genetic mapping of hybrid male sterility?

Species-specific microsatellites were further tested for their utility in separating fertile and sterile BC4-S. I surveyed microsatellite variation in 16 BC4-S fertile, 2 semi-fertile (less sperm than usual or no motility) and 24 BC4-S sterile males. Though there was no clear distinction between fertile and sterile males in terms of microsatellite variability, several potentially interesting positions were found. In these loci, single peaks indicating homozygosity were present in fertile BC4-S and BC4-M males, but a certain degree of heterozygosity (both peaks present) was observed in sterile individuals.

At position 172 (D.sim) or 170 (D.mau) in multiplex 1 (FAM-labeled), which is closest to DMU1043090, all fertile BC4-S have a single peak at position 172 bp; all fertile BC4-M have a single peak at position 170 bp. Most sterile flies have a single peak as fertile BC4-S, but two samples (8.3%) show combined peaks from both 170 and 172 (Figure 6.5).

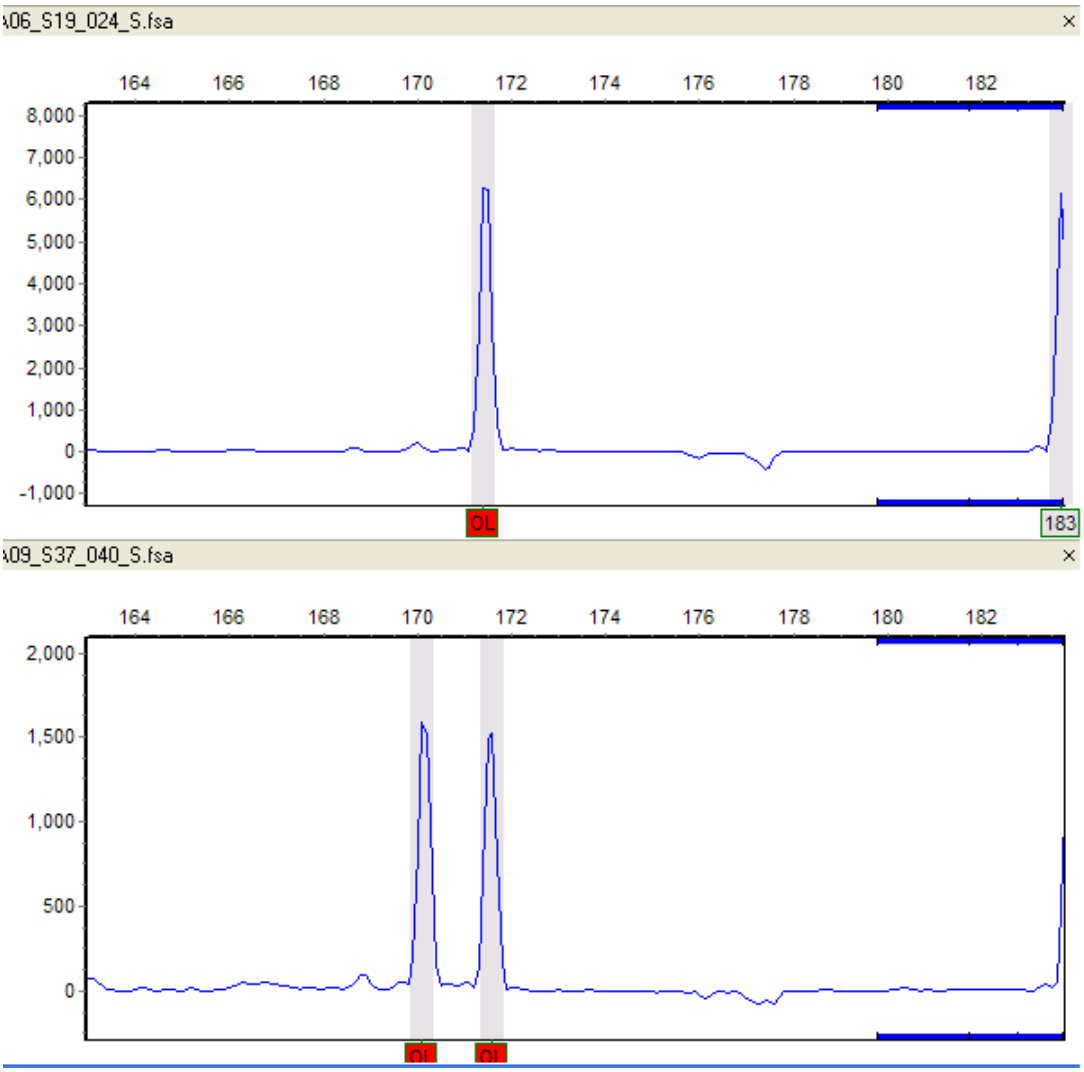


Figure 6.5 Comparison of allele size difference between fertile and sterile males of BC4-S (Position 172 vs.170,172) (Figure on top is a representative of a single peak in fertile BC4-s, the bottom one is a representative of double peaks in sterile BC4-S)

At position 148 (D.sim) or 150 (D.mau) in multiplex 1 (TAMRA-labeled), which is closest to drohoxnk4, majority of the fertile BC4-S have a single peak at position 148; but one (#22) has both peaks at 148 and 150. All fertile BC4-M have a single peak at position 150 bp. Most BC4-S sterile flies have a single peak as in fertile BC4-S, but four (#1, #21,#24,#28) show combined peaks for both position 148 and 150.

Similar comparisons are also used to examine fertile and sterile BC4-M males. Because the sample size is not big enough (6 fertile, 1 SF, 3 sterile), only one position at 202 labeled with

HEX in multiplex 2 seems to show a difference between F and S. Four fertile and one semi-fertile show a single peak at 202; two others have no peaks. All three sterile males have no peak at that position. This peak is in the size range of AC006414 (Figure 6.6).

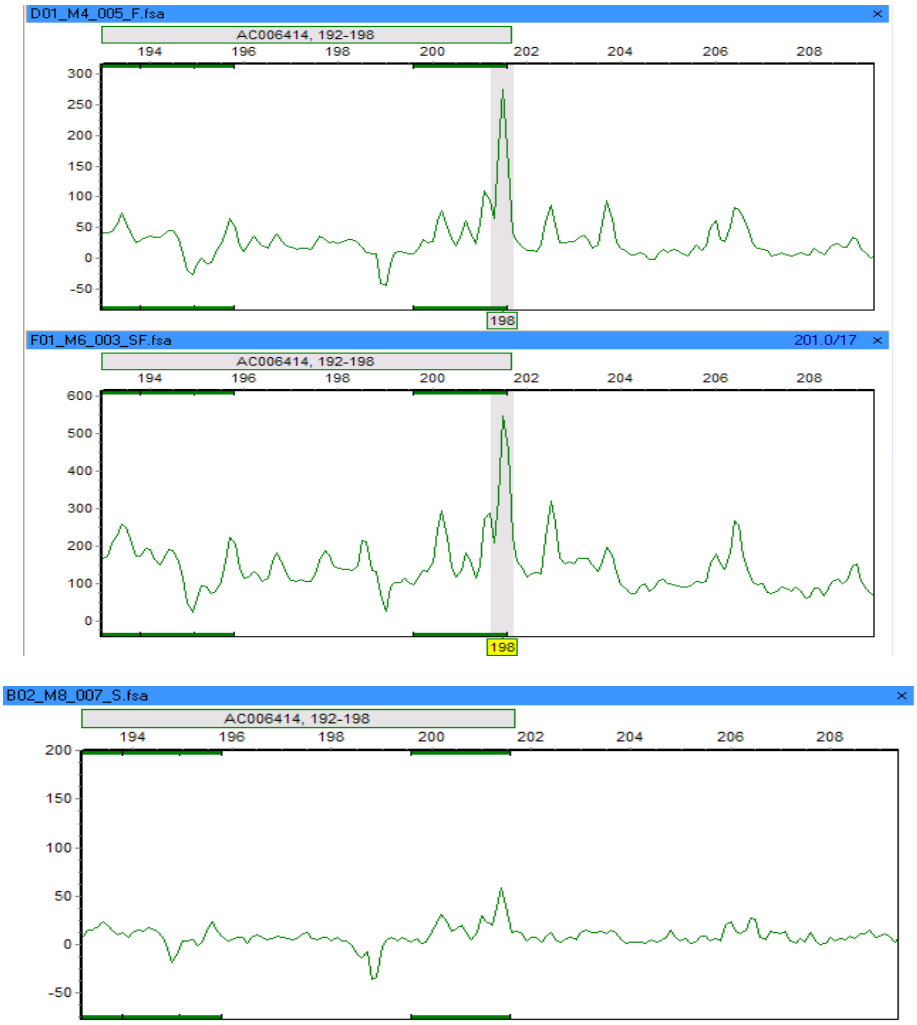


Figure 6.6 Comparison of allele size difference between fertile and sterile BC4-M at Position 202 (Fertile, SF, Sterile from top to bottom, respectively)

6.3.3 Summary of markers with combined peaks in sterile males

After examining all eleven markers, four loci tended to include combined peaks in sterile but not in fertile (Table 6.4, Figure 6.7).

Table 6.4 Microsatellite markers that are possibly associated with male sterility

Locus name	Cytological location	Genetic location
DMU43090, 180-186	99D6-D9	3–132.8
AC006414	89A1-A5	3–94.7
drohoxnk4	91E1-E3	
dmtrxiii	88B3	3–100.3

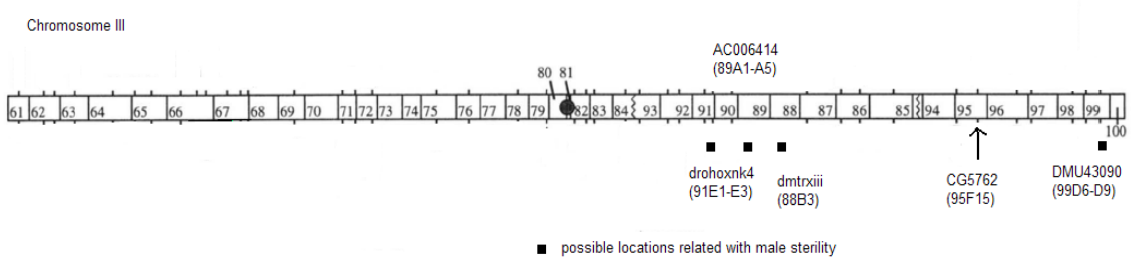


Figure 6.7 Locations of the markers in Chromosome III. (Markers positions are shown with square, the arrow pointed to gene CG5762)

Notably these four microsatellite loci are the closest markers to CG5762 of all species-specific markers described above except dftm125. Cytological location of dftm125 is 95C6-C8, an adjacent region to CG5762 (95F15) but dftm125 did not turn out to be informative in terms of separating fertile and sterile males. Interestingly, all BC4-M regardless of their fertility/sterility pattern showed double peaks at this position, in contrast to BC4-S that had no peaks at all (Figure 6.8).

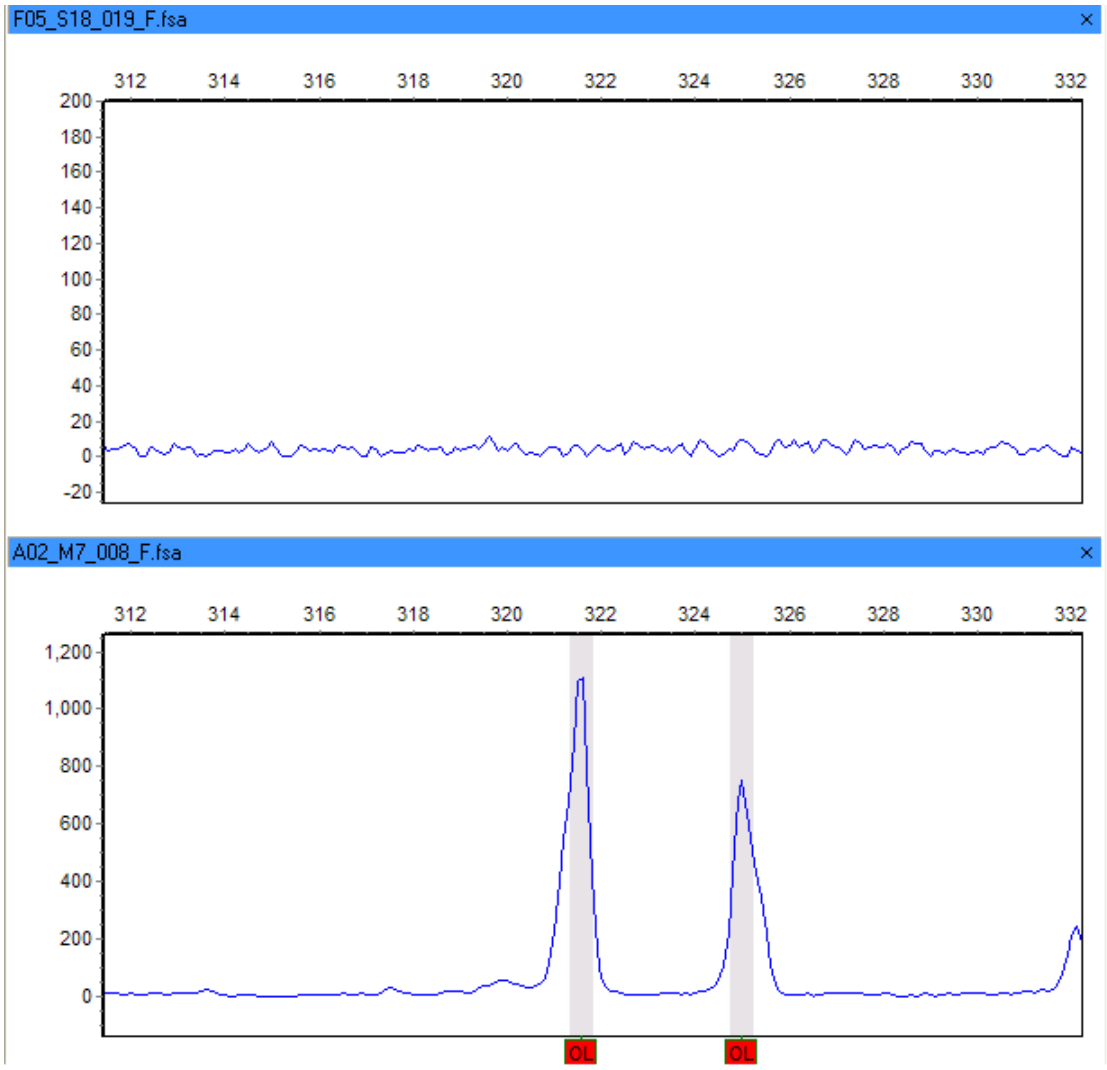


Figure 6.8 Comparison of allele size difference between D. silulans and D. mauritiana (FAM labeled Position 322, 325 in multiplex 1, which is closest to dmtf125)

6.3.4 Heterozygosity in BC4-S sterile males

Besides the type of combined peaks from both *D. simulans* and *D. mauritiana* as mentioned above that may be related to hybrid sterility, some sterile BC4-S also exhibited double peaks, but none of them were present in BC4-M (Figure 6.4(c) vs 4(e)). These loci may indicate heterozygous loci in pure *D. simulans* species. The heterozygous peaks) were checked individually in all BC4-S sterile males. The results indicate that each individual has at least one heteryzogous peak (either the heterozygous loci in pure *D. simulans* or combined from *D.*

mauritiana), but the locations vary from each other. Heteryzogous peaks occur in some locations with high frequency (>10) (red dot in Figure 6.9) but are rare in other (2-3 times; green dot). I discarded all locations in which heterozygous peaks occured only once, as it may have happened by chance alone. If a heterozygous peak occurs more than twice, it may suggest a possible region for sterility related dysfunctions in hybrid.

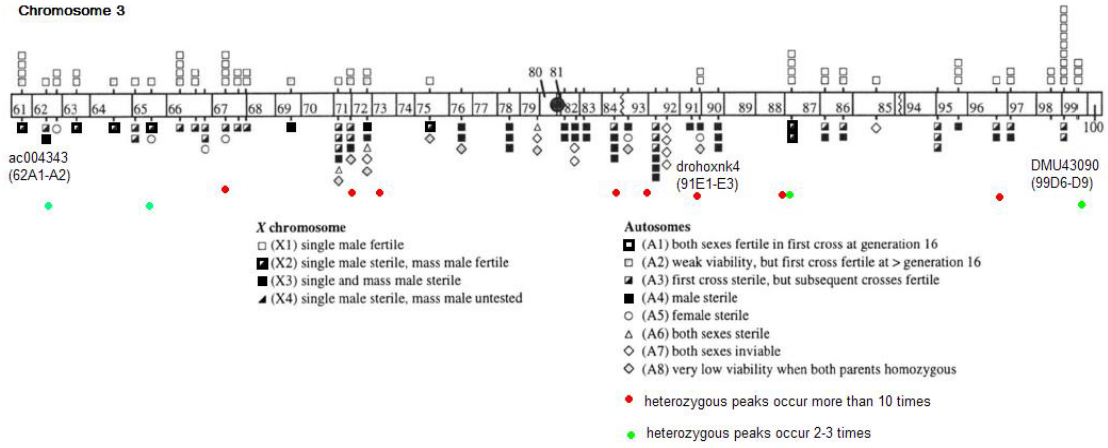


Figure 6.9 Chromosome III sterility markers from True et al. (1996) and the present study (locations marked with black squares are regions associated with male sterility in True et al. (1996); locations with red/green circle are marker locations where heterozygous peaks occured more than twice in the present study)

6.4 Discussion

Out of 24 markers we originally used from Colson's paper (1999) that differentiated *D. simulans* and *D. sechellia*, only 11 markers were found here to separate *D. simulans* and *D. mauritiana*. This lower differentiation seems to support the view that *D. simulans* and *D. mauritiana* have diverged more recently than *D. simulans* and *D. sechellia*.

Most analyzed microsatellite loci were monoallelic in both *D. simulans* and *D. mauritiana*, but those that were polymorphic (more than one allele) were more so in *D. simulans*. For example, *D.simulans* population has two alleles (208bp, 210bp) in dmtxiii locus, but *D.mauritiana* has only one allele (213bp). One possible explanation is that *D. mauritiana* is derived from an ancestral *D. simulans* population, so they shared many ancient polymorphisms but *D. simulans* as an

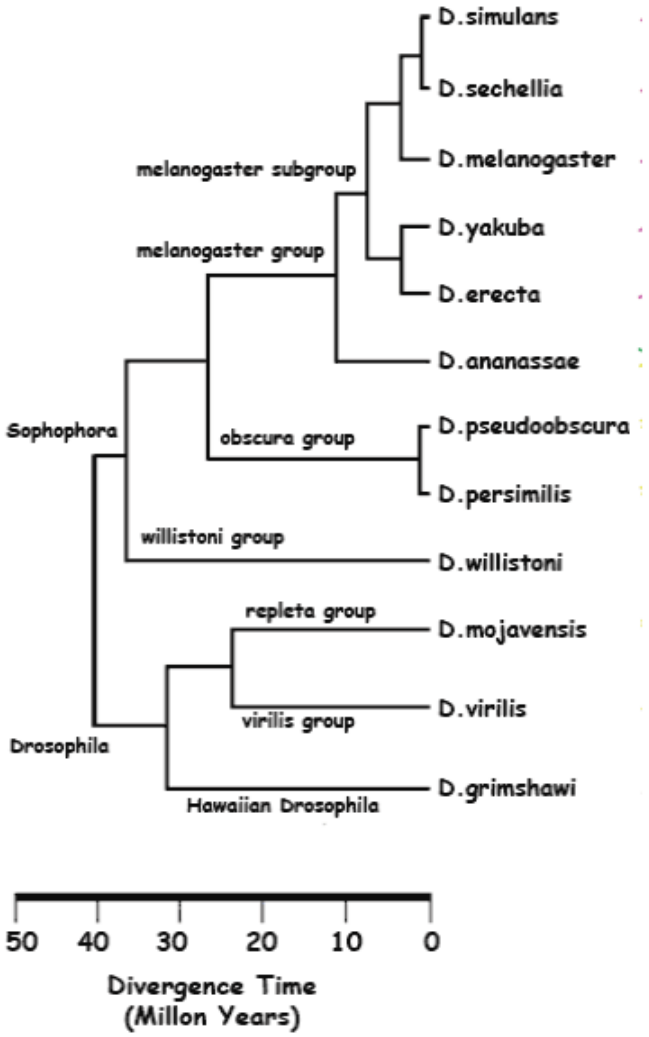
abundant and cosmopolitan species is characterized by higher genetic variability than *D. mauritiana* (Kliman and Hey 1993).

The application of species-specific microsatellite markers to mapping of hybrid sterility indicates that the sterility genes are possibly closer to cytological position 88-91 or position 99 on 3R arm. Gene CG5762 is between them, but the marker (dftm125) adjacent to it did not show variance between fertile and sterile males. Therefore, analysis of more samples with closer genetics markers will be necessary to investigate this question in more detail.

The heterozygosity results in BC4-S sterile males show a consistency with True *et al.* (1996) for most of the locations, but I also find two possible locations around 3-65 and 3-99 mapping to two ends of chromosome III. I propose that intra-species heterozygosity in backcrossed individuals (not introgressed from a different species) may also affect fitness of hybrid males because this may result in less transcripts produced from certain alleles. Although endogenous heterozygosity is fully functional in pure species, insufficient expression may have more dramatic consequences downstream in globally imbalanced transcription stoichiometry on a hybrid genetic background.

APPENDIX A

PHYLOGENETIC TREE OF DROSOPHILA (FROM FLYBASE)


D.simulans
D.sechellia
D.melanogaster
melanogaster subgroup
D.yakuba
melanogaster group
D.erecta
D.ananassae
Sophophora
obscura group
D.pseudoobscura
D.persimilis
willistoni group
D.willistoni
repleta group
D.mojavensis
D.virilis
virilis group
Drosophila
D.grimshawi
Hawaiian Drosophila
50 40 30 20 10 0
Divergence Time
(Millon Years)

APPENDIX B

SUMMARY OF EXISTING GENETIC ANALYSES OF REPRODUCTIVE ISOLATION BETWEEN DROSOPHILA SPECIES (MODIFIED FROM COYNE AND ORR'S TABLE (1998))

Species pair	Trait	Number of genes	Reference
D. heteroneura x D. silvestris	head shape	9	Templeton 1977
D. melanogaster x D. simulans	hybrid inviability	≥9	Pontecorvo 1943.
	female pheromones	≥5	Coyne 1996(a)
D. mauritiana x D. simulans	hybrid male sterility	≥15	Wu *et al.* 1996
	hybrid female sterility	≥4	True *et al.* 1996
	hybrid inviability	≥5	True *et al.* 1996
	male sexual isolation	≥2	Coyne 1996(b)
	female sexual isolation	≥3	Coyne 1996(b)
	genital morphology	≥9	True *et al.* 1997
	shortened copulation	≥3	Coyne 1993
D. mauritiana x D. sechellia	female pheromones	≥6	Coyne and Charlesworth 1997
D. simulans x D. sechellia	hybrid male sterility	≥6	Hollocher and Wu 1996
	hybrid inviability	≥2	Hollocher and Wu 1996
	female sexual isolation	≥2	Coyne 1992
D. mojavensis x D. arizonae	hybrid male sterility	≥3	Pantazidis *et al.* 1993
	male sexual isolation	≥2	Zouros 1981
	female sexual isolation	≥2	Zouros 1981
D. pseudoobscura x D. persimilis	hybrid male sterility	≥9	Orr 1989
	hybrid female sterility	≥3	Orr 1989
	sexual isolation	≥3	Noor 1997
D. pseudoobscura U.S.A. x Bogota	hybrid male sterility	≥5	Orr 1989
D. buzatti x D. koepferae	hybrid male inviability	≥4	Carvajal *et al.* 1996
	hybrid male sterility	≥7	Naveira 1991
D. subobscura x D. madeirensis	hybrid male sterility	≥6	Khadem and Krimbas 1991
D. virilis x D. littoralis	hybrid female viability	≥5	Mitrofanov and Sidorova 1981
D. virilis x D. lummei	male courtship song	≥4	Hoikkala and Lumme 1984

	hybrid male sterility	≥6	Heikkinen and Lumme 1991
D. hydei x D. neohydei	hybrid male sterility	≥5	Heikkinen and Lumme 1991
	hybrid female sterility	≥2	Schafer 1978
	hybrid inviability	≥4	Schafer 1979
D. virilis x D. texana	hybrid male sterility	≥3	Lamnissou *et al.* 1996
D. auraria x D. biauraria	male courtship song	≥2	Tomaru *et al.* .1995

APPENDIX C

IMPROVED PROCEDURE OF RNA EXTRACTION

1. Cut single male fly into head and body parts. Store the head immediately at -80ºC for DNA extraction later.

2. Place the fly body sample in 500µ/ RNase-free tube and homogenize in 70µ/ TRI reagent using a polytron pestle.

3. Incubate the homogenate at room temperature (RT) for 5 minutes to dissociate the nucleoproteins from the nucleic acids.

4. Add 14µ/ chloroform to the homogenate.

5. Cover the sample, shake vigorously for ~20 seconds.

6. Incubate at RT for 10minutes.

7. Centrifuge the mixture at ≥ 10,000 x g for 15 minutes at 4ºC.

8. Carefully transfer the aqueous phase (the top phase) (containing the RNA) into a clean RNase-free tube. (Up to 20µ/)

9. Add 35µ/ RNase-free water and mix well.

10. Add 0,5µ/ glocoblue and mix well.

11. Add 70µ/ isoproponol and mix well.

12. Incubate at RT for 10 minutes.

13. Centrifuge at ≥ 10,000 x g for 15 minutes at 4ºC to pellet the RNA.

14. Decant the supernatant and wash the pellet (blue color) with 70µ/ cold 75% ethanol by vortexing.

15. Centrifuge at ≥ 10,000 x g for 5 minutes at 4ºC. Discard the supernatant.

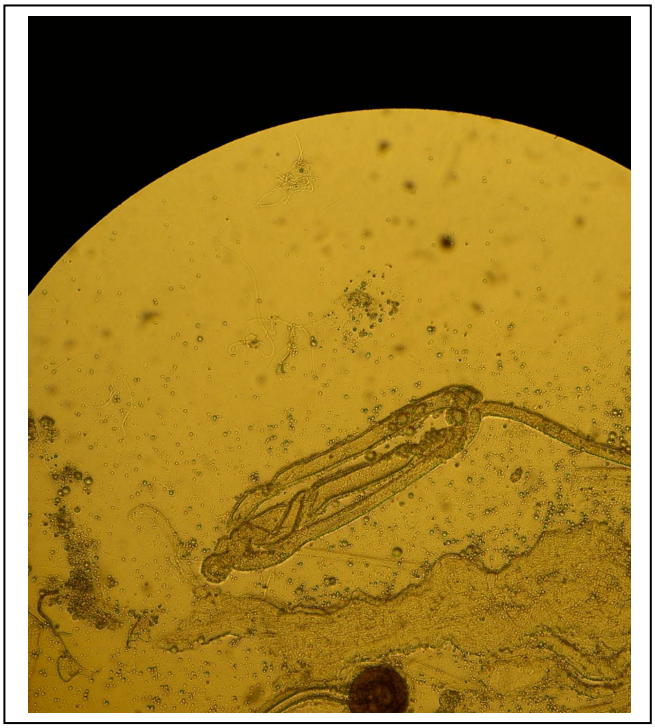
16. Air dry the pellet for about 30 minutes.

17. Resuspend the RNA in 10µ/ RNase-free water and mix well.

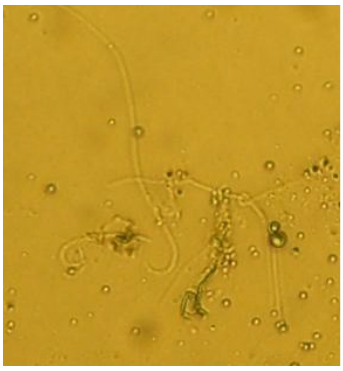
18. Check the quantification of RNA by NanoDrop.

APPENDIX D

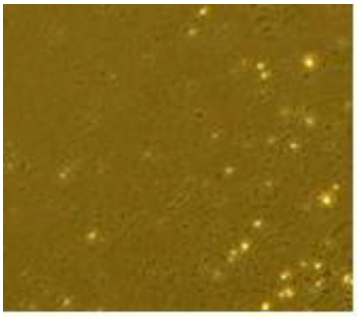
IMAGE OF FEMALE PRODUCTIVE TRACTS AFTER COPULATION WITH SPERMS TRANSFERRED



(a)



(b)



(c)

Microscopy image of female productive tracts (a)
zoom in image with sperms presence (b); zoom in image without sperms (c)

REFERENCES

Adams, M. D., S. E. Celniker, *et al.* (2000) The genome of Drosophila melanogaster. Science 287: 2185-2195.

Ahmadian AB, Gharizadeh AC, Gustafsson AC, Sterky F, Nyren P *et al.* .(2000) Single-nucleotide polymorphism analysis by pyrosequencing. Anal. Biochem. 280: 103–110.

Ahmadian A, Lundeberg J, Nyren P, Uhlen M, Ronaghi M. (2000). Analysis of the p53 tumor suppressor gene by pyrosequencing. Biotechniques 28:140–144.

Alipaz JA, Wu C-I, Karr TL (2001) Gametic incompatibilities between races of Drosophila melanogaster. Proceedings of the Royal Society of London Series B-Biological Sciences, 268, 789-95.

Andrews, J., G. Bouffard, *et al.* (2000). Gene discovery using computational and microarray analysis of transcription in the Drosophila melanogaster testis. Genome Research 10(12): 2030.

Arbeitman MN, Furlong EE, Imam F *et al.* (2002) Gene expression during the life cycle of Drosophila melanogaster. Science, 297, 2270-2275.

Arbeitman MN, Fleming AA, Siegal ML, Null BH, Baker BS (2004) A genomic analysis of Drosophila somatic sexual differentiation and its regulation. Development, 131, 2007-2021.

Bastock, M. and A. Manning (1955). The courtship of Drosophila melanogaster. Behaviour: 85-111.

Batra, R., K. Charizanis, *et al.* (2010). Partners in crime: bidirectional transcription in unstable microsatellite disease. Human Molecular Genetics 19(R1): R77.

Batra, R., K. Charizanis, *et al.* (2010). Partners in crime: bidirectional transcription in unstable microsatellite disease. Human Molecular Genetics 19(R1): R77.

Bodmer M and Ashburner M. 1984. Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of Drosophila. Nature 309:425-430.

Bradley KC, Boulware MB, Jiang H, Doerge RW, Meisel RL, Mermelstein PG (2005) Changes in gene expression within the nucleus accumbens and striatum following sexual experience. Genes Brain and Behavior, 4, 31-44.

Brody T, Stivers C, Nagle J, Odenwald W (2002) Identification of novel Drosophila neural precursor genes using a differential embryonic head cDNA screen. Mechanisms of Development, 113, 41-59.

Bucan M, Abel T (2002) The mouse: genetics meets behaviour. Nature Review Genetics, 3, 622-630.

Capy P, Veuille M, Paillette M, Jallon JM, Vouidibio J, David JR (2000) Sexual isolation of genetically differentiated sympatric populations of Drosophila melanogaster in Brazzaville, Congo: the first step towards speciation? Heredity, 84, 468-475.

Cáceres M, Lachuer J, Zapala MA *et al.* (2003) Elevated gene expression levels distinguish human from non-human primate brains. Proceedings of the National Academy of Sciences of the United States of America, 100, 13030-13035.

Capy P, Veuille M, Paillette M, Jallon JM, Vouidibio J, David JR (2000) Sexual isolation of genetically differentiated sympatric populations of Drosophila melanogaster in Brazzaville, Congo: the first step towards speciation? Heredity, 84, 468-475.

Carvajal AR, Gandarela MR, Naveira HF. (1996) A three-locus system of interspeci¢c incompatibility underlies male inviability in hybrids betwen Drosophila buzzattii and D. koepferi. Genetica 98: 1-19.

Cash AC, Whitfield CW, Ismail N, Robinson GE (2005) Behavior and the limits of genomic plasticity: power and replicability in microarray analysis of honeybee brains. Genes Brain and Behavior, 4, 267-271.

Celniker SE, Wheeler DA, Kronmiller B *et al.* (2002) Finishing a whole-genome shotgun: release 3 of the Drosophila melanogaster euchromatic genome sequence. Genome Biology, 3, 0079.

Chapman T, Liddle LF, Kalb JM, Wolfner MF, Partridge L (1995) Cost of mating in Drosophila melanogaster females is mediated by male accessory gland products. Nature, 373, 241-244.

Chapman T, Bangham J, Vinti G *et al.* (2003) The sex peptide of Drosophila melanogaster: female post-mating responses analyzed by using RNA interference. Proceedings of the National Academy of Sciences of the United States of America, 100, 9923-9928.

Charlesworth, B., Coyne, J.A. and Barton, N.H. (1987) The relative rates of evolution of sex chromosomes and autosomes. Am. Nat.130:113-146

Cherezov, V., D. Rosenbaum, *et al.* (2007). High-resolution crystal structure of an engineered human 2-adrenergic G protein coupled receptor. Science 318(5854): 1258.

Cheung VG, Conlin LK, Weber TM *et al.* (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. Nature Genetics, 33, 422-425.

Chippindale AK, Gibson JR, Rice WR (2001) Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in Drosophila. Proceedings of the National Academy of Sciences of the United States of America, 98, 1671-1675.

Clark AG, Aguade M, Prout T, Harshman LG, Langley CH (1995) Variation in sperm displacement and its association with accessory gland protein loci in Drosophila melanogaster. Genetics, 139, 189-201.

Clausen J, Kech DD, Hiesey WM. (1940). Experimental studies on the nature of species. I. Effect of varied environments on western North America plants. Washington DC. Carnegie Institute of Washington.

Clayton DF (2004) Songbird genomics: methods, mechanisms, opportunities, and pitfalls. Annals of the New York Academy of Sciences, 1016, 45-60.

Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR (1999) A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in Drosophila. Neuron, 22, 203-204.

Cohn VH, Thompson MA, Moore GP.(1984). Nucleotide sequence comparison

Colson, I. and D. Goldstein (1999). Evidence for complex mutations at microsatellite loci in Drosophila. Genetics 152(2): 617.

Colson, I., S. Macdonald, *et al.* (1999). Microsatellite markers for interspecific mapping of Drosophila simulans and D. sechellia. Molecular ecology 8(11): 1951-1955.

Couto, A., M. Alenius, *et al.* (2005). Molecular, anatomical, and functional organization of the Drosophila olfactory system. Current Biology 15(17): 1535-1547.

Cowles CR, Hirschhorn JN, Altshuler D, Lander ES (2002) Detection of regulatory variation in mouse genes. Nature Genetics, 32, 432-437.

Coyne, J. A. (1992). Genetics of sexual isolation in females of the Drosophila simulans species complex. Genet. Res. Camb. 60: 25-31.

Coyne, J. A. (1993). The genetics of an isolating mechanism between two sibling species of Drosophila. Evolution 47: 778-788.

Coyne, J. A. 1996(a). Genetics of differences in pheromonal hydro-carbons between Drosophila melanogaster and D. simulans. Genetics 143: 353-364.

Coyne, J. A. 1996(b). Genetics of sexual isolation in male hybrids of Drosophila simulans and D. mauritiana. Genet. Res. 68: 211-220.

Coyne JA, Elwyn S (2006a) Does the desaturase-2 locus in Drosophila melanogaster cause adaptation and sexual isolation? Evolution, 60, 279-291.

Coyne JA, Elwyn S (2006b) Desaturase-2, environmental adaptation, and sexual isolation in Drosophila melanogaster. Evolution, 60, 626-627.

Coyne JA, Orr HA (1989) Patterns of speciation in Drosophila. Evolution, 43, 362-381.

Coyne JA, Orr HA (1997) Patterns of speciation in Drosophila revisited. Evolution, 51, 295-303.

Coyne, J. & Charlesworth, B. (1997) Genetics of a pheromonal di¡erence a¡ecting sexual isolation between Drosophila mauritiana and D. sechellia. Genetics 145: 1015-1030.

Coyne, J. and H. Orr (2004). Speciation, Sinauer Associates Sunderland, MA.

Dallerac R, Labeur C, Jallon JM, Knipple DC, Roelofs WL, Wicker-Thomas C (2000) A delta 9 desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in Drosophila melanogaster. Proceedings of the National Academy of Sciences of the United States of America, 97, 9449-9454.

Dobzhansky T. (1936) Studies on hybrid sterility. II. Localization of sterility factors in Drosophila pseudoobscura hybrids. Genetics 21: 113-135.

Duerr RH *et al.* (2006). A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene. Science 314:1461-1463

Estoup, A., P. Jarne, *et al.* (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Molecular ecology 11(9): 1591-1604.

Fakhrai-Rad H, Pourmand N, Ronaghi M. (2002). Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms, Hum. Mutat. 19: 479–485.

Fang S, Takahashi A, Wu C-I (2003) A mutation in the promoter of desaturase2 is correlated with sexual isolation between Drosophila behavioral races. Genetics, 162, 781-784.

Frank, S. (1991). Divergence of meiotic drive-suppression systems as an explanation for sex-biased hybrid sterility and inviability. Evolution 45(2): 262-267.

Garcia AC, Ahamdian A, Gharizadeh B, Lundeberg J, Ronaghi M, Nyren P. (2000). Mutation detection by pyrosequencing: sequencing of exons 5 to 8 of the p53 tumour supressor gene, gene 253: 249–257.

Gene-Ontology-Consortium (2001) Creating the gene ontology resource: design and implementation. Genome Research, 11, 1425-1433.

Golub TR, Slonim DK, Tamayo P *et al.* (1999) Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. Science, 286, 531-537.

Grant V. (1994). Modes and origins of mechanical and ethological isolation in angiosperms. Proc. Natl. Acad. Sci. USA 91:3-10

Goldstein, D. and A. Clark (1995). Microsatellite variation in North American populations of Drosophila melanogaster. Nucleic Acids Research 23(19): 3882.

Greenberg AJ, Moran JR, Coyne JA, Wu C-I (2003) Ecological adaptation during incipient speciation revealed by precise gene replacement. Science, 302, 1754-1757.

Greenberg AJ, Moran JR, Wu C-I (2006) Proper control of genetic background with precise allele substitution: A comment on Coyne and Elwyn. Evolution, 60, 623-625.

Greenspan RJ, Ferveur JF (2000) Courtship in Drosophila. Annual Review of Genetics, 34, 205-232.

Grozinger CM, Sharabash NM, Whitfield CW, Robinson GE (2003) Pheromone-mediated gene expression in the honey bee brain. Proceedings of the National Academy of Sciences of the United States of America, 100, 14519-14525.

Guenet JL, Nagamine C, Simon-Chazottes D, Montagutelli X, Bonhomme F (1990) *Hst-3*: an X-linked hybrid sterility gene. Genetical Research 56:163-165

Gu J, Gu X (2003) Induced gene expression in human brain after the split from chimpanzee. Trends in Genetics, 19, 63-65.

Haerty, W., and Singh RS. (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.

Haldane, JBS. (1922). Sex ratio and unisexual sterility in hybrid animals. Genetics 12: 101-109

Hallem, E. and J. Carlson (2004). The odor coding system of Drosophila. TRENDS in Genetics 20(9): 453-459.

Heifetz Y, Lung O, Frongillo EA Jr, Wolfner MF (2000) The Drosophila seminal fluid protein Acp26Aa stimulates release of oocytes by the ovary. Current Biology, 10, 99-102.

Heikkinen E and Lumme J. (1991). Sterility of male and female hybrids of Drosophila virilis and Drosophila lummei. Heredity 67:1-11.

Henderson NR, Lambert DM (1982) No significant deviation. from random mating of world-wide population of Drosophila melanogaster. Nature, 300, 437–440.

Hey, J., and R. M. Kliman, (1993) Population genetics and phylogenetics of DNA sequence variation at multiple loci within the Drosophila melanogaster species complex. Mol. Biol. Evol. 10: 804–822.

Higuchi, R., C. Fockler, G. Dollinger, and R. Watson. (1993). Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Biotechnology (NY) 11:1026-1030.

Hoikkala A and Lumme J. (1984). Genetic control of the difference in male courtship sound between D. virilis and D. lummei. Behav. Genet. 14: 827-845.

Hollocher, H. and Wu, C-I. (1996).The genetics of reproductive isolation in the Drosophila simuians clade: X versus autosomal effects and male versus female effects. Genetics. 143(3):1243-1255

Hollocher, H., C.-T. Ting, F. Pollack, and C.-I Wu (1997a). Incipient speciation by sexual isolation in Drosophilamelanogaster: Variation in mating preference among natural isolates. Evolution 51: 1175-1181.

Hollocher, H., C.-T. Ting, M.-L. Wu, and C.-I Wu (1997b). Incipient speciation by sexual isolation in Drosophilamelanogaster: Extensive genetic divergence without reinforcement. Genetics 147:1191-1201

Hosack DA, Dennis G, Sherman BT, Lane HC, Lempicki RA (2003) Identifying biological themes within lists of genes with EASE. Genome Biology, 4, P4.

Irizarry RA (2005) From CEL files to annotated lists of interesting genes. Biostatistics, 4, 249-264.

Jaenike J. (2001).Sex chromosome meiotic drive. Annu. Rev. Ecol. Syst. 32:25-49

Jarvis, ED, Smith VA, Wada K *et al.* (2002) A framework for integrating the songbird brain. Journal of comparative physiology. A, Neuroethology, sensory, neural, and behavioral physiology, 188, 961-980.

Jiang M, Ryu J, Kiraly M, Duke K, Reinke V, Kim SK (2001) Genome-wide analysis of developmental and sex-regulated gene expression profiles in Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America, 98, 218-223.

Jiggins, CD, Linares M, Naisbit RE, Salazar C, Yang ZH, Mallet J. (2001). Sex-linked hybrid sterility in butterfly. Evolution55:1631-1638

Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G (2001) The contributions of sex, genotype and age to transcriptional variance in Drosophila melanogaster. Nature Genetics, 29, 389-395.

Johnson NJ. *et al.* (1993). The effects of interspecific Y chromosome replacements on hybrid sterility within the Drosophila simuians clade, Genetics 135:443-453

Kaplan WD, Tinderholt VE, Gugler DH (1962) The number of sperm present in the reproductive tracts of Drosophila melanogaster females. Drosophila Information Service, 36, 82.

Kelley, L. and M. Sternberg (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols 4(3): 363-371.

Kelly WG, Schaner CE, Dernburg AF, Lee MH, Kim SK, Villeneuve AM, Reinke V (2002) X-chromosome silencing in the germline of C. elegans. Development, 129, 479-492.

Kendziorski C, Irizarry RA, Chen KS, Haag JD, Gould MN (2005) On the utility of pooling biological samples in microarray experiments. Proceedings of the National Academy of Sciences of the United States of America, 102, 4252-4257.

Keyser, P., K. Borge-Renberg, *et al.* (2007). The Drosophila NFAT homolog is involved in salt stress tolerance. Insect biochemistry and molecular biology 37(4): 356-362.

King MC, Wilson AC (1975) Evolution at two levels in Humans and Chimpanzees. Science, 188, 107-116.

Kirkpatrick M and Ravigne V. (2002). Speciation by natural and sexual selection: models and experiments. Am. Nat. 159:S22-S35

Kliman, R. and J. Hey (1993). DNA sequence variation at the period locus within and among species of the Drosophila melanogaster complex. Genetics 133(2): 375.

K. G. & Oliver, B. (2000). Gene discovery using computational and microarray analysis of transcription in the Drosophila melanogaster testis. Genome Research 10,1841–1842.

Kliman R, Andolfatto P, Coyne JA, Depaulis F, Kreitman M, Berry AJ, McCarter J, Wakeley J, Hey J.(2006).The Population Genetics of the Origin and Divergence of the Drosophila simulans Complex Species. Genetics 156: 1913-1931

Khadem M, Krimbas CB. (1991). Studies of the species barrier between Drosophila subobscura and D. madeirensis. 1. The genetics of male hybrid sterility. Heredity 67: 157-165.

Knapik, E., A. Goodman, *et al.* (1998). A microsatellite genetic linkage map for zebrafish (Danio rerio). Nature Genetics 18(4): 338-343.

Knowlton N, Nate JL, Guzman HM, Rowan R. (1997). Direct evidence for reproductive isolation among the three species of the Montrastraea annularis complex in Central America (Panama and Honduras). Mar. Biol. 127:705-711

Knudsen S (2002) A Biologist's Guide to Analysis of DNA Microarray Data. Wiley-Liss, New York.

Korol A, Rashkovetsky E, Iliadi K, Michalak P, Ronin Y, Nevo E (2000) Nonrandom mating in Drosophila melanogaster laboratory populations derived from closely adjacent ecologically contrasting slopes at Evolution Canyon . Proceedings of the National Academy of Sciences of the United States of America, 97, 12637-12642.

Kuzin B, Tillib S, Sedkov YA, Mizrokhi LI, Mazo A (1994) The Drosophila trithorax gene encodes a chromosomal protein and directly regulates the region-specific homeotic gene fork head. Genes & Development, 8, 2478-2490.

Lamnissou KM, Loukas M, Zouros E. (1996). Incompatibilities between Ychromosome and autosomes are responsible for male hybrid sterility in crosses between Drosophila virilis and Drosophila texana. Heredity 76: 603-609.

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

Lee JE, Choi JH, Lee JH, Lee MG. (2005). Gene SNPs and mutations in clinical genetic testing: haplotype-based testing and analysis. Mutation Research 573(1-2):195-204

Li C, Wong WH (2001a) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proceedings of the National Academy of Sciences of the United States of America, 98, 31-36.

Li C, Wong WH (2001b) Model-based analysis of oligonucleotide arrays: Model validation, design issues and standard error application. Genome Biology, 2, 1-11.

Li C, Wong WH (2003) DNA-Chip Analyzer (dChip). In: The Analysis of Gene Expression Data: Methods and Software (ed. Parmigiani G, Garrett ES, Irizarry RA, Zeger SL), pp. 120-141. Springer, New York.

Liang P, Pardee AB (1997) Differential display: a general protocol. In: Methods in molecular biology: Differential display methods and protocols (ed. Liang P, Pardee AB), pp. 3-12. Humana Press, New Jersey.

Liu LZ, Slininger PJ.(2007). Universal external RNA controls for microbial gene expression analysis using microarray and qRT-PCR. Journal of Microbiological Methods. 68:486-496.

Lo SH, Wang ZN, Hu Y, Yang HH, Gere S, Buetow KH, Lee MP. (2003). Allelic Variation in Gene Expression Is Common in the Human genome. Genome Research. 13:1855-1862.

Lynch JD. (1978). The distribution of leopard frogs (Rana Blairi and Rana pipiens) (Amphibia, Anura, Ranidae) in Nebraska. J. Herpetol. 12:157-162

Mackay T F, Heinsohn SL, Lyman RF, Moehring AJ, Morgan TJ, Rollmann SM (2005) Genetics and genomics of Drosophila mating behavior. Proceedings of the National Academy of Sciences of the United States of America, 102, 6622-6629.

Markow TA, O'Grady PM (2005) Evolutionary genetics of reproductive behavior in Drosophila: connecting the dots. Annual Review of Genetics, 39, 263-291.

Mayr E. (1942). Systematics and the origin of species. Columbia University Press, New York.

Mayr E (1963). Animal Species and Evolution. Belknap Press, Cambridge, MA.

Masly, J. and D. Presgraves (2007). High-resolution genome-wide dissection of the two rules of speciation in Drosophila. PLoS Biol 5(9): e243.

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL (2003) Rapid evolution of male-biased gene expression in Drosophila. Proceedings of the National Academy of Sciences of the United States of America, 100, 9894-9899.

Mello CV, Vicario DS, Clayton DF (1992) Song presentation induces gene expression in the songbird forebrain. Proceedings of the National Academy of Sciences of the United States of America, 89, 6818-6822.

Michalak P, Noor MA (2003). Genome-Wide Patterns of Expression in Drosophila Pure Species and Hybrid Males. Mol Biol Evol 20: 1070-1076.

Michalak, P., and M. A. F. Noor. (2004). Association of misexpression with sterility in hybrids of Drosophila simulans and D. mauritiana. J. Mol. Evol. 59:277-282.

Michalak P, Noor MAF (2006) Genetics of reproductive isolation and species differences in model systems. In: Evolutionary Genetics (ed. Fox C, Wolf JB), pp. 387-398. Oxford University Press, USA.

Mitrofanov VG. and Sidorova NV. (1981). Genetics of the sex ratio anomaly in Drosophila hybrids of the virilis group. Theor. Appl.Genet. 59: 17-22.

Moehring KC, Teeter, Noor MAF. (2007) Genome-wide patterns of expression in Drosophila pure species and hybrid males.II. Examination of multiple species hybridizations, platforms, and life-cycle stages. Molecular Biology and Evolution 24(1):137-145

Morton, R., M. Choudhary, *et al.* (2004). A reanalysis of protein polymorphism in Drosophila melanogaster, D. simulans, D. sechellia and D. mauritiana: effects of population size and selection. Genetica 120(1): 101-114.

Muller, H.J. (1940) Bearing of the Drosophila work on systematics, in The New Sysremotics (Huxley, J.S., ed.), pp. 185-268, Clarendon Press

Muller H J (1942). Isolating mechanisms, evolution, and temperature. Biol. Symp. 6: 71-125.

Naveira H, Fontdevila A. (1991). The evolutionary history of Drosophila buzzatii. XXI. Cumulative action of multiple sterility factors on spermatogenesis in hybrids of D. buzzatii and D. koepferae. Heredity 67: 57-72.

Nordstrom T, Nourizad K, Ronaghi M, Nyren P. (2000). Method enabling pyrosequencing on double-stranded DNA, Anal. Biochem. 282:186–193.

Noor MAF. (1997) Genetics of sexual isolation and courtship dysfunction in male hybrids of Drosophila pseudoobscura and D. persimilis. Evolution 51: 809-815.

Noor, M. (2005). Patterns of evolution of genes disrupted in expression in Drosophila species hybrids. Genetics Research 85(02): 119-125.

Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM (2004) Common pattern of evolution of gene expression level and protein sequence in Drosophila. Molecular Biology and Evolution, 21, 1308-1317.

Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. Nat Genetics, 32, 261-266.

Oliver B, Parisi M (2004) Battle of the Xs. Bioessays, 26, 543-548.

Orr HA. (1989). Localization of genes causing postzygotic isolation in two hybridizations involving Drosophila pseudoobscura.Heredity 63: 231-237.

Orr HA, Masly JP, Presgraves DC (2004) Speciation genes. Current Opinion in Genetics and Development, 14, 675-679.

Orr, H. and Turelli M (2001). The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. Evolution 55(6): 1085-1094.

Ortiz-Barrientos D, Counterman BA, Noor MAF (2004) The genetics of speciation by reinforcement. Public Library of Science (PLoS) Biology 2:e416.

Ortiz-Barrientos D, Counterman B, Noor MAF. (2007) Gene expression divergence and the origin of hybrid dysfunctions. Genetica 129(1): 71-81

Opler PA, Baker HG, Frankie GW. (1975). Reproductive biology of some Costa Rican Cordia species (Boraginaceae). Biotropica 7:234-237

Pantazidis AC, Galanopoulos VK, Zouros E. (1993). An autosomal factor from Drosophila arizonae restores normal spermatogenesis in Drosophila mojavensis males carrying the D. arizonaeYchromosome. Genetics 134: 309-318.

Panhuis TM, Swanson WJ, Nunney L (2003) Population genetics of accessory gland proteins and sexual behavior in Drosophila melanogaster populations from Evolution Canyon. Evolution, 57, 2785-2791.

Parisi M, Nuttal R, Naiman D, Bouffard G, Malley J, Andrews J, Eastman S, Okiver B (2003) Paucity of genes on the Drosophila X chromosome showing male-biased expression. Science, 299, 697-700.

Paulson DR. (1974). Reproductive isolation in damselfies. Syst. Zool. 23:40-49.

Peterson TML, Luckhart S. (2006). A mosquito 2-Cys peroxiredoxin protects against nitrosative and oxidative stresses associated with malaria parasite infection. Free Radic Biol Med. 40(6):1067-1082.

Phadnis, N. and H. Orr (2009). A single gene causes both male sterility and segregation distortion in Drosophila hybrids. Science 323(5912): 376.

Pontecorvo, G. (1943). Viability interactions between chromosomes of Drosophila melanogaster and Drosophila simulans. J. Genet. 45:51-66.

Preuss TM, Caceres M, Oldham MC, Geschwind DH (2004) Human brain evolution: insights from microarrays. Nature Reviews Genetics, 5:850-860.

Rankin CH (2002) From gene to identified neuron to behaviour in Caenorhabditis elegans. Nature Reviews Genetics, 3, 622-630.

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL (2003) Sex-dependent gene expression and evolution of the Drosophila transcriptome. Science, 300, 1742-1745.

Ranz JM and Machado CA. (2006) Uncovering evolutionary patterns of gene expression using microarrays. Trends Ecol Evol 21:29–37

Reinke V, Smith HE, Nance J *et al.* (2000). A global profile of germline gene expression in C. elegans. Molecular Cell, 6, 605-616.

Rice WR, Linder JE, Friberg U, Lew TA, Morrow EH, Stewart AD (2005) Inter-locus antagonistic coevolution as an engine of speciation: assessment with hemiclonal analysis. Proceedings of the National Academy of Sciences of the United States of America, 102, 6527-6534.

Rifkin SA, Kim J, White KP (2003) Evolution of gene expression in the Drosophila melanogaster subgroup. Nature Genetics, 33, 138-144.

Robinson GE, Grozinger CM, Whitfield CW (2005) Sociogenomics: Social life in molecular terms. Nature Reviews Genetics, 6:257-270.

Ronaghi M, Uhlen M, Nyren P. (1998). A sequencing method based on real-time pyrophosphate, Science 281: 363.

Ronaghi M, Nygren M, Lundeberg J, Nyren P. (1999). Analyses of secondary structures in DNA by pyrosequencing, Anal. Biochem. 267: 65–71.

Ronaghi M. (2003) Pyrosequencing for SNP genotyping, Methods Mol. Biol. 212: 189–195.

Ronaghi M, Elahi E. (2002) Pyrosequencing for microbial typing, J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 782: 67–72.

Ronaghi M, Nygren M, Lundeberg J, Nyren P. (1999) Analyses of secondary structures in DNA by pyrosequencing, Anal. Biochem. 267: 65–71.

Ronaghi, M., Uhle´n, M., and Nyre´n, P. (1998) A sequencing method based on real-time pyrophosphate. Science 281, 363–365.

Saetre P, Lindberg J, Leonard JA *et al.* (2004) From wild wolf to domestic dog: gene expression changes in the brain. Brain Research Molecular Brain Research, 126, 198-206.

Sasa M, Chippindale PT, Johnson NA. (1998). Patterns of postzygotic reproductive isolation in frogs. Evolution, 52, 1801-1810.

Sandler, L. and E. Novitski (1957). Meiotic drive as an evolutionary force. American Naturalist 91(857): 105-110.

Schmittgen, T.D. (2001). Real-time quantitative PCR. Methods 25:383-385.

Schafer U. (1978). Sterility in Drosophila hydei_D. neohydei hybrids. Genetica 49: 205-214.

Schafer U. (1979). Viability in Drosophila hydei_D. neohydei hybrids and its regulation by genes located in the sex heterochromatin. Biol. Zentralblatt 98: 153-161.

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. Chromosoma 109(6): 365-371.melanogaster provides evidence for allele-specific mutation rates. Molecular biology and evolution 15(10): 1269.

Schlotterer, C. and D. Tautz (1992). Slippage synthesis of simple sequence DNA. Nucleic Acids Research 20(2): 211.

Schug, M., T. Mackay, *et al.* (1997). Low mutation rates of microsatellite loci in Drosophila melanogaster. Nature Genetics 15(1): 99-102.

Schug, M., K. Wetterstrand, *et al.* (1998). The distribution and frequency of microsatellite loci in Drosophila melanogaster. Molecular ecology 7(1): 57-70.

Seehausen O, van Alphen JJM, Witte F (1997) Cichlid Fish Diversity Threatened by Eutrophication That Curbs Sexual Selection. Science, 277, 1808-1811.

Shaw KL. (1996). Polygenic inheritance of a behavioral phenotype: interspecif genetics of song in the Hawaiian cricket genus laupala.Evolution 50:256-266

Sobrino B, Brión M, Carracedo A.SNPs in forensic genetics: a review on SNP typing methodologies Forensic Science International 154(2-3): 181-194

Spieth HT, Hsu TC. (1950). The Influence of Light on the Mating Behavior of Seven Species of the Drosophila melanogaster Species group. Evolution. 4:316-325

Spieth, H. (1974). Courtship behavior in Drosophila. Annual Review of Entomology 19(1): 385-405.

Streisinger, G., Y. Okada, *et al.* (1966). Frameshift mutations and the genetic code, Cold Spring Harbor Laboratory Press.

Sturtevant, A. (1915). Experiments on sex recognition and the problem of sexual selection in Drosoophilia. Journal of Animal Behavior. Vol 5(5): 351-366.

Sturtevant AH. 1920. Genetic studies on Drosophila simulans. I. Introduction. Hybrids with Drosophila melanogaster. Genetics 5:488-500

Sucher NJ, Awobuluyi M, Choi YB, Lipton SA (1996) NMDA receptors: from genes to channels. Trends in Pharmacological Sciences, 17, 348-355.

Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF (2001) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. Proceedings of the National Academy of Sciences of the United States of America, 19, 7375-7379.

Takahashi A, Tsaur SC, Coyne JA, Wu C-I (2001) The nucleotide changes governing cuticular hydrocarbon variation and their evolution in Drosophila melanogaster. Proceedings of the National Academy of Sciences of the United States of America, 98, 3920-3925.

Templeton, A. R. (1977) Analysis of head shape di¡erences between two interfertile species of Hawaiian Drosophila. Evolution 31: 630-641

Ting CT, Takahashi A, Wu CI. (2001). Incipient speciation by sexual selection: Concurrent evolution at multiple loci. Proc Natl Acad Sci USA 98: 6709–6713.

Tomaru M, Matsubayashi H, OgumaY. (1995). Heterospeci¢c inter-pulse intervals of courtship song elicit female rejection in Drosophila biauraria. Anim. Behav. 50: 905-914.

True, J.R., Weir, B.S. and Laurie, C.C. (1996). A genome-wide survey of hybrid incompatibility factors by the Introgression of marked segments of Drosophila mauritiana chromosomes into Drosophila simulans, Genetics 142:819-837

True, J. R., Liu, J., Stam, L. F., Zeng, Z.-B. & Laurie, C. C. (1997). Quantitative genetic analysis of divergence in male secondary sexual traits between Drosophila simulans and D. mauritiana. Evolution 51: 816-832.

Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America, 98, 5116-5121.

Uddin M, Wildman DE, Liu G *et al.* (2004) Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. Proceedings of the National Academy of Sciences of the United States of America, 101, 2957-2962.

Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axe R (1999) A spatial map of olfactory receptor expression in the Drosophila antenna. Cell, 96, 725–736.

Vosshall LB, Wong A, Axel R (2000) An Olfactory Sensory Map in the Fly Brain. Cell, 102, 147-159.

Wang H, McArthur ED, Sanderson SC, Graham JH, Freeman DC. (1997). Narrow hybrid zone between two subspecies of big sagebrush (Artemisia tridentata: Asteraceae). IV. Reciprocal transplant experiments. Evolution 51:95-102

Whitehead, A., and D. L. Crawford. (2006). Variation within and among species in gene expression: raw material for evolution. Mol. Ecol. 15:1197-1211.

Whitfield CW, Cziko AM, Robinson GE (2003a) Gene expression profiles in the brain predict behavior in individual honey bees. Science, 302, 296-299.

Whitfield CW, Brillet C, LeConte Y, Robinson GE (2003b) Behavioral plasticity and gene expression in the brain: microarray analysis of the influences of environment and genotype on behavioral maturation in the honeybee. Society for Neuroscience Abstracts, 627, 15.

Wittwer, C.T., K.M. Ririe, R.V. Andrew, D.A. David, R.A. Gundry, and U.J. Balis. (1997). The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. BioTechniques 22:176-181.

Wolfner MF, Harada HA, Bertram MJ *et al.* (1997) New genes for male accessory gland proteins in Drosophila melanogaster. Insect Biochemistry and Molecular Biology, 27, 825-834.

Woolley SC, Sakata JT, Crews D (2004) Tracing the evolution of brain and behavior using two related species of whiptail lizards: Cnemidophorus uniparens and Cnemidophorus inornatus. Institute of Laboratory Animal Resources (Ilar) Journal, 45, 46-53.

Wu, C-I. and Davis, A.W. (1993). Evolution of postmating reproductive isolation: The composite nature of Haldane's rule and its genetic bases, Am. Nat 142:187-212

Wu C-I., Hollocher H, Begun DJ, Aquadro CF, Xu Y, Wu ML (1995) Sexual isolation in Drosophila melanogaster: a possible case of incipient speciation. Proceedings of the National Academy of Sciences of the United States of America, 92, 2519-2523.

Zeng L-W. and Singh RS. (1993). The genetic basis of Haldane's rule and the nature of asymmeMc hybrid male sterility between Drosophila simuians, D. mauritiana and D. secheiiia, Genetics 134:251-260

Zouros E. (1981) The chromosomal basis of sexual isolation in two sibling species of Drosophila: D. arizonensis and D. mojavensis.Genetics 97: 703-718.

BIOGRAPHICAL INFORMATION

Daina Ma got her bachalor's degree in Hebei University, China in 2005. Since then she joined in the Quantitative Biology PhD program in University of Texas at Arlington. Her study is mainly focused on gemonic bases of species isolation mechanisms using fruit flies as the model.