Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Tianai Sun 4/11/2014

Speciation driven by internal genetic conflicts: a test in *Drosophila* by an introgression study

by

Tianai Sun

Yun Tao
Adviser

Department of Biology

Yun Tao
Adviser
Arri Eisen
Committee Member
Paul Lennard
Committee Member

4/11/2014

Speciation driven by internal genetic conflicts: a test in *Drosophila* by an introgression study

By

Tianai Sun

Yun Tao, PhD
Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Biology

2014

Abstract

Speciation driven by internal genetic conflicts: a test in *Drosophila* by an introgression study

Tianai Sun

Speciation, the evolution of reproductive isolations (RI) between two incipient species, has long been a focus of evolutionary genetics. Hybrid male sterility (HMS) is typically the fastest evolved one among all RI mechanisms. The evolution of HMS has been thought to be the consequence of adaptive evolution, but the nature of the adaptation remains elusive until recently a major role is attributed to intragenomic conflicts, as illustrated in meiotic drive systems. Typically, a meiotic drive system consists of distorter, responder and suppressor. Chromosome harboring the distorter gains fitness advantage by favoring the transmission of its own on the cost of its homolog. This must be assisted by the tightly linked responder to distinguish itself from its homolog. Meanwhile, suppressor(s) is required to evolve as a counterbalancing force of the distorter because meiotic drive causes fitness loss to the genes not closely linked to it. Due to the lack of recombination between the X and Y, meiotic drive is most likely to evolve on sex chromosomes. The "drive theory" posits that the perpetual conflicts between the three elements disturb the balance of meiosis genes, thus leading to fast evolution of HMS. One prediction of the "drive theory" is that the same set of genes are capable of controlling both sex ratio and fertility in the male hybrids between two very incipient species such as *Drosophila albomicans* and *D. nasuta*. Quantitative trait loci (QTL) mappings have implicated 5 regions that underlie either sex ratio or HMS, or both. To increase the accuracy of mapping, I used molecular markers to assist the introgression from *D. albomicans* with various combinations of the five QTL into the *D. nasuta* background. As expected, every individual region expresses dual function of sex ratio and male fertility. For one QTL region (suppressor) spanning the centromere of the 2nd chromosome, I generated shorter introgressions for fine mapping. The result suggests that this suppressor region can be further partitioned to three factors, allworking to decrease sex ratio while increase fertility. We therefore conclude that genes affecting sex ratio also contribute to male fertility, as predicted by the "drive theory".

Speciation driven by internal genetic conflicts: a test in *Drosophila* by an introgression study

By

Tianai Sun

Yun Tao, PhD
Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Biology

2014

Acknowledgements

I would like to sincerely thank Dr. Yun Tao, who offered me this great opportunity to finish my honor thesis and became involved in his lab. Without his incessant help, I don't know how far I can go; Dr. Arri Eisen for his generous help on my presentation and coordination with the Honor program; Dr. Paul Lennard for his inspiring advices on revision of the thesis; my lab colleagues, Linbin Zhang and Hailian Xiao for their omnipotent technical supports whenever I encountered trouble; Dr. Pat Marstellar for her coming to my final defense and supporting me with encouragements; my roommate and friends for their feedbacks on my thesis and supports; my parents for reminding me of how much I love what I am doing and encouraging me to persist to the end; all staffs in biology department for their assistance in my academic life.

Table of Contents

Introduction

Hybrid Male Sterility and Haldane's Rule

Speciation is the fundamental mechanism leading to biodiversity, including evolution of our own species from other primates. By definition, it takes place when two genomes become incompatible in the hybrids either as hybrid sterility or inviability. Conventionally, hybrid incompatibility (HI) is thought to be the consequence of divergent adaptive evolutions of the two genomes (species) to external factors. However, in this paper I test a hypothesis that internal factors—genetic conflicts within a genome—are also making key contributions to HI.

An observation called Haldane's rule provides basic framework to understand the evolution of HI. J. B. S. Haldane observed that the HI phenotypes are more likely to be expressed in the heterogametic sex (males in XY or females in ZW sex determining systems) than the homogametic sex (Haldane 1922). The explanations for Haldane's rule, though not fully established, are thought to have four candidates: the dominance theory, the "large X" effect, the "faster male" and the "drive" theory. The dominance theory states that heterogametic sex is more likely to be affected by the deleterious HI factors on the X because they are generally recessive (Turelli and Orr 2000). However, the dominance theory only provides a physiological model (which might be true, particularly for hybrid inviability) but does not have any evolutionary implications. The "large X" effect is the observation that hybrid male sterility (HMS) genes are more enriched on the X than autosomes (Tao *et al*. 2003; Masly and Presgraves 2007), possibly caused by easier fixation of X-linked advantageous but recessive mutations (Charlesworth *et al*. 1987). The "faster male" is based on the observations that HMS evolves faster than other two kinds of HI including hybrid female sterility and hybrid inviability, probably due to sexual selection (Wu and Davies 1993). But the "faster male" hypothesis works against Haldane's rule in the ZW system and

thus unlikely to be universally applicable (Tao and Hartl 2003).

The last candidate, the "drive" theory, posits that genetic conflicts between the sex chromosomes and autosomes over sex ratio play predominant roles in genome evolution. This theory also provides ultimate evolutionary explanations for the two observed patterns of "faster male" and "large X" effects (Meiklejohn and Tao 2010). In this paper, my goal is to test this theory with experimental data. Before introducing our experiment, I will discuss the basic compositions of the meiotic drive system: selfish genetic elements and modifiers.

Selfish Genetic Elements

As conventionally accepted, Mendel's first law elaborates that during a typical meiosis in the nucleus of a diploid, two homologous chromosomes disjoin equally with random transmission to the next generation. However, selfish genetic elements can twist the Mendelian mechanism to gain more than 50% chance of transmission (Sandler and Novitsiky 1957). Even though selfish genetic elements usually decrease the host's fitness, they still can maintain their presence in a population, and the diversity of selfish genetic elements goes beyond one's imagination (Hurst and Werren 2001). Among all types of selfish genes, transposable elements and segregation distorters (meiotic drive) are the two most publicized because of their importance in biological functions and evolution. Moreover, theoreticians proposed a causal link between meiotic drive and speciation two decades ago (Hurst and Pomiankowski 1991; Frank 1991).

Meiotic drive has attracted many geneticists' eyes since the mid 20th century when Larry Sandler first conceptualized it. He observed "nonrandom disjunction" of re-arranged chromosomes in *Drosophila melanogaster* and named it as "meiotic drive", implying a phenomenon happening during meiosis (Sandler and Novitsky 1957). At about the same time, the Segregation Distortion (SD) system in *D. melanogaster* was discovered (Hiraizumi and

Crow 1957). Extensive studies on SD in *Drosophila* and the t-complex in *Mus* provided more insights into the mechanism and evolution of meiotic drive (Sandler *et al*. 1959). Morphologically, in the late 70s, abnormality caused by meiotic drives was observed to occur at post-meiotic stages when sperms started to condense (Tokuyasu *et al*. 1977; Lyon 2003). This observation strikingly disproved Sandler's conjecture and switched the focus to the post-meiotic stages. However until then, the molecular etiology of meiotic drive was not known. Recently, research on t-complex disclosed some parts of the mystery. One of the distorters is identified as *Fgd2*, which encodes a G protein regulating the signaling pathway and was speculated to be involved in nuclear condensation. Its transcription occurs at early stages of meiosis I during the spermatogenesis (Bauer *et al*. 2007). The SD distorter is a gene sharing the same molecular mechanism and signaling pathway as *Fdg2*, which also causes post-meiotic abnormality (Kusano, *et al*. 2003).

The three components mentioned next are necessary for a meiotic drive system to work: the distorter, the responder and the suppressor. The distorter skews the chance of transmission of itself to more than 50%. It only works when it can tell its own allele from its homolog, because otherwise suicide may occur. The responder is the DNA cue assisting the distorter to achieve this goal. If the responder is insensitive, then the distorter will not cause harm to its carrier. Thus tight linkage between distorter and insensitive responder is a prerequisite for any meiotic drive to evolve. In both SD and t-complex, the distorter(s) and insensitive responders are locked together by chromosomal inversions. The third component, the suppressor, is then required to evolve because meiotic drive is generally harmful for all the genes not closely linked to the distorters and need to be controlled when the population is severely off balance. Thus, many genes in a genome would potentially evolve as suppressors to counterbalance the distorter.

The tight linkage between the distorter and the responder required for a meiotic drive

system to evolve is most easily to be satisfied on sex chromosomes, because of a lack of recombination between the heteromorphic X and Y chromosomes. Indeed this has already been supported by the discoveries of close to 20 meiotic drive systems on sex chromosomes so far (Jaenike 2001). However, certain discovery bias might exist because sex-linked meiotic drives manifest sex ratio distortion (SRD) in the offspring, a phenotype can be easily scored; in contrast, autosomal meiotic drive cannot be observed directly.

Fisher's principle provides evolutionary explanation for equal sex ratio in bisexual populations, which is that natural selection pushes equal investment to both sexes and thus maximizes the fitness of the genes controlling sex ratio (Fisher 1930). This insight seems to have been gained by Karl Düsing long before Fisher (Düsing 1883). Later in 1967, Hamilton pointed out two fundamental assumptions underlying Fisher's principle: the population must be large and panmictic; the controlling genes are autosomal (Hamilton 1967). For sex-linked genes, the Fisherian sex ratio does not apply. For example, the optimal sex ratio for the X-linked genes in XY male is 100% female while for Y-linked genes it is 0% female. As what Hamilton (1967) first noticed, SRD can cause much greater disturbance to the whole genome because SRD violates Fisher's principle of equal sex ratio for most genes in a genome and therefore much more serious evolutionary consequences are anticipated.

Modifiers of SRD

There is an intrinsic conflict between the sex chromosomes and autosomes over sex ratio because their optima differ. Therefore, the initial arising of SRD inevitably induces defensive responses by evolving autosomal suppressors, which also triggers the evolution of the second distorter. The alternate evolution of distortion-suppression can repeat perpetually (Hurst and Pomiankowski 1991). Because this system only works in the heterogametic sex (more precisely, the XY male or ZW female meiosis), males are predicted to be on the faster

track of evolution than females (in XY system). In fact, SRD might also ultimately cause meiotic sex chromosome inactivation (MSCI), a ubiquitous phenomenon observed in XY male and ZW female across a wide range of taxa from *Drosophila* to *Mus* (Lifschytz *et al*. 1972; Turner 2007), for the purpose to silence SRD on the X. In *Mus*, there is a causal linkage between the disruption of MSCI and hybrid male sterility (Campbell et al. 2013). However, whether all X-linked HMS genes are related to MSCI or not is still to be studied. The "drive theory" predicts that "faster male" evolution pushes faster evolution of HMS genes because SRD happens in XY males. (Of course, "faster female" evolution is expected to happen in ZW females but there is no empirical data as supports yet).

SRD also provides a plausible explanation for the observed "large X" effect (Meiklejohn and Tao 2010). Because the genetic conflicts over sex ratio are between sex chromosomes and autosomes, genetic alterations should be roughly shared 50% to 50% by them. As sex chromosomes contain fewer genes than autosomes, for example in *D. melanogaster* the X chromosome covers around 22.4Mbp as opposed to ~97.9Mbp covered by autosomes (St. Pierre *et al*. 2014), the X is expected to be enriched for genetic changes driven by SRD as well as HMS genes eventually. Thus the "large X" effects are observed in multiple cases.

From crosses between two populations or two incipient species, we expect to see the presence of incompatibilities between two independently evolved SRDs in the hybrid offspring. Further, if there is a causal link between SRD and HMS, many genes expressing HMS should also express SRD in the hybrid between two species. This is the central prediction for the "drive theory" and the focus of my empirical study reported here.

The *Drosophila* Model

D. nasuta and *D. albomicans* are two closely related fruit fly species and their

speciation is very recent because HMS has just begun to emerge (Kitagawa 1990). Furthermore, sex ratio distortion was found in the F1 hybrid males produced by the cross between certain strains of *D. albomicans* and *D. nasuta* (Yang *et al*. 2004). These evidences thus provide me with an ideal model to test the "drive theory" related to SRD and HMS.

These two species have different distributions. *D. albomicans* is found in the area from Japan to East India, while *D. nasuta* is from East Africa to continental India (Kitagawa *et al*. 1982). Moreover, these two species differ conspicuously in karyotype. *D. nasuta* inherited the ancestral karyotype (2n=8) with an X/Y pair and three autosomal pairs (2, 3 and 4), while *D. albomicans* have reduced number of chromosomes (2n=6) as a result of Robertsonian fusions between the sex (X and Y) and the 3rd chromosomes to form the X-3/Y-3 neo-sex chromosomes (Ranganath and Ramachandra 1994). Also notably, that the 3rd chromosome in *D. nasuta* is a large acrocentric consisting of two Muller's elements (C and D). In lab experiments with standard procedures, these two species can be easily crossed and strong HMS only appears in F2 or latter generations.

Preliminary Data

In the following section, I will describe preliminary data collected before or concurrently with my thesis research.

Inbred Lines for Genetic Mapping

Genetic mapping requires pure lines that are fixed for DNA variants and free of inversions to avoid blockage of recombination. For this reason, three inbred lines were constructed by 15 generations of sibling pairings. Two lines, alb2 and shl2, were constructed from the *D. albomicanss* stocks Miyakojima (Okinawa, Japan) and Shiilong (India), respectively, and the third line nas3 was constructed from the *D. nasuta* stock G86

(Mauritius). Unfortunately all three lines are polymorphic with inversions on either one or both of the X-3 and the 2nd chromosomes. With molecular marker-assisted selection, we successfully constructed sublines free of inversions, alb267 and nas314, but failed to do so for shl2. But we succeeded in associating polytene band sequences with molecular markers on the two haplotypes in Shl2. The polytene band sequence of nas314 has an inversion on the 3rd chromosome as compared to alb267, thus blocking recombination between them on the whole 3rd chromosome and introducing difficulties to genetic mapping for about 40% of the genome between these two lines. Fortunately, one of the haplotypes in shl2 (shl2-hap1) is homosequential to alb267, which has the standard banding sequences of *D.albomicans* according to Lin *et al*. (1974) and offers a remedy for our further study (Figure 1).

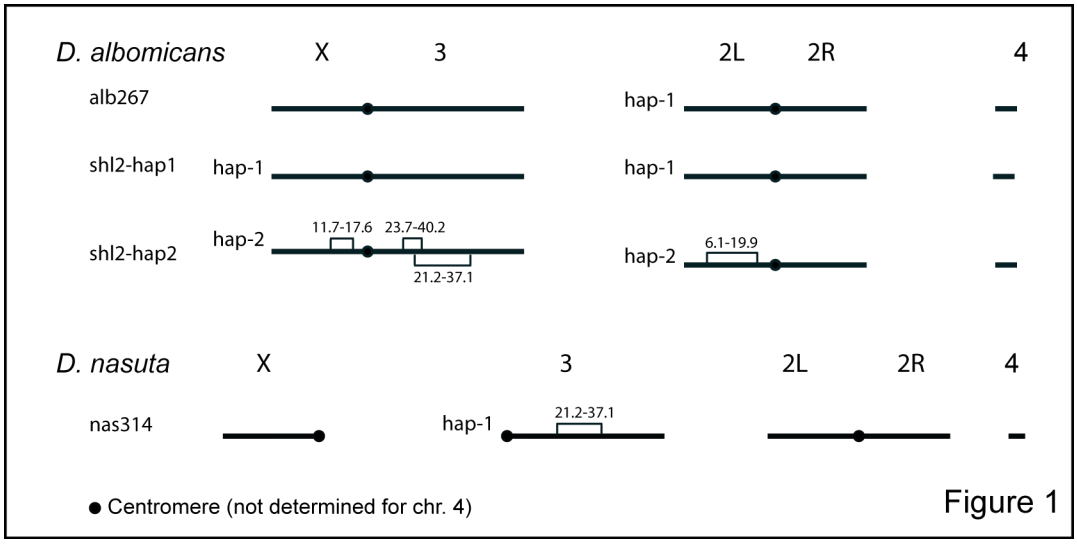

Figure 1. Inbred lines generated by 15 generations of sibling pairings. Inversions are marked with break points. Alb267 and nas314 are free of inversions, while shl2 contains inversion polymorphisms. Compared to alb267, shl2-haplotype 1 has the same sequence, while on the 3rd chromosome of nas314, there is an inversion (Zhang *et al*. manuscript submitted).

Fertility and Sex Ratio of Hybrid Males and Females

By an exhaustive-mating scheme that allows all sperms (eggs) to fertilize with eggs (sperm) from the tester females (males), we were able to quantify fertility and sex ratio for the parent by counting all the offspring, males and females separately, with an unprecedented precision. Some inter-specific F1 hybrid males were severely depressed for fertility (#total offspring) while strongly elevated for sex ratio (female%), such as the F1 males from alb2♀ x nas3♂ and shl2♀ x nas3♂ (Figure 2). But the others expressed close to 0.5 sex ratio and slightly depressed fertility (nas3♀ x shl2♂) or even higher fertility (nas3♀ x alb2♂) as compared to the intra-specific control crosses (alb2 x shl2, both directions). We can interpret the observed fertilities in Figure 1 by three antagonistic causes: inbreeding depression, heterosis and outbreeding depression (aka HI). Both males and females from the three inbred lines (alb267, shl2 and nas314) had reduced fertilities, suggesting strong inbreeding depression in these inbred lines. In contrast, F1 hybrids from alb2 x shl2 in both reciprocal crosses had much higher fertility, suggesting heterosis. In the inter-specific hybrids, we expect to see reduced fertility if there is HI – obviously it is true for two F1 males (from alb2♀ x nas3♂ and shl2♀ x nas3♂), suggesting slight effects of HI (HMS). On the other hand, F1 males from their reciprocal crosses (nas3 ♀ x alb2 ♂ and nas3 ♀ x alb2 ♂) expressed almost normal fertility, suggesting no significant amount of HMS in these two crosses. This asymmetry of F1 fertility further indicates that only a small amount of HMS genes are accumulated in these two species (Turelli and Moyle 2007). This observation is also consistent with the young age of this pair of species.

Overall, the F1 hybrid females from either intra-specific or inter-specific crosses expressed similar fertilities and sex ratios unlike the situations in F1 hybrid males. The discrepancy between male and female follows Haldane's rule that heterogametic sex (F1 males in *Drosophila*) is always less fertile or viable than homogametic sex (F1 female in

Drosophila) (Figure 2).

In sum, the above experiments showed that there are genetic divergences between *D. albomicans* and *D. nasuta* among genes controlling male fertility and sex ratio. These divergences seemed to be larger between alb2 and nas3 than between shl2 and nas3. However, to uncover the causal relationship between SRD and HMS, we still need to dissect the genetic architecture of both phenotypes. For this purpose, a set of three QTL mapping experiments were executed between appropriate lines, as summarized below.

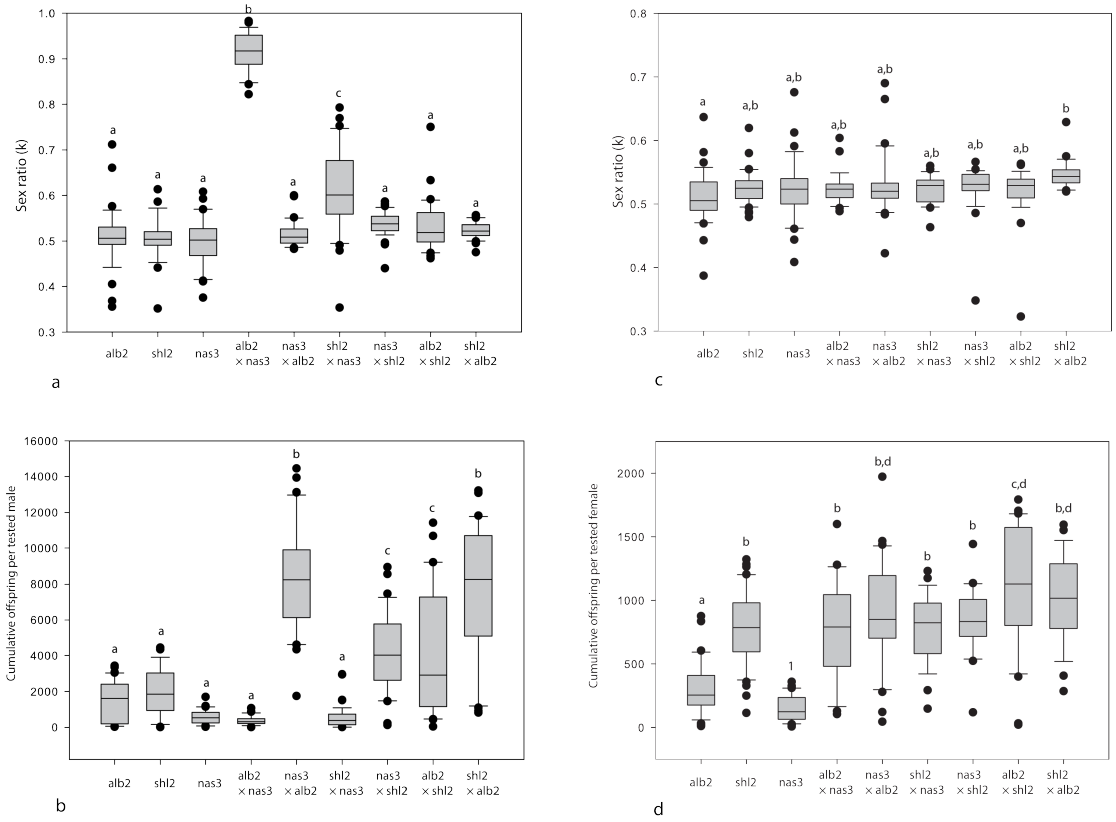

Figure 2. Reduced male fertility and distorted sex ratio found in some interspecific hybrid males. We obtained F1 (male or female) from crosses indicated on the abscissa and tested the phenotypes (fertility and sex ratio) using an exhaustive mating protocol. **(a-b)** Phenotypes of F1 hybrid males. **(c-d)** Phenotypes of F1 hybrid females. Genotypes in each panel are grouped according to significant level of divergence (same letter indicates same group, ANOVA followed by TukeyHSD, $P < 0.05$). Overall, strong SRD and HMS are expressed in two interspecific hybrid males from alb2♀ x nas3♂ and shl2♀ x nas3♂. See text for more comments on inbreeding depression, heterosis and HI (Zhang *et al*. manuscript submitted).

Quantitative Trait Loci (QTL) Mapping of SRD and HMS

Three QTL mapping experiments were executed in order to determine the location and effect of genes contributing to SRD and HMS observed above between alb2 (or shl2) and nas3 (Figure 3). The first QTL experiment (Exp1) was based on the cross between alb267 and shl2-hap1, the second (Exp2) was between alb267 and nas314 and the third (Exp3) was between shl2-hap1 and nas314. The inter-specific divergences were obtained from Exp2 and Exp3, but the whole third chromosome could not recombine in these two crosses due to inversions (D6 region). In other words, we could not separate factors around this region on the map. To gain better resolution on the third chromosome, we crossed alb267♀ to shl2♂ and selected alb267/shl2-hap1♀ to generate recombinants between these two sets of homosequential chromosomes by crossing the F1 ♀ to nas314♂. By this method we were able to penetrate the third chromosome, as well as map the intra-specific variations between alb267 and shl2-hap1 in terms of HMS and sex ratio genes.

From Exp1, we mapped five QTLs for SRD, including four X-3-linked distorters (D1 to D4) and one major suppressor on the 2nd chromosome (S1). Another non-significant suppressor (S2) was also implicated. For fertility, we mapped two loci (HMS1 and HMS2), of which HMS1 overlapped with D1 locus with certainty but the other suppressor (HMS2) was located near but not overlapping with S1 locus.

From Exp 2, we mapped four QTLs for SRD — two on the X-3 chromosome (D5 and D6) and two (S3 and S4) on the 2nd chromosome, and six QTLs for fertility as well —HMS3, 4 and 5 on the X-3 chromosome and HMS6, 7 and 8 on the 2nd chromosome. However the inversion on the 3rd chromosome prevented meiotic cross-over from occurring on this chromosome, thus D6 might cover multiple factors for SRD.

From Exp 3, three loci (D7, D8 and S5) for SRD and 6 loci (HMS9-14) for fertility were mapped between shl2-hap1 and nas314, but all effects were weak.

According to the “drive theory”, the QTL for HMS should also have the phenotype of SRD. The three QTL mapping results roughly followed the prediction but some discrepancies were also obvious. For example, we failed to map the corresponding loci for fertility around the D2-4 regions in Exp1. The discrepancies might disprove our hypothesis, or it might be caused by the inaccuracy of the QTL mapping due to its reliance on statistical analysis. For example, there was a large region around the centromere of the 2nd chromosome severely suppressed for cross-over, which put the accuracy of the mapping in question. In addition, SRD is a trait only measurable in fertile males. Therefore, QTL for SRD might be undetectable if they also cause strong HMS.

All in all, to increase the power and resolution of the genetic mappings on SRD and HMS was of necessity for our future study. Therefore in my thesis research, I focused on this goal and used an introgression approach to obtain finer mappings for both SRD and HMS genes from D1 to D4 and S1 regions.

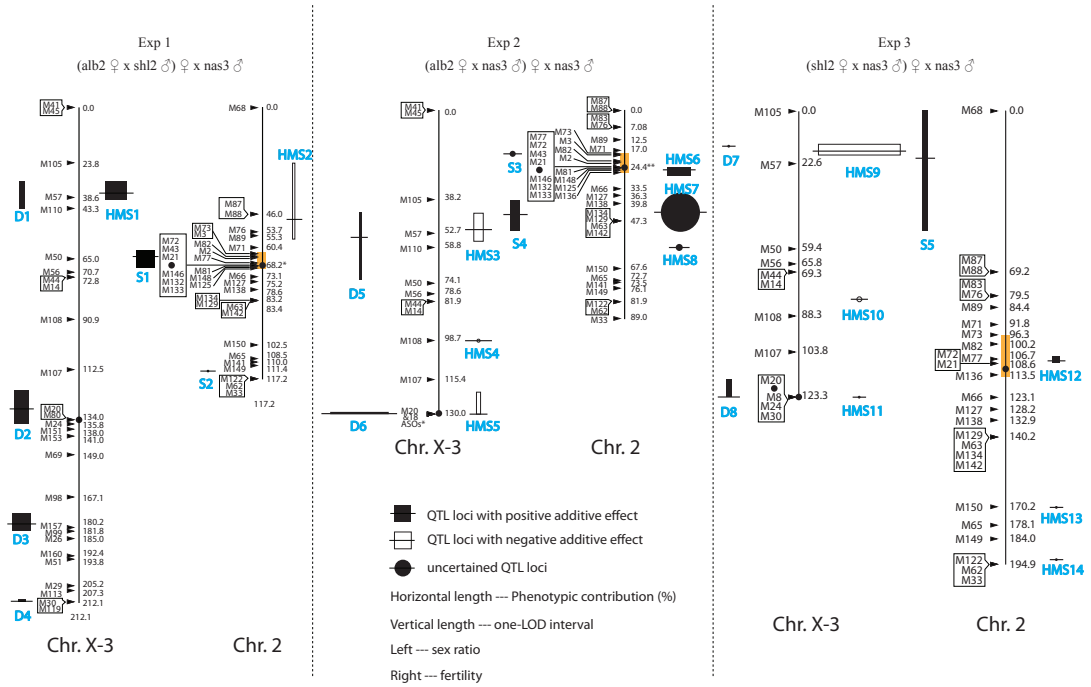

Figure 3. QTL mapping results. ASO markers were used for constructing linkage groups. For each experiment, boxes/circles on the left side represent factors for SRD, while those on the right side represent factors for fertility. The width of each box/circle exhibits how much contribution it gives. Boxes represent confident mapping while circles show uncertain ones.

Exp1: Mapping between alb267 and shl2-hap1. Five significant QTLs were mapped (D1-D4, S1) for SRD, with one additionally weak one (S2), while two were mapped for fertility (HMS1, HMS2).

Exp2: Mapping between alb267 and nas314. Four QTLs (D5, D6, S3, S4) were mapped for SRD, and six (HMS3, 4, 5 and HMS6, 7, 8) for fertility. HMS7 was mapped without certainty. On this map, recombination on the whole 3rd chromosome was suppressed, so were one region on the X and four others regions on the 2nd chromosome.

Exp3: Mapping between shl2 and nas314. Three loci were mapped for SRD and six for HMS, which was similar as the result of Exp2 but with weaker effects.

(Zhang *et al*. manuscript submitted).

Material and Methods

Drosophila Strains and Husbandry

All the *Drosophila* strains were reared in vials with standard cornmeal-molasses food. Laboratory temperature was set at 23°C±1°C. To test hybrid male fertility and sex ratio, each single hybrid male was mated to three nas314 virgin females in its individual vial (aged five days before cross) for seven days before being collected. Tested males were frozen (-20°C) for further genotyping and females were discarded. To generate new recombinants, each single hybrid female was crossed to three nas314 males for seven days before parents were cleared. The hybrid female parents were saved and stored at -20°C for genotyping if that was needed. Dental plugs were inserted in food to enhance offspring development. All offspring were counted (males and females separately) every other day until the 20th day after setup.

Introgression

A marker-assisted approach to introgress chromosomal segments spanning each of the five QTL (D1-D4, and S1) regions from alb267 to nas314 background was used. Two (or more than two for S1 locus) specific RFLP markers closely linked to each QTL were followed to track the introgression (Table 1). Thirty-two possible genotypes were expected if all random combinations of the five QTLs (D1-S1) were possible, but in reality less combinations were obtained because some of the possible recombination were less likely to occur.

Pure alb267♀ was crossed to pure nas314♂ at the first generation. Around 20 hybrid females were collected to set up the next generation (second) when each individual hybrid female was crossed to three nas314 males. For the third generation, hybrid males were collected and each was crossed to three nas314 virgins in separate vials for seven days before

the single male parent was collected and frozen for genotyping. According to the genotype, vials with desired combination of markers were kept and the offspring were sexed and counted until the 20th day since setup. Female virgins were also collected from these strains to set up the 4th generation. This general procedure was followed by all latter generations, with hybrid males and hybrid females collected alternately in consecutive generations (Figure 4). For example, if the hybrid had alb267 genes (dark in the graph) spanning the whole D1-D4 and S1 region (e.g., the F2 ♂), this combination was named D1D2D3D4; S1. If cross-over occurred as such to recombine out the D1 gene (the markers for D1 were all of nas314 now), then it was named D2D3D4; S1 (e.g., F4 ♂).

Genotyping

DNA from the collected samples was extracted following the standard protocol for Quick and Dirty DNA Extraction. DNA samples were frozen at -20°C if not being immediately processed. Standard protocol for PCR reactions was followed (see Appendix I for details). Different primers for each marker were shown in Table 1. PCR products were cut by appropriate restriction enzymes (see appendix II for restriction enzyme reaction protocol). The RFLP patterns were then detected under UV after running agrose gels for ~30 minutes/105V.

Table 1. Markers used for introgressing alb267 into nas314

locus	Marker ID	primer	enzyme	RFLP
D1	PKM57	F: AGCATTGCAGTGTGCAATC R: GACCACATCGTTGTTGATGC	HaeIII	alb267: 172+575 nas314: 747
	PKM105	F: AAAATGCATCGACATGCTCA R:AACAAGTATGGGGAAAGCAA	AflII	alb: 311+377 nas:691
D2	PKM20	F: CTGGCCGGAGTACAGAAGAA R: GCTGATCGTTGAAAAGCACA	AatII	alb267: 421+364 Nas314: 789
	PKM107	F: ATTGGCGAGCTGGAAATATG R: ACATTGAACACATCCGACGA	BsiEI	alb267: 116+393 nas314: 509
D3	PKM99	F: CCCTATGCCGAAATAGCTGA R: ACTCCTGCATGGGATTGAAG	HaeIII	alb267: 103+483 nas314: 586
	PKM98	F:GACGAAGAGGCATCAGCTCT R:GGCAGGGCATAACCTTCATA	HinPI	alb267:73+3+89+343 nas314:73+3+89 +110+233
D4	PKM30	F: GTGTGAGCGCTCTCTCCTCT R: GGAGATGCCAGTGTTGGACT	HindIII	alb267: 88+598 nas314: 687
S1	PKM3	F:GCTTCTCTTTGGGCAGATGA R:TCGAGACGCTACAAAAAGCA	AflII	alb267:516 nas314: 444+72
	PKM77	F: ATACTGGGCAAGGTGATTGG R:CGATTAGTTTGATATTCCAGTTCTCA	XbaI	alb267: 333+87 nas314: 414
	PKM72	F: GTGGCATTGGTCTGGACAC R: GGGTTGATGGAGTAGGCAGT	AvaII	alb267: 139+447 nas314: 620
	PKM66	F:GTAAGTCGCTCCGCAAGAAG R:TTGCCTTCCAGAATGTAGCC	Xho I	alb267:643 nas314:354+289
	PKM127	F:GCGTTTGCCTCAATATTCCT R:ATCTTCCGGGGCGATAATAC	HindIII	alb267:637 nas314: 155+482
	PKM63	F: GAGAGCAGCCAAATCCTCAC R: GATTGGCCCAGATTGTGTCT	HinP1	alb267: 588 nas314: 268+322
	PKM150	F:TTTCGCTATCAAACCGATCC R:GGCACACACACTTCCTCAAA	BsmAI	alb267:131+378 nas314:509

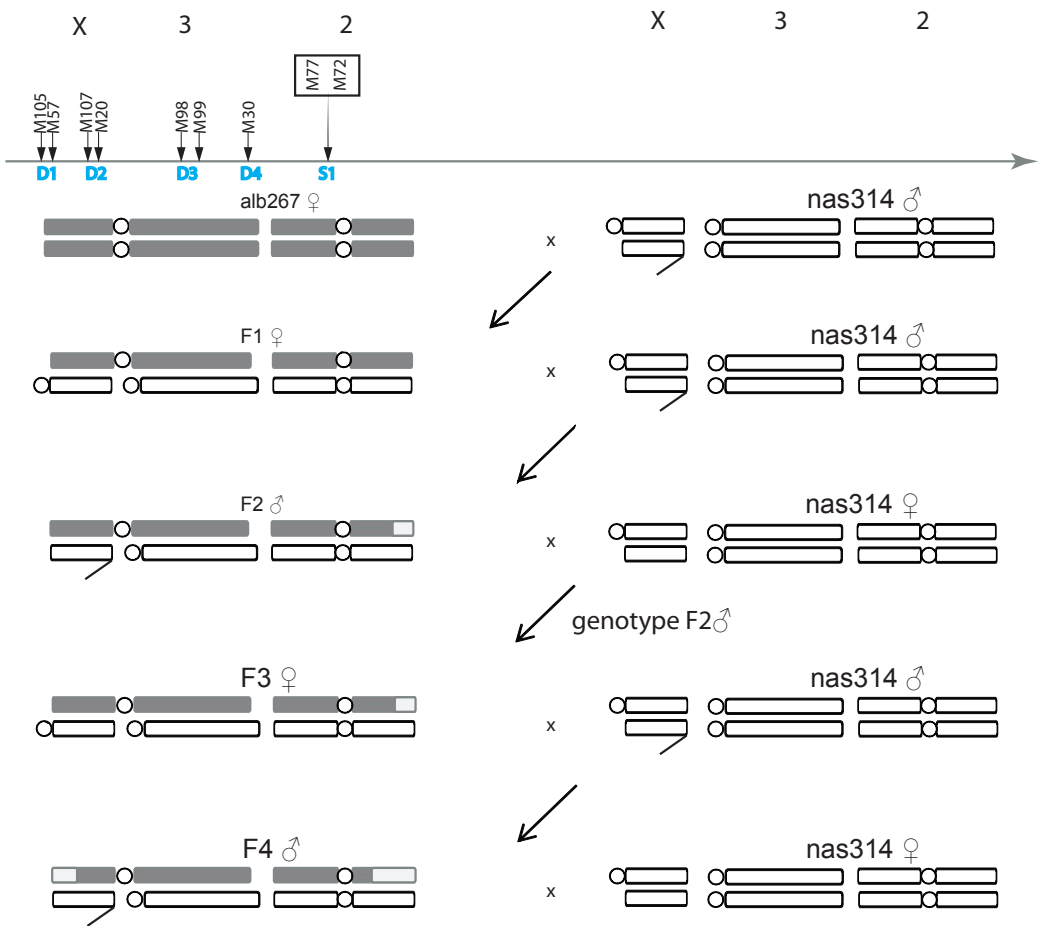

Figure 4. Introgression scheme. Each locus was tested by two flanking markers (marked as in the figure) except D4 that is located at the tip of chr.3. If two flanking markers for one locus show the same result as that of alb267 alleles, this QTL in between is assumed to be present.

Fine Mapping of the S1-S2 Region

Shorter introgressions of alb267 into nas314 around the S1 region were constructed by screening hundreds of single males from their hybrid mothers with more markers (shown on the map in Figure 7B). Combinations shown in Figure 7C were selected. These combinations were then crossed to D2D3D4 (strong distorters). And after being screened, males with D2D3D4; S1* (S1*: shorter introgressed combinations of S1) were then tested for phenotypes (male fertility and sex ratio) using the method described above.

Statistical Analysis

All the statistical analyses for fertility and sex ratio were performed in R. Fertility of two samples were compared with the Wilcoxon test. For more than two samples, fertility by rankings was analyzed with standard ANOVA followed by *post hoc* TukeyHSD. Generalized linear model (GLM) was used to compare sex ratios from original counts of males and females with binomial distribution and logit link function. *Post hoc* analysis was made possible by using the R package “multcomp”.

Results

Introgression

The functions of the constituent elements in a meiotic drive system must be tested in appropriate genetic contexts. For this reason, we introgressed DNA fragments from *D. albomicans* into *D. nasuta* background (Figure 4) with combinations of the five QTLs (D1-D4 and S1 in Figure 3-Exp1). Males of these introgression genotypes were tested for fertility (f: number of offspring per male parent) and sex ratio (k: percentage of females) in their offspring. The function of individual locus can be inferred from contrasting two genotypes with only one locus different (Figure 5 and 6).

The function of D1 can be obtained by contrasting D1D2D3D4; S1 (f = 55.7 ± 8.0; k = 0.87 ± 0.02) to D2D3D4; S1 (f = 97.6 ± 9.2; k = 0.72 ± 0.03). Under the genetic background of D2D3D4; S1, the presence of D1 decreases fertility significantly ($p = 0.002$, Wilcoxon test) while increases sex ratio ($p < 2 \times 10^{-16}$, logistic regression). Based on the same logic, the function of D2 can be illustrated by contrasting the fertility of D2D3D4 (f = 2.8 ± 0.5) to that of D3D4 (f = 51.9 ± 17.7) ($p = 1\times 10^{-6}$), and comparing their sex ratios (D2D3D4: k = 0.87 ± 0.02; D3D4: k = 0.70 ± 0.05; p = 0.02). Under the background of D3D4, D2 works to decrease fertility while increase sex ratio significantly. For D3, the contrast relied on is between D2D3 (f = 8.1 ± 1.4, k = 0.77 ± 0.02) and D2 (f = 68.5 ± 6.4, k = 0.51 ± 0.01). The function of D3 also can be concluded as to decrease fertility but increase sex ratio and this result is supported by significant p values (f: $p=5.8 \times 10^{-14}$; k: $p = 2 \times 10^{-16}$). For D4, similar result is obtained from the contrast between D2D3 and D2D3D4. With p = 0.046 for fertility as well as p = 0.001 for sex ratio, D4 significantly decreases fertility while increases sex ratio as expected under the background of D2D3.

Finally, the function of S1 as a suppressor is also determined. Based on the

comparison between D2D3D4; S1 and D2D3D4, suppressor S1 plays its role in increasing fertility while decreasing sex ratio with the presence of D2D3D4 (f: $p = 1.4 \times 10^{-11}$; k: $p = 2.9 \times 10^{-6}$). Moreover, the contribution of S1 to both fertility and sex ratio is opposite to those of D1-D4.

Thus, each of the five major QTL regions has been shown to have dual functions affecting male fertility and sex ratio, and their effects are consistent with the theoretical expectation of the "drive theory" as elaborated in Introduction.

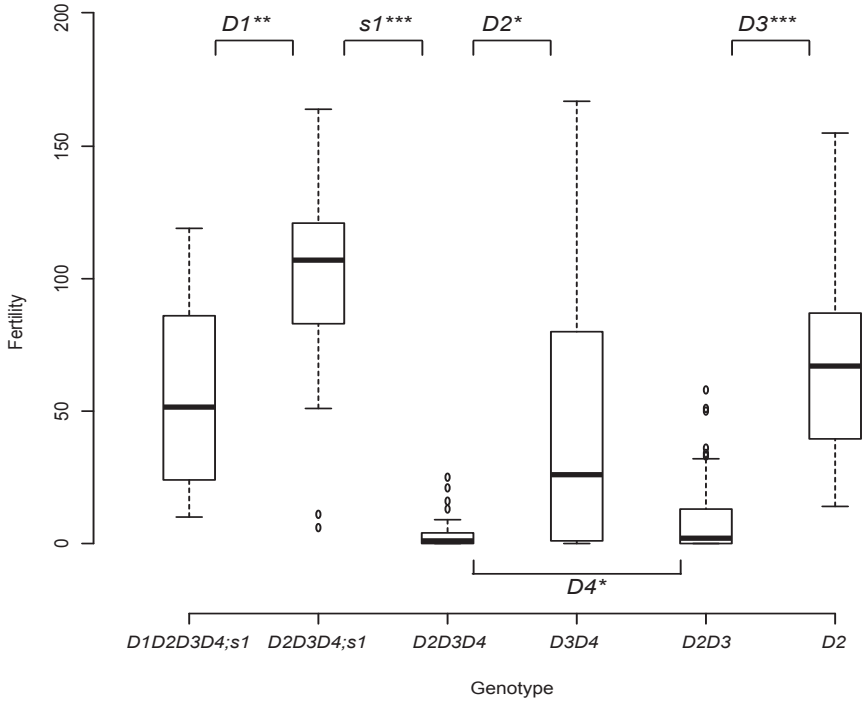

Figure 5. Boxplots of fertilities (offspring per male) of various introgression males. Asterisks represent statistical significance (* P< 0.05; ** P < 0.01; *** P< 0.001). By comparing two genotypes with only one locus different, we were able to determine the function of that locus.

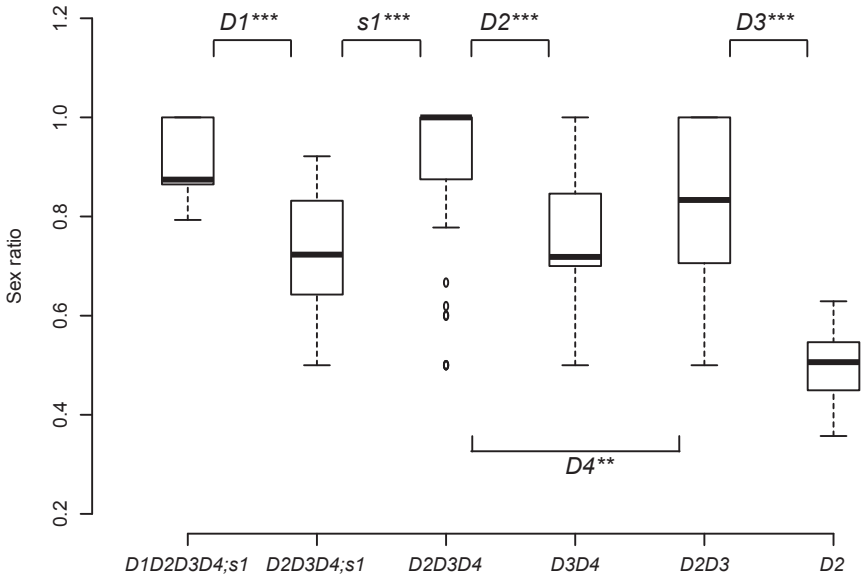

Figure 6. Boxplots of sex ratios (proportion of females) of various introgression males. Asterisks represent statistical significances as in Figure 5.

Fine Mapping

We noticed that the HMS6-8 region (spanning the centromere of the 2nd chromosome) was quite ambiguous in terms of the number, location and effects of the QTLs (Figure 3-Exp2). This uncertainty might be attributed to the low recombination rate nearby, and the complexity of genetic architecture in this region as well. To resolve this problem, we screened the whole 2nd chromosome with more markers (PKM83-150) looking for shorter introgressions and tested their fertility and sex ratio (Figure 7).

These shorter introgressions are classified into 3 distinct groups based on both fertility (ANOVA on fertility rankings: $p < 2 \times 10^{-16}$ for the whole 9 lines, combined with *post hoc* analysis with Tukey HSD) and sex ratio (Logistic regression, $p << 0.001$) (Figure 7c). Group 1 consists of the lines 191 and 44 with the highest fertility (overall $f = 155.2 \pm 4.7$) but lowest sex ratio ($k = 0.627 \pm 0.010$). Group 2 consists of the lines 64, 57, 209, 12, 91 and 26 with the intermediate phenotypes of $f = 34.9 \pm 2.1$ and $k = 0.819 \pm 0.008$. Group 3 only contains line 71 and its fertility ($f = 8.4 \pm 1.8$) is the lowest but still higher than that of the genotype D2D3D4 ($f = 2.4 \pm 1.5$, $p < 0.01$). However, the sex ratio of line 71 ($k = 0.889 \pm 0.045$) does not differ from that of the genotype D2D3D4 ($k = 0.889 \pm 0.045$, $p = 0.112$). This deviation from expectation can be explained by the very weak or null SRD suppression effect of this locus, or the very small progeny sizes in both lines compared (fertilities of D2D3D4 and line 71 are both near zero and the small sizes of progeny provide very little evidence for sex ratio difference). A possible solution to this problem is to test these shorter introgressions with a weak distorter (for example, D2 only) to allow more offspring to be reproduced.

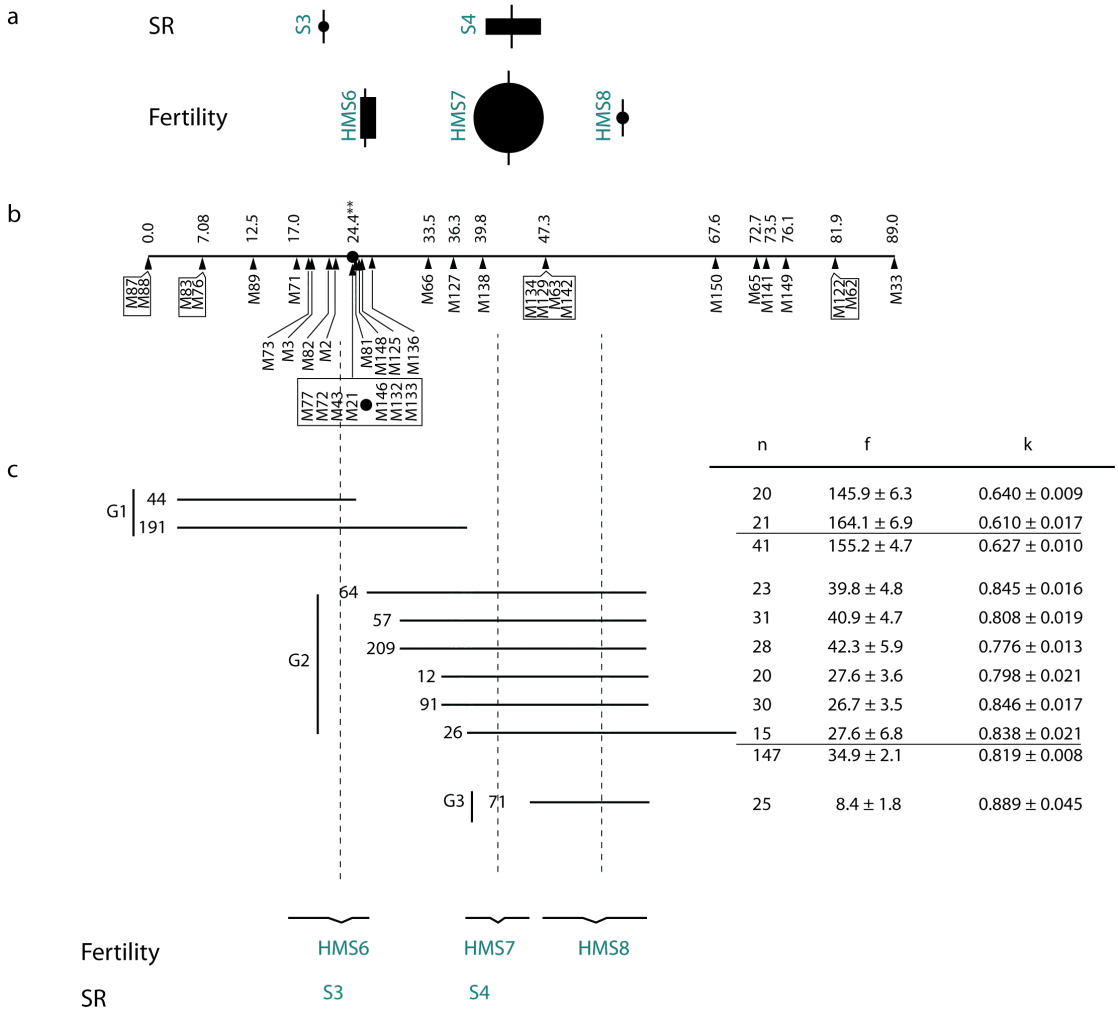

Figure 7. Fine mapping of the M83-M150 region. (a) The QTLs implicated in this region from Exp2 of QTL mappings (Figure 3): HMS6-8 as HMS factors and S3-4 as SRD factors. (b) Genetic map of the 2nd chromosome. (c). Results of fine mapping. The dashed lines and the curly brackets at the bottom points to the most likely positions of QTL and their ranges based on fine mapping data. The three loci implicated in fine mapping are remarkably consistent to those obtained from QTL mapping.

Discussion

Speciation is the origin of biodiversity including our human species. Its relevance to answer the ultimate humanity questions "Where Do We Come From? What Are We? Where Are We Going?" can never be more fundamental. Until today, our knowledge of the molecular and evolutionary mechanisms of speciation is still limited. Even though the "origin of species" has been a major research focus for evolutionary biologists ever since Darwin posed the question more than 150 years ago, great advances were made possible only with the assistance of genetic analysis methodologies, particularly in the last three decades. Many of the recent researches were motivated to address the cause(s) for Haldane's rule (1922), asking "Why do males evolve so fast?" (Wu and Davies 1993). Several major theories appear to be plausible, among which the "drive theory" seems to provide the best accounts for empirical observations (Tao and Hartl 2003). Importantly, the "drive theory" posits a type of non-adaptive evolution—the intragenomic conflicts, particularly in the form of sex ratio meiotic drives— to be responsible for the evolutionary change underlying the fast evolution of male fertility genes. Researchers have accumulated indirect supports for the "drive theory" in the last decade (Cocquet *et al*. 2012; Phadnis *et al*. 2009; Tao *et al*. 2001), but more direct evidence to support it is still in demand. Therefore, I aimed in my thesis research to gather more concrete evidence to test the "drive theory".

The two major results from the introgression experiments and the fine mapping experiment strongly support the "drive theory" because each of the five genes of HMS also expresses SRD. The dual functions of these genes on SRD and HMS are functional in "opposite" directions: if one distorter increases sex ratio then it must also decrease male fertility; if the suppressor decreases sex ratio then it must also increase male fertility, even if the suppressor is finely split to three smaller regions (One of the three does not follow the

pattern but we believe this is an issue of small sample size and lack of detection power) (Figures 5-7). Moreover, the dual functions can be easily explained by the physiological mechanism of meiotic drive. In the few meiotic drive systems well studied so far, visible abnormalities at post-meiotic development stages contribute to selective weeding of sperm that presumably carry the Y-chromosome with sensitive responder(s) (Tao *et al.* 2007), although the molecular etiology might be attributed to meiotic stage (Bauer *et al.* 2007). In addition to reducing the amount of functional sperms, it is also likely that the SRD genes have pleiotropic effects of reducing fertility in terms of the total amount of gamete production and sperm competence by inducing pathological reactions to occur during spermatogenesis anyway. If the driving force for evolving HMS is indeed SRD, we expect to see SRD and HMS go hand in hand in hybrids – an expectation satisfied completely by our experimental evidence. Moreover, all distorters are X-linked and suppressors are on autosomes in our mapping, a result once again fully consistent with the "drive theory".

One more important feature of our study is that we uncovered an on-going evolution of SRD system within *D. albomicans*, and this on-going evolution is generating HMS effects as we compared alb2 to shl2 in terms of their SRD/HMS polymorphisms (Figure 2). This is a case of speciation in action and provides a great opportunity to study "speciation genes". Historically, the term "speciation genes" are referred to any genes underlying the divergence between two species, but many of the divergent genes underlying reproductive isolation (RI) are not necessarily playing any role in the initial development of RI and might have evolved only after speciation has been well established. So these genes are better called "divergence genes". Currently in the field of speciation genetics, a major challenge is to close the gap between "divergence genes" and "speciation genes". More specifically, solid evidences must be sought to support the causal role of any "divergence gene" in the process of speciation. For testing the "drive theory", though several previous studies have attributed sex ratio

meiotic drive to be the cause of hybrid male sterility, this study is the first one to catch them in action.

As QTL mapping relies on statistical analysis and the accuracy is in doubt, one mapped QTL locus might contain multiple factors with weak effects or it might truly contain one factor. Therefore, the result of QTL mapping requires other evidences to support, especially when the P value is not very small. In order to solve this issue, we have further generated fine mapping for the S1 region, which has been mapped to consist of at least three factors for fertility ambiguously in the QTL mapping (Figure 7A). The significant difference in terms of fertility and sex ratio between the three groups indicates that the whole S1 region indeed consists of at least 3 discrete factors for fertility as well as two or three factors for sex ratio (Figure 7C). This result corresponds to what QTL mapping suggests and enhances the reliability of the QTL mappings in our study. However, there is one deviation at HMS8 region due to its lack of contribution to sex ratio. We come up with an explanation that the failure to show its complete function is due to the small progeny size. As only a few offspring are produced from the males in question, sex ratio distortion is hard to observe. Thus this deviation is not sufficient to reject our hypothesis. Importantly, even with such a higher magnification, each locus still expresses its dual function on both sex-ratio distortion and fertility determination (barring the deviation at HMS8), which further provides strong support for the “drive theory”. Even if each QTL contains several minor genes for HMS/sex ratio, we can still anticipate that the genetic architecture is universal and probably can be applied to all the cases in *Drosophila*.

Even though the fine mapping is generated with high magnification, it still does not have enough power to tell (1) whether the pair of S and HMS loci (S3/HMS6, S4/HMS7) are the same or not; and (2) whether HMS8 or the region surrounding this locus has suppressing effects on sex ratio. If we take into account the perfect consistence between factors for sex

ratio and fertility as shown in Figure 5 and 6, we can at least interpret the data in Figure 7 as that sex ratio and fertility are controlled simultaneously.

The result of our study also provides a plausible solution to one long-standing controversy with regards to the role of chromosomal rearrangement in speciation. Chromosomal rearrangements (fusion, fission, translocations, pericentric and paracentric inversions) are often observed in hybrids reproduced by two closely related species, and thus are believed by some to have played a major role in establishing RI due to the concurrence of a reduced fitness in F1 heterozygotes (White and Kankel 1978; King 1993). However, this view is easily attacked by the doubt that if heterozygotes are less fit, the new karyotype is less likely to spread in population and get fixed to form new species. This fundamental flaw to the theory proposed by White and King makes many other evolutionary biologists skeptical of their theory. Based on our study regarding to the genetic architecture of HMS between *D. albomicans* and *D. nasuta,* we have shown that the present RI is significantly affected by certain genic factors, instead of only chromosomal rearrangements. Although we do not have data, we may confidently predict that the fitness of heterozygotes from two parents with different karyotypes is probably not significantly low, but the current heterozygotes have much reduced fitness because of genic HI factors. Therefore, the argument supporting the chromosomal speciation theory is not valid. Nevertheless, we do not have direct evidence to disprove that chromosomal rearrangements have played major roles in speciation, either.

Future Directions

This study opens up possibilities of several diverse directions for future research to focus on. First, the intra-specific effects of distorters and suppressors of the meiotic drive system can be determined using a similar approach in this study. We expect to see that the

meiotic drive system carrying its function within species can also cause some levels of male sterility among strains with different geographic origins. We can achieve this goal by introgressing alb2 into shl2 background or vice versa, to show that this system is still active within *D. albomicans*.

Second, given the universality of the genetic architecture underlying HMS/sex ratio between *D. albomicans* and *D. nasuta*, it is possible to even more accurately map and position clone some of the underlying genes. The study of their functions will lead to profound understanding of their molecular etiology. For example, we may follow their expression pathways, study their protein products, and determine the molecular mechanisms for sex ratio control as well as fertility.

More generally, it is also possible to connect the "drive theory" to population genetics to infer the allele dynamics and the interaction between intragenomic conflicts and selection. The population studies can be extended to genomics study as well to scan the whole genome for a broader view of the allele dynamics accompanying speciation.

If these above potential studies are accomplished, our understanding of speciation in genetics will move forward with a giant leap.

Appendix I

Protocol for PCR reaction (Adapted from "*Molecular Cloning*")

For one reaction (12.5λ), mix as the following suggests:

10×buffer	1.25λ
25mM Mg^{2+}	0.2λ
10mM dNTP	0.25λ
primer F/R (10μM for each primer)	0.4λ
Taq DNA polymerase	0.1λ
Template	1.0λ
ddH_2O	9.3λ
Total	12.5λ

Appendix II

Protocol for enzyme reaction

For one reaction (10λ), mix as the following suggests:

10×CutSmart buffer	1λ
PCR product	4λ
Enzyme (20U/λ)	0.5λ
ddH_2O	4.5λ
Total	10λ

*For enzyme HindIII, replace CutSmart buffer with 2×buffer

*For enzyme with 10U/λ concentration, change its volume to 0.75λ and decrease ddH_2O volume to 4.25λ. Other volumes stay the same.

References

Bauer, H., Véron, N., Willert, J., Herrmann, B.G. (2007). "The t-complex-encoded guanine nucleotide exchange factor Fgd2 reveals that two opposing signaling pathways promote transmission ratio distortion in the mouse". *Genes Dev* 21(2): 143–147.

Campbell, P., Good, J.M., Nachman, M.W. (2013). "Meiotic Sex Chromosome Inactivation Is Disrupted in Sterile Hybrid Male House Mice". *Genetics* 193: 819-828.

Charlesworth, B., Coyne, J. A., Barton, N. H. (1987). "The relative rates of evolution of sex chromosomes and autosomes". *The American Naturalist* 130 (1): 113–146.

Cocquet, J. *et al.* (2012). "A genetic basis for a postmeiotic X versus Y chromosome intragenomic conflict in the mouse". *PLoS Genet.* **8**, e1002900.

Düsing, K. (1883). "Die Factoren, welche die *Sexualitat entscheiden*". *Jenaische Zeitschrift für Naturwissenschaft,* 16, 428-464.

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

Fisher, R.A. (1930). *The Genetical Theory of Natural Selectio*n. Oxford: Clarendon Press.

Haldane, J. B. S. (1922). "Sex ratio and unisexual sterility in hybrid animals". *J. Genet.* 12: 101–109.

Hamilton, W.D. (1967). "Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology". *Science* 156(3774):477-88.

Hiraizumi, Y., Crow, J.F. (1957). "The amount of dominance of recessive lethals from natural populations of *D. melanogaster*". *Drosophila Inform. Serv*. 31 : 123.

Hurst, L.D., Pomiankowski, A. (1991). "Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena". *Genetics* 128(4): 841-858.

Hurst, G.D.D., Werren, J.H. (2001). "The role of selfish genetic elements in eukaryotic evolution". *Nature Reviews Genetics* 2, 597-606.

Jaenike, J. (2001). "Sex chromosome meiotic drive". *Annual Review of Ecology and Systematics* 32: 25-49.

King, M. (1993). *Species evolution*. Cambridge Univ. Press, Cambridge, U.K.

Kitagawa, O.Y-I. (1990). "Incipient reproductive isolation between *Drosophila masuta* and *Drosophila albomicans"*. *Genet Sel Evol* 22: 31-46.

Kitagawa, O. Y-I., Wakahama, Y., Fuyama, Y., Shimada, E., Takanashi *et al*. (1982). "Genetic studies of the Drosophila nasuta subgroup, with ntoes on distribution and morphology". *Jpn. J. Genet.* 57: 113-141.

Kusano, A., Staber, C., Chan, H.Y.E., Ganetzky, B. (2003). "Closing the (Ran) GAP on segregation distortion in *Drosophila"*. *BioEssays* 25: 108-115.

Lifschytz, E., Lindsley, D.L. (1972). "The role of X-chromosome inactivation during spermatogenesis". *Proc. Natl. Acad. Sci.U.S.A* 69, 182-186.

Lin, F.J., Tseng, H.C., Wang, T.C. (1974). "Standard map of the salivary gland chromosomes of *Drosophila albomicans Duda"*. *Dros. Inf. Serv*. 51: 42-43.

Lyon, M.F. (2003). "Transmission ratio distortion in mice". *Annu Rev. Genet* 37:393-408.

Masly, J. P., Presgraves, D. C. (2007). "High-resolution genome-wide screen for hybrid incompatibilities validates the large-X-effect in *Drosophila"*. *PloS Biology* 5, e243.

Meiklejohn, C.D., Tao, Y. (2010). "Genetic conflict and sex chromosome evolution". *Trends Ecol Evol* 25(4): 215-223.

Phadnis, N., Orr, H. A. (2009). "A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids". *Science* **323**, 376-379.

Ranganath, H.A., Ramachandra, N.B. (1994). "Pattern of sexual isolation between parental races (Drosophila nasuta and D. albomicans) and the newly evolved races (Cytorace I and II)". *Indian Journal of Experimental Biology* 32, 98-102.

Sandler, L., Hiraizumi, Y., Sandler, Y. (1959). "Meotic drive in natural populations of Drosophila melanogaster. I. The cytogenetic basis of segregation-distortion". *Proc Natl Acad Sci U S A.* 45(9): 1412–1422.

Sandler, L., Novitsky, E. (1957). "Meiotic drive as an evolutionary force". *Am. Nat*. 91, 105-110.

St. Pierre, S.E., Ponting, L., Stefancsik, R., McQuilton, P., and the FlyBase Consortium (2014). "FlyBase 102 - advanced approaches to interrogating FlyBase". *Nucleic Acids Res.* 42(D1):D780-D788.

Tao, Y., Chen, S., Hartl, D. L. & Laurie, C. C. (2003) "Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. I. Differential accumulation of hybrid male sterility effects on the *X* and autosomes". *Genetics* 164, 1383-1397.

Tao, Y., Hartl, D.L. (2003) "Genetic dissection of hybrid incompatibilities between Drosophila simulans and Drosophila mauritiana . III. Heterogeneous accumulation of hybrid incompatibilities, degree of dominance and implications for Haldane's rule". *Evolution* 57: 2580 - 2598.

Tao, Y., Hartl, D. L. & Laurie, C. C. (2001) "Sex-ratio segregation distortion associated with reproductive isolation in *Drosophila". Proc. Natl. Acad. Sci. USA* **98**, 13183-13188.

Tao, Y., Masly, J. P., Araripe, L., Ke, Y. & Hartl, D. L. (2007) "A *sex-ratio* system in *Drosophila simulans*. I. An autosomal suppressor". *PloS Biol.* **5**, e292.

Tokuyasu, K.T., Peacock, W.J., Hardy, R.W. (1977). "Dynamics of spermiogenesis in *Drosophila melanogaster* VII. Effects of segregation distorter chromosome" . *Journal of ultrastructure research* 58: 96-107.

Turelli, M., Moyle, L.C. (2007). "Asymmetric postmating isolation: Darwin's corollary to Haldane's rule". *Genetics* 176: 1059-1088.

Turelli, M. & Orr, H. A. (2000). "Dominance, epistasis and the genetics of postzygotic isolation". *Genetics* 154, 1663-1679.

Turner, J.M. (2007). “Meiotic sex chromosome inactivation”. *Development* 134(10):1823-31.

White, K., Kankel, D. R. (1978). “Patterns of cell division and cell movement in the formation of the imaginal nervous system in *Drosophila melanogaster”*. *Developmental biology*, 65(2), 296-321.

Wu, C.I., 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.

Yang, Y.Y., Lin, F.J., Chang, H.Y. (2004). “Sex ratio distortion in hybrids of *Drosophila albomicans* and *D. nasuta”*. *Zoological Studies* 43(3): 622-628.

Zhang, L., Sun, T., Woldesellassie, F., Xiao, H., Tao, Y. (2014). “*Sex-ratio* meiotic drive causes the evolution of hybrid male sterility”. (submitted manuscript).