10-23-2020 1:30 PM

Genetic basis of hybrid sterility between Drosophila pseudoobscura and D. persimilis

Alannah J.P. Mattice, *The University of Western Ontario*

Supervisor: Moehring, Amanda J., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Biology

Abstract

Speciation is the underlying process that leads to formation of new species, and therefore is the basis of biodiversity. Genes involved in each stage of speciation, such as those involved in interspecies sterility, remain elusive. Male hybrid sterility and postzygotic isolation between *Drosophila pseudoobscura* and *D. persimilis* was examined in this study through backcrossing of female hybrids into each parental line (introgression), selecting for a sterile sperm phenotype, needle-eye sperm. Sperm phenotypes did not separate through backcrossing; instead, males presented with multiple sperm phenotypes. A relationship between the phenotypes observed and the potential genes involved was examined through whole genome sequencing and SNP analysis of the DNA of 20 introgressed male hybrid samples. One finding was SNPs for hybrid sperm sterility were species specific. Also, sperm sterility and heteromorphism appear to be controlled by many loci. Further analysis of SNPs isolated in this study has the strong potential to identify candidates for loci involved in formation of needle-eye sperm, and postzygotic male hybrid sterility in other species.

Keywords

Speciation, hybrid, sterility, spermatogenesis, sperm, postzygotic isolation, next generation sequencing.

Summary for Lay Audience

Speciation is the process of two populations of organisms of the same species evolving over time until they are unable to reproduce with each other. Some species have not completely separated, and are still able to create viable, but oftentimes sterile, hybrid offspring. A common example of hybrid sterility comes from horses and donkeys, who separated approximately 7.7-15 million years ago (Huang *et al.* 2015). When a male donkey and a female horse reproduce, they sire a mule. All male mules are sterile and most female mules are sterile. In rare cases female mules are fertile when mated to a horse or donkey (Savory 1970).

Similar to horses and donkeys, the crossing of two species of fruit flies, *Drosophila pseudoobscura* and *D. persimilis*, produce all sterile male hybrids. However, in the case of these fruit flies, all female hybrids are fertile. These two species of fruit flies also diverged more recently, 0.55 million years ago. These sterile hybrid male fruit flies can still produce sperm, but these sperm are not able to fertilize female eggs to make more hybrids. Fruit flies are used because they are less expensive to maintain, have shorter life cycles, and can be in a tightly controlled environment. My research focused on genetic differences cause the male fruit flies to be sterile. Hybrids receive genetic material (DNA) from both parent species. The DNA of both fly species studied here is split into two pairs of five separate chromosomes, X/Y, 2, 3, 4, and dot. The pairs of each chromosome can interact with each other through proteins. Instead of ten separate assembly lines for proteins, pairs of chromosomes are connected to each other by networks integral to protein production and cell function. In hybrids, the chromosomes are unlikely to all function properly because each species has differentiated chromosomes

that might not be able to form proper pairs. The failure of some of these networks could be the basis of sterility. My study supported the species-specific differences in the pieces of the network contributing to hybrid sterility. This work can be continued to identify specific points in the DNA that lead to hybrid sterility and applied to other species.

Acknowledgments

I would like to thank Dr. Amanda Moehring for her help and guidance during my thesis and during the writing of this thesis. Dr. Katharine Korunes at Duke University performed the assembly of the raw whole genome sequences, variant calling, and principle component analysis. I would like to thank my two thesis advisors; Dr. Kathleen Hill and Dr. Elizabeth MacDougall-Shackleton for their help from my first committee meeting through to the final meeting at the end of my thesis. I would also like to thank Karen Nygard for all her help in using the Zeiss fluorescent microscope and camera software at the Western Ontario University Biotron. Dr. Anne Simon also deserves many thanks for helping me finish this thesis as an interim supervisor during the COVID-19 pandemic.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

BC	Backcross
BC11*per*	Backcross to *D. persimilis* for 11 generations
BC11*pse*	Backcross to *D. pseudoobscura* for 11 generations
BDM	Bateson-Dobzhansky-Muller
DAPI	4’,6-diamidino-2-phenylindole
HMR	Hybrid Male Rescue
IGS	Intergenic spacer regions
LHR	Lethal Hybrid Rescue
NE	Needle-eye
NGS	Next generation sequencing
OsdH	*Odysseus-site Homeobox*
PCA	Principle Component Analysis
PSE	*Drosophila pseudoobscura*
PER	*Drosophila persimilis*
QTL	Quantitative trait loci

Chapter 1

1 Introduction

When two species interbreed, the resulting hybrid offspring are often sterile and therefore unable to reproduce. Identifying the genetic basis of hybrid sterility (is not a new question. Much work has been done to understand how this mixture of DNA in a hybrid can result in hybrids unable to form more hybrids, even though the parent species were able to mate and fertilize an egg with sperm from a separate species (Coyne 1992; reviewed in Presgraves 2010; Turelli *et al.* 2001). Although extensive research has been done to understand the genetic basis of hybrid sterility, genes linked to sterility have only been found for some species (reviewed in Presgraves 2010), and in particular backcross generations, while first-generation sterility and the basis of sterility in other species pairs largely have no answer. Species pairs such as *Drosophila pseudoobscura* and *D. persimilis* are one such pair where much work has been done to understand why only male hybrids are sterile, yet the exact genes involved are still unknown (Dobzhansky 1934; Machado *et al*. 2007; Noor *et al.* 2001). This thesis will be focusing on postzygotic isolation and explanations for heterogametic (two different sex chromosomes, e.g. XY) hybrid sterility. In particular, this thesis examines sperm heteromorphism (more than one sperm morph), its link to spermatogenesis, and how sperm heteromorphism is presented in male hybrids of *Drosophila pseudoobscura* and *D. persimilis*.

1.1 Speciation

The world's biodiversity has been defined using multiple species concepts, with the most widely accepted being the biological species concept. According to this definition, a species is a population that is reproductively isolated from other populations (reviewed

in: Mayr 1982). Both allopatric (geographic isolation of populations) and sympatric (without geographic isolation) speciation fall under the biological species concept. New species can be formed when one of these populations separates into two populations which, over time, becomes reproductively isolated, which is the process of speciation. Speciation allows for the vast biodiversity on Earth through the creation of new species while reproductive isolation maintains the separate lineages within this biodiversity.

1.1.1 Speciation mechanisms

Lack of interbreeding between two populations can occur due to pre- or postzygotic isolation, which occur before or after the fertilization of the zygote, respectively. For example, prezygotic isolation can come about due to temporal isolation where the populations mate at different times in the day (Muller 1942). If the populations no longer recognize each other as suitable mates, gene flow will decrease between the populations. There are several modes of speciation that fall under prezygotic or postzygotic isolation. Prezygotic isolation occurs when two populations are unable to form zygotes (Turelli *et al.* 2001). These cases involve species that may be isolated by temporal, geographical, mechanical, gametic, and behavioral mechanisms (Turelli *et al.* 2001). Compared to postzygotic isolation, where two populations can form hybrid zygotes, prezygotic isolation is considered to result in less gene flow because of the lack of mating between species and therefore less genetic recombination in offspring. The smaller amount of gene flow resulting from prezygotic isolation therefore serves to maintain a stronger genetic barrier between species. Although hybrids can be formed during postzygotic isolation, they are usually sterile or inviable (die before reproductive age; Turelli *et al.* 2001). Postzygotic isolation can arise from intrinsic causes, where the basis of sterility or

inviability are genetic, such as Dobzhansky-Muller incompatibilities (Turelli *et al.* 2001). Postzygotic isolation could be extrinsic (Turelli *et al.* 2001), where the environment influences hybrid fitness, such as a hybrid that is not fit in either parent species environments.

My thesis focuses on postzygotic isolation, where two species will mate but the hybrid formed between them is inviable or sterile. Although there is evidence for genetic incompatibilities underlying heterogametic hybrid sterility, the specific genes involved in most cases remain elusive (e.g., Civetta, 2016; Storchova *et al*. 2004). The two Drosophila species used in my thesis, *D. pseudoobscura* and *D. persimilis*, were chosen for this study not only because of their male hybrid sterility, but also because they produce heteromorphic sperm consisting of fertilizing and non-fertilizing sperm. The production of multiple sperm types within one ejaculate (heteromorphism) has been noted in other species but this reproductive trait is poorly understood (Till-Bottraud *et al*. 2005). For example, genes responsible for sperm heteromorphism (more than one type of sperm) are unknown for *Drosophila pseudoobscura* and *D. persimilis*, although regions in the chromosomes have been narrowed down through quantitative trait loci (QTL) mapping (Machado *et al*. 2007; Noor *et al.* 2001). Therefore, this species pair was used to investigate genes for both hybrid sterility and sperm heteromorphism.

1.1.2 Postzygotic isolation: Hybrid sterility

The inability to bear offspring or produce viable gametes is known as sterility. For females, this could be improper formation of eggs (Erdelyi and Szabad 1998), or abnormalities with reproductive organs (Sun and Spradling 2013). Immotile sperm, malformed sperm, malformed testes lacking sperm are common characteristics of male

sterility (Dobzhansky 1934; Pilder *et al.* 1997). When the hybrids formed between two populations are sterile, known as hybrid sterility, a barrier to gene flow is present. Without gene flow between them, populations are genetically isolated (Ehrman 1962), and their respective alleles can evolve along separate evolutionary paths.

Identifying the genetic basis of hybrid sterility can give understanding to the process of speciation. For separate species to be able to make viable hybrids, the two species must first recognize each other as potential mates, successfully mate and fuse egg and sperm, and then have the gene from the two lineages able to interact and function within the resulting hybrid. Species pairs with longer divergence time are more likely to make inviable hybrids or are unable to produce zygotes at all (Orr 1995; Turelli *et al.* 2001). There are questions as to how much genetic change is likely to result in speciation, hybrid sterility, hybrid inviability, or gametic incompatibility. The Drosophila genus offers great opportunities for speciation studies due to the species within this genus generally having a short generation time, the availability of different species pairs spanning a range of divergence times, and fully sequenced genomes for multiple species (Hales *et al.* 2015).

Several theoretical models have been used to help explain hybrid sterility and inviability (discussed further below). Haldane's rule was proposed to explain why hybrid sterility and inviabilityare more often present in the heterogametic sex (Haldane 1922). This model was expanded upon multiple times as genetic information became more accessible. The Bateson-Dobzhansky-Muller (BDM) model is one such case, where the greater understanding of how genes can change over time and how important

gene/protein interactions are for cell function helped explain hybrid sterility from a molecular level (Bateson 1909; Dobzhansky 1937; Muller 1940, 1942).

1.1.3 Hybrid sterility and Haldane's rule

According to the BDM model, genetic differences accumulate in two populations who come from a common ancestor (Bateson 1909; Dobzhansky 1937; Muller 1940, 1942). Alleles can mutate, duplicate, or be deleted over time in a population. New alleles can accumulate in each population, and alleles that were in common between the two populations can be lost, resulting in divergence at the DNA level. The lack of gene flow due to geographic isolation in combination with these genetic changes allows a population to become distinct from other populations of the same or closely related species. The different alleles within each genome function properly in their own genetic background but may be incompatible with the other population's genetic backgrounds. The variations in allele content and their incompatibility can become apparent when the two species' genomes come together in hybrids and cause abnormalities, like reduced growth rates (McDaniel *et al.* 2008). In a hybrid, the novel alleles can have negative genetic interactions with the other genetic background or can be missing genetic interactors from the same genetic background. The reduced fitness of hybrids can subsequently act as a selective pressure for against of interspecies mating. Genetic divergence that causes reproductive isolation via the reduced fitness of hybrid offspring can therefore further reinforce gene flow barriers between species, increasing genetic divergence and further reducing fitness in hybrids (Ayala *et al*. 1974).

There are many instances where only one hybrid sex is sterile (Davis *et al.* 2015; Ehrman 1962; Good *et al.* 2008). If only one hybrid sex is sterile, inviable, or missing, it

is the heterogametic (XY or ZW) sex, a phenomenon known as Haldane's rule (Haldane 1922). This curious observation of lowered heterogametic hybrid fitness could not be explained solely by improper interactions between the X (or Z) chromosome of one species and the Y (or W) chromosome of another species because of instances where Haldane's rule is observed in haplodiploid systems in the hemizygous sex (Koevoets and Beukeboom 2009). It also is not due to a particular sensitivity of male spermatogenesis, as ZW females are sterile and ZZ males are fertile (Haldane 1922; Laurie 1997).

The dominance theory was formulated to help explain the trend of unidirectional hybrid sterility in the heterogametic sex in Haldane's rule (Orr 1993). The BDM model explains how genes can diverge in separate lineages and lead to deleterious interactions when they come together (Bateson 1909; Dobzhansky 1937; Muller 1940, 1942). The BDM model also argues that one of the interacting alleles would need to be dominant in order to see an incompatibility in a first-generation hybrid. The Dominance model expands upon the BDM model in order to explain how dominance interactions can cause a disproportionate effect in the heterogametic sex (Turelli and Orr 1995). In this model, a recessive allele on the sex chromosome has a deleterious interaction with a dominant allele on an autosome. If the allele on the X chromosome is recessive, the negative effects would not be seen in female hybrids who are XX because recessive X-linked alleles from one species are masked by the dominant X-linked alleles of the other species (Figure 1). However, males, who are hemizygous for the X chromosome, will show the negative effects of interactions between alleles on the X chromosome and dominant autosomal alleles (Stevens 1905). This same model can be applied to species that are ZW/ZZ.

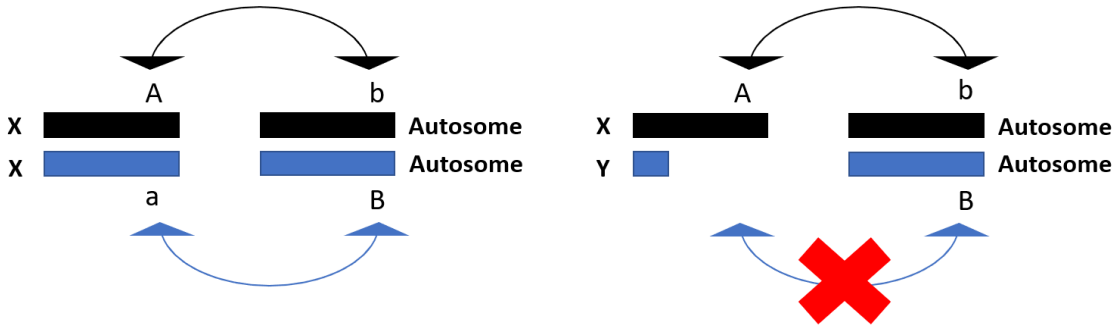


Figure 1 Dominance model representation of genetic incompatibility in male hybrids. Black represents chromosomes from one species and blue represents chromosomes from a different species. The ancestral species alleles were AABB. After the two populations split, the black species had 'b' rise to fixation (AAbb) and the blue species had 'a' reach fixation (aaBB). On the left is a female hybrid with interactions taking place between alleles on the X and autosome for each species chromosomes present. On the right is a male hybrid where the interaction between alleles from the blue species chromosome cannot take place.

There are few empirical examples where the genes underlying the dominance model have been identified. Those that have been identified contribute to the postzygotic barrier of hybrid inviability. For example, when *Drosophila melanogaster* females mate with *D. simulans* males, no sons are produced (Lachaise *et al.* 1986; Sturtevant 1920). The recessive gene *Hybrid male rescue* (*Hmr*) from the X chromosome of *D. melanogaster* and the dominant gene *Lethal hybrid rescue* (*Lhr*) from chromosome 2 of *D. simulans* have a negative interaction in hybrids (Brideau *et al.* 2006). In pure species *D. melanogaster* HMR protein forms a complex with LHR and binds to DNA near the

centromere, an interaction that is affected by the dosage of each protein (Thomae *et al.* 2013). *Hmr* in *D. melanogaster* (Hmr^{mel}) is expressed at much higher levels than *D. simulans* (Hmr^{sim}), while the reverse is true for *Lhr*, which has higher expression in *D. simulans* (Lhr^{sim}; Thomae *et al.* 2013). Male hybrids bearing a Hmr^{mel} and a *D. simulans Lhr* therefore have much higher expression of both proteins. Although LHR from *D. simulans* still binds to *D. melanogaster* HMR in hybrids, the shift in amount of protein expression results in an improper interaction between the two components. The higher level of HMR/LHR complex results in binding of the complex to abnormal areas of the chromosome, affecting transcription in those areas. If one of these particular alleles of two genes are mutated, there is a rescue of male hybrid viability (Hutter *et al.* 1990). Hmr^{mel} is a dosage compensatory gene, which results in higher expression in males because they only have expression from the Hmr^{mel} whose expression is higher than Hmr^{sim}. Because of this, there are higher levels of HMR in hybrid males than there are in hybrid females (Thomae *et al.* 2013). It is thought that this higher excess of HMR/LHR complex in hybrid males compared to hybrid females is the reason for the greater effect on male inviability, explaining the Haldane's rule effect of this gene combination. This imbalance of complex quantity also helps explain why male hybrids with Hmr^{sim} and Lhr^{mel} are viable due to there being lower expression of these orthologs (Thomae *et al.* 2013). This example shows how genes and their functions can evolve separately in two species, and the resulting species isolation through hybrid inviability can arise through recessive-dominant interactions that affect the heterogametic, but not homogametic, sex.

1.1.4 Genetic basis of hybrid sterility in Drosophila

One of the primary model systems used to study the genetic basis of hybrid sterility is the genus Drosophila (Coyne and Orr 1989; Coyne and Orr 1997; Orr and Presgraves, 2000). Drosophila species have been used in empirical studies that have informed theoretical models for the genetic basis of postzygotic isolation (Dobzhansky 1934; Koopman 1950; Orr and Presgraves 2000). Drosophila's short generation time allows for the observation of many individuals from different generations. There are also multiple species that can pair and form viable hybrids. These hybrids have been studied as to why they are fertile, sterile, or why parent of origin influences the viability and fertility of offspring (Bayes and Malik 2009; Civetta and Singh 1995; Coyne 1985; Palopi and Wu 1994; Phadnis and Orr 2009; Phadnis 2011; Ting *et al.* 1998; Wu and Davis 1993). Hybrid sterility studies using Drosophila species have tried to identify what at the molecular level causes reproductive failure in hybrids (Bayes and Malik 2009; Orr and Irving 2001; Phadnis and Orr 2009). Fully sequenced genomes for multiple fruit fly species aid in determining which genetic variants cause sterility in interspecies hybrids (Myers *et al.* 2000; Hahn *et al.* 2007).

1.1.4.1 *D. simulans* and *D. mauritiana*

The genetic basis of male hybrid sterility has been well-studied in the closely related species pair *D. simulans* and *D. mauritiana* (Figure 2; Bayes and Malik 2009; Ting *et al.* 1998). The X-linked gene *Odysseus-site Homeobox* (*OsdH*) affects hybrid sterility between these two species. The satellite-DNA binding protein produced by *OdsH* differs in abundance and localization during spermatogenesis between the two Drosophila species (Bayes and Malik 2009). OsdH localizes to additional locations on the Y

chromosome in *D. mauritiana* compared to *D. simulans* (Bayes and Malik 2009), and this difference could be an intrinsic postzygotic isolating mechanism. Male hybrid sterility could be caused by the gain of function of decondensation of the *D. simulans* Y chromosome through the interaction with *D. mauritiana* OsdH.

To understand the role of species-specific genes in hybrid sterility, researchers often use particular mating paradigms to cross genetic material from one species into the genetic background of the other species. Introgression consists of crossing viable inter-species hybrid individuals back with one of the parental species, for several generations (Harrison and Larson 2014). One study performed introgression followed by assays of gene expression on the resulting sterile vs. fertile males. They found that introgressed sterile males had *D. mauritiana OsdH,* while fertile males had *D. simulans OsdH*. Further, introgressed *OsdH* from *D. mauritiana* into the *D. simulans* genetic background led to misexpression of 14% of autosomal genes that are normally expressed in the testes of *D. simulans* males (Lu *et al*. 2010). The abnormal expression of OsdH, which is expressed in the beginning stages of spermatogenesis, affects the autosomal genes responsible for the later stages of spermatogenesis and results in hybrid sterility when there is a mismatch between OdsH and the autosomes. Interestingly, both OsdH and HMR cause sterility through genetic conflict between genes on the X chromosome and genes on other chromosomes, and both repress satellite DNA expression. The repetitive elements associated with satellite sequences may therefore have a deeper connection to how species become genetically distinct (Bayes and Malik 2009; Brideau *et al.* 2006).

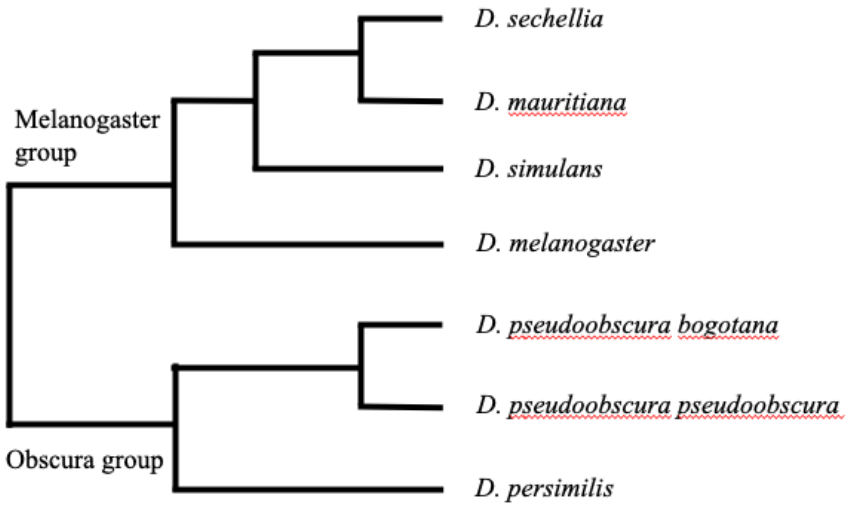


Figure 2 Partial phylogenetic tree of Drosophila species. The lines do not indicate time since divergence, instead they indicate only a divergence occurrence. Members of the obscura group and melanogaster group are represented (figure adapted from: Jezovit *et al.* 2017).

1.1.4.2 *D. pseudoobscura pseudoobscura* and *D. p. bogotana*.

Another gene linked to hybrid sterility in Drosophila is *Overdrive* (Orr and Irving 2001; Orr and Irving 2005; Phadnis and Orr 2009; Phadnis 2011), which affects hybrids between the subspecies *Drosophila pseudoobscura* and *D. pseudoobscura bogotana*. The *Overdrive* gene evolves at a fast rate, is within the inverted region of the X chromosome, and is expressed in the testes (Noor *et al.* 2007; Phadnis and Orr 2009). *Overdrive* has a large effect on hybrid sterility and has dominant interactors on the second and third chromosome, but each interactor individually has a small effect on sterility (Phadnis 2011). This X-autosomal interaction therefore supports the BDM incompatibility model

of hybrid sterility, but it does not support the Dominance model because Overdrive does not induce sterility in females when homozygous (Phadnis and Orr 2009).

The *Overdrive* gene also affects segregation distortion (also called 'meiotic drive') in hybrid. This is when there is abnormal segregation of particular chromosomes during meiosis, and can be observed as a skew in the sex ratio of offspring when it is the sex chromosomes that are affected. Hybrid males who become weakly fertile after aging, and when mated with females from either pure species or with hybrid females produce almost entirely female offspring (Orr and Irving 2005). Overdrive was found to be part of this distortion along with interactors on the 2nd and 3rd chromosome, separate from the interactors linked to male hybrid sterility (Phadnis 2011). Sex chromosome segregation distortion, rather than male offspring lethality, is the likely cause of the high proportion of female offspring (Orr and Irving 2005). The molecular interaction that takes place between *Overdrive* and its partners is still not clear for both sterility and segregation distortion.

1.2 The species pair *D. pseudoobscura* and *D. persimilis*

The species pair *Drosophila pseudoobscura* and *D. persimilis* have been used as a model for hybrid sterility because they are easy to rear in the laboratory and form sterile male hybrids but fertile female hybrids. The presence of fertile female hybrids allows for recombinants to be produced from hybrids, which is very useful for finding genetic loci linked to sterility (e.g. Machado *et al*. 2007; Noor *et al.* 2001). These closely-related species diverged about 0.55 million years ago (Wang *et al*. 1997). Their initial divergence occurred allopatrically (geographic isolation), but some populations of these species now live sympatrically (same geographic location) (Wang *et al.* 1997). These species form

hybrids in the wild, although the occurrence of hybrids is infrequent and male hybrids are sterile (Dobzhansky 1973; Noor *et al.* 2001). Females of the two species are morphologically almost identical, with exception that males of the two species differ in the shape of their external genitalia (Rizki 1951).

1.2.1 Genomes

Both *D. pseudoobscura* and *D. persimilis* have sequenced genomes. The two species have five chromosomes: X, Y, two, three, four, and the dot chromosomes (the fifth). These two species are genetically separated by multiple inversions on different chromosomes. These inversions restrict gene flow in the area of the inversions and the surrounding area (Machado *et al.* 2007). Inversions on the top and bottom arms of the X chromosome and on the second chromosome show higher amounts of divergence than elsewhere on those chromosomes (Noor *et al.* 2007). In the case of *D. persimilis*, there is a lower number of polymorphisms in the 2nd chromosome inversion compared to the 2nd chromosome of *D. pseudoobscura*, and it is thought this inversion is fixed (Machado *et al.* 2007; Noor *et al.* 2007). There is an inversion on the third chromosome, but the divergence in this inversion between *D. pseudoobscura* and *D. persimilis* is lower than the divergence in sequence in inversions on the X and second chromosomes (Noor *et al.* 2007).

1.2.2 Heteromorphic sperm in *D. pseudoobscura* and *D. persimilis*

Something unique about this species group is that males of both parental species produce heteromorphic sperm, which are sperm of different shapes or sizes produced at the same time (Pitnick *et al.* 2008). Indeed, in the genus Drosophila, only the *obscura* and *affinis* groups contain species that have sperm heteromorphism. Males in these groups

produce a longer fertile sperm morph, called eusperm, and a shorter non-fertile sperm morph, called parasperm (Holman *et al.* 2008; Holman & Snook, 2008; Moore *et al.* 2013; Snook *et al.* 1994). The different sperm morphs are usually characterized by the difference in length between eusperm and parasperm (Joly and Lachaise 1994; Snook 1997). Not only are there differences in the total length of the sperm, but there are also differences in both head and tail length between eusperm and parasperm (Snook 1997; Alpern *et al.* 2019). Recently, two parasperm morphs in *D. pseudoobscura* were characterized, parasperm 1 and parasperm 2 (Alpern *et al.* 2019).

While it was previously thought that there was only one parasperm morph with variation in size (Snook 1997), two parasperm morphs in D. pseudoobscura were recently characterized: parasperm 1 and parasperm 2. The length of eusperm for *D. persimilis* and *D. pseudoobscura* is about 300 μm, while parasperm 1 is about 55 μm and parasperm 2 is about 100 μm (Snook 1997; Alpern *et al.* 2019). This length difference is mostly caused by sperm tail length, since the nucleus lengths are approximately 30 μm, 15 μm, and 10 μm for eusperm, parasperm 2, and parasperm 1 respectively. Note that the heads of eusperm and parasperm contain the same amount of genetic content, but parasperm heads are a fraction of the size of eusperm heads. There is evidence that the parasperm in these species groups do not serve as just a 'cheap filler' in the ejaculate, but aid in sperm competition and protecting eusperm from female spermicides present in the female reproductive tract (Alpern *et al.* 2019). Specifically, parasperm 2 was seen in higher proportions when male competition was perceived by the copulating males, whereas both types of parasperm were positively correlated with eusperm survival in the presence of female reproductive tract proteins (Holman and Snook 2008; Alpern *et al.* 2019).

1.2.3 Spermatogenesis in *D. pseudoobscura* and *D. persimilis*

Spermatogenesis is the cellular process that results in the production of sperm cells. To be more specific, spermatogenesis is the development of mature spermatozoa from germline stem cells through meiosis and mitosis (Fuller 1993). There are multiple stages to this process and what happens at each stage is species specific. For example, the number of mitotic divisions and the total number of sperm produced can differ among species (Dobzhansky 1934).

Spermatogenesis starts and ends in the testes of *Drosophila pseudoobscura* and *D. persimilis* (Dobzhansky 1934; Fuller 1993). To begin, germ-line stem cells divide and differentiate into primary spermatogonial cells (Fuller 1993). Following this is five mitotic divisions, meiosis, elongation, and individualization (Dobzhansky 1934). At the apical end of the testes, a germ-line stem cell enters mitosis and becomes a primary spermatogonial cell (Fuller 1993). This primary spermatogonial cell separates from the hub where germ-line stem cells reside and is enclosed by two cyst cells. This enclosure, or capsule will surround the sperm during spermatogenesis (Fuller 1993). Now known as the secondary spermatogonia, the spermatogonial cell in the cyst undergoes five mitotic divisions, resulting in 32 primary spermatocytes, unlike *D. melanogaster's* four mitotic divisions (Dobzhansky 1934, Fuller 1993). At this stage, primary spermatocytes grow in size and replicate mitochondria (Dobzhansky 1934). DNA also replicates at this time as the cells continue to the premeiotic S phase and many of the genes are transcribed in preparation for differentiation after meiosis (Fuller 1993).

Mature primary spermatocytes enter meiosis and some of the steps in Drosophila meiosis are similar to typical meiosis. One difference is that the X and Y chromosomes

do not cross over for some Drosophila (Larracuente *et al.* 2010). In *Drosophila pseudoobscura* and *D. persimilis* the only region of homology between the two chromosomes are intergenic spacer regions (IGS). This differs from *D. melanogaster*, whom has both rRNA regions and IGS on the X and Y used for pairing during meiosis (Fuller 1993). A difference to canonical meiosis is that the nuclear membrane is not disintegrated (Dobzhansky 1934; Fuller 1993). Mitochondria of *D. pseudoobscura* align outside of the nucleus on either side to allow for equal separation. Chromosomes are attached to spindle fibers and separated to either side of the nucleus and the nucleus is pinched into two. The primary spermatocytes partially separate but a cytoplasmic bridge remains, connecting the now secondary spermatocytes (Dobzhansky 1934; Fuller 1993). Meiosis II follows with the separation of sister chromatids and another partial cell division, keeping the cytoplasm bridges intact and resulting in 128 spermatids.

Spermatids then reorganize mitochondria, assemble the axoneme, and begin elongation and individualization (Fuller 1993). The axoneme attaches to the nucleus of the spermatid. At this point the flagellar axoneme grows, allowing for the tail extend from this. During tail elongation, mitochondria are incorporated into the axoneme and tail. The nucleus of the spermatid elongates into a thin rod shape, which becomes the head of the sperm. Chromatin condenses during this stage, and the nucleus loses some of its volume, allowing for the slender rod head (which can be seen in Figure 3 - Fuller 1993). In order to condense, histones are removed from DNA and replaced with protamines, which functions as a DNA-binding protein is to more highly condense chromatin to fit into the nucleus of the sperm (reviewed in: Kanippayoor *et al.* 2013). After the tail has elongated, the two sperm morphs can be seen in the testes of *D.*

pseudoobscura and *D. persimilis*. Before this stage however, the eusperm and parasperm have not been distinguished from each other (Njogu *et al.* 2010). Spermatid individualize as the cytoplasmic bridges connecting each spermatid move down the length of the bundle. A bulge is formed as cytoplasm is removed from the spermatids and removed as waste. Each membrane can then cover the spermatid as its own sperm plasma membrane. Sperm tails are coiled, and mature sperm are ready for fertilization (Fuller 1993).

Because eusperm and parasperm develop in separate bundles, the mechanism(s) controlling this differentiation in phenotype must occur early on in spermatogenesis. What these developmental triggers are for two sperm morphs are still unknown, but knowing that the triggers affect an entire sperm bundle gives a clue as to when in spermatogenesis to search for a difference in gene expression, with a likely cellular stage being in primary spermatogonium.

1.2.4 Errors in spermatogenesis leading to sterility

Genes leading to sterility in *D. pseudoobscura* and *D. persimilis* hybrids are still unknown but there are candidate genes in other species that could lead to sterility in these hybrids. For example, the previously mentioned mutation to *OdsH* can lead to premeiotic failures in spermatogenesis in the interspecies hybrids of *D. mauritiana* and *D. simulans* (Bayes and Malik 2009). Genetic studies in hybrids are difficult, with few individual sterility loci identified, and so a richer source of candidate loci is needed to look at what causes spermatogenic failures *within* a species. Some insight can be gleaned from studies of the species *D. melanogaster*, a heavily-used model genetic organism, and a relative of *D. pseudoobscura* and *D. persimilis*.

The *D. melanogaster* recessive gene *rae1*, when mutated, leads to errors in male meiosis and spermatogenesis (Volpi *et al.* 2013). This gene is recessive and on the second chromosome of *D. melanogaster*. Flies with mutant *rae1* have errors in multiple spermatogenesis stages, including abnormal nuclei formation in primary spermatocytes, nonuniform nuclei and mitochondria in post-meiotic spermatids, and improper spermatid differentiation. During meiosis, chromosomes do not completely condense having a significant reduction in histone H3 phosphorylated at serine 10, chromatin shows improper alignment, and there are nondisjunction and chromatin bridges (Volpi *et al.* 2013). Although these sperm have issues during meiosis, the sperm are able to elongate, similarly to what is seen in sterile male hybrids from *D. pseudoobscura* and *D. persimilis*. Improper segregation and chromatin bridges are also observed in these hybrids, so it is possible that *rae1* is also involved in the sterile sperm phenotype for *D. pseudoobscura* / *D. persimilis* hybrids (Kanippayoor 2017).

Errors leading to sterility can also happen as late as the differentiation stage after spermatid elongation. For example, a loss of function mutation to *D. melanogaster PFTAIRE interacting factor 1A* (*Pif1A*) causes a disruption of the removal of cytoplasmic bridges and unneeded cytoplasmic components (Yuan *et al.* 2019). Male *D. melanogaster* with this mutation are sterile because of the post-meiotic effect of incomplete individualization has on the sperm (Yuan *et al.* 2019). This gene is a homolog to a human spermatogenesis gene, *CCDC157,* that has been linked to human male sterility (Reinke *et al.* 2013). Reduced number of sperm and what is thought to be incomplete separation of sperm in *D. pseudoobscura* and *D. persimilis* hybrids could be accounted for by a disruption in cytoplasmic bridge removal (Kanippayoor 2017). There are well-described

genes for human spermatogenesis whose homologues in Drosophila can be compared to the human counterpart to see how similar their function is (Fuller 1993). Even though there are differences in spermatogenesis between humans and Drosophila (Kanippayoor *et al.* 2013), human CCDC157 and Drosophila Pif1A are similar in 3D structure and both have higher transcription in the testes compared to other cells in the body (Yuan *et al.* 2019).

1.3 Genetic basis of *D. pseudoobscura*/ *D. persimilis* hybrid sterility

The genomes of *D. pseudoobscura* and *D. persimilis* differ by five inversions. There is no successful crossover between inverted regions of the two species' genomes in the hybrid genome (Machado *et al.* 2007). The lack of gene flow in these inverted areas allows for the opportunity for alleles to diverge separately in *D. pseudoobscura* and *D. persimilis*. These inverted regions, therefore, harbor polymorphisms specific to one species because there is no gene flow within the inverted regions between the two species. Perhaps not surprisingly, the inverted regions of the genome harbor candidate hybrid sterility genes.

1.3.1 Cellular characterization of sterility in *D. pseudoobscura* and *D. persimilis*

An abnormal sperm phenotype is seen in the hybrids of multiple species pairs, including *D. pseudoobscura* and *D. persimilis* (Kanippayoor 2017; Kanippayoor *et al.* 2020). Sterile hybrid sperm are non-motile and distinguished by a hole in the head of the sperm, giving it the appearance of a needle-eye (Figure 3).

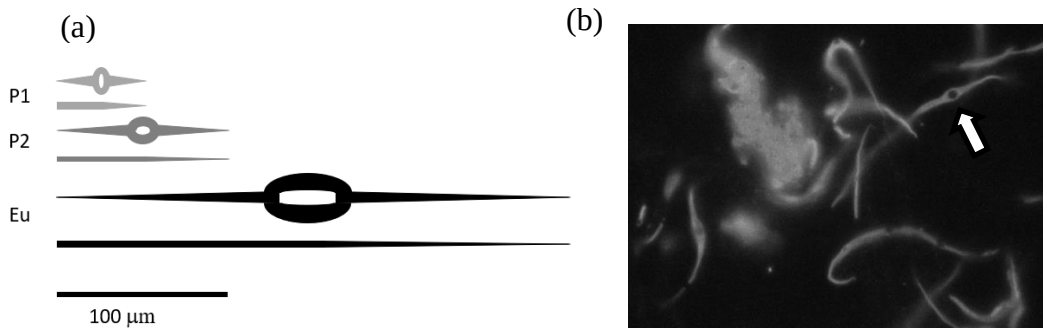


Figure 3 Illustration of sperm head phenotypes and microscope image of NE sperm. (a) Cartoon of sperm head morphologies. P1 is parasperm 1 (55μm), P2 is parasperm 2 (100 μm), and Eu is eusperm (300 μm). Two morphologies are shown for each: needle-eye (NE; top) and wt (bottom). (b): Image of sperm heads from a hybrid (second backcross) male from the *D. pseudoobscura* female and *D. persimilis* male cross at 100X magnification. The white arrows point to a NE sperm head. Wild-type eusperm and parasperm do not have this hole or "eye" present in the head of sperm.

Individuals with this sperm phenotype were found to have half the normal amount of sperm per sperm bundle and two tails per sperm, indicating an error in spermatogenesis. Having two tails would impede the ability of these "needle-eye sperm" to participate in fertilization (Tokuyasu 1974). What is not known about this hybrid sterility phenotype is what gene(s) are causing improper sperm formation in hybrids.

1.3.2 Cellular basis of sterility in other species pairs

Defects in spermatogenesis leading to male sterility in Drosophila can be observed at various stages in spermatogenesis, as discussed above. It is unknown if one cellular mechanism, affecting many stages, or if many different cellular mechanisms cause sterility. What is similar between some of these examples is the testes-specific expression of sterility-associated genes (Bayes and Malik 2009; Noor *et al.* 2007; Phadnis and Orr 2009; Yuan *et al.* 2019). The testes-specific expression of sterility-associated genes is also seen in mammals. The *Prdm9* gene in hybrid mice, which is only expressed in the testes linked to sterility and spermatogenesis dysfunction (Mihola *et al.* 2009; Nishino *et al.* 2019). What is also similar between mice and Drosophila is the nondisjunction of chromosomes during meiosis of spermatogenesis (Nishino *et al.* 2019; Volpi *et al.* 2013). Another study in mosquito hybrids found nondisjunction in hybrid sperm (Liang and Sharakhov 2019). The testes on these male mosquito hybrids were underdeveloped and sperm did not mature properly due to nondisjunction and chromatin condensation failure, resulting in large spermatids with two times as much chromosome content. Although the genes leading to this are likely not the same, this mechanism leading to sterility may be shared.

1.4 Sperm heteromorphism

Something unique about this species group is that males of both species produce heteromorphic sperm, which are sperm of different shapes or sizes produced at the same time (Pitnick *et al.* 2008). Sperm heteromorphism has evolved independently multiple times in separate taxa. Sperm heteromorphism has been noted in crustaceans, mollusks, fish, and insects (Till-Bottraud *et al*. 2005). For some of the species with sperm

heteromorphism, relatives within the same genus do not present with sperm heteromorphism. Indeed, in the genus Drosophila, only the obscura and affinis groups contain species that have sperm heteromorphism. For example, male *D. melanogaster* produce one type of sperm, which is used to fertilize eggs, while its relative *D. pseudoobscura*, produces three types of sperm.

In some cases, there is a fertilizing sperm type and a non-fertilizing sperm type, as in *D. pseudoobscura* and *D. persimilis*. How similar or different the non-fertilizing sperm are to the fertilizing sperm depends on the species. In some cases, like lepidopteran species, the non-fertilizing sperm is anucleated, lacking a nucleus (Lai-Fook 1982). Other organisms have nucleated non-fertilizing and fertilizing sperm morphs (Pasini *et al.* 1996). If a non-fertile sperm morph evolved multiple times, it must be providing a reproductive benefit to males. It has been proposed that the benefit stems from sperm competition with other males (Alpern 2013; Alpern *et al.* 2019). So far, genes or gene expression that allows multiple sperm morphs remains elusive for Drosophila.

Drosophila pseudoobscura and *D. persimilis* males produce a longer fertile sperm morph, called eusperm, and a shorter non-fertile sperm morph, called parasperm (Holman *et al.* 2008; Holman & Snook, 2008; Moore *et al.* 2013; Snook *et al.* 1994). The different sperm morphs are usually characterized by the difference in length between eusperm and parasperm (Joly and Lachaise 1994; Snook 1997). Not only are there differences in the total length of the sperm, but there are also differences in both head and tail length between eusperm and parasperm (Snook 1997; Alpern *et al.* 2019).

Recently, paraspem was found to be two separate parasperm morphs in *D. pseudoobscura*, parasperm 1 and parasperm 2 (Alpern *et al.* 2019). The length of eusperm for *D. persimilis* and *D. pseudoobscura* is about 300 mm, while parasperm 1 is about 55 mm and parasperm 2 is about 100 mm (Snook 1997; Alpern *et al.* 2019). This length difference is mostly caused by sperm tail length, since the nucleus lengths are approximately 30 mm, 15 mm, and 10 mm for eusperm, parasperm 2, and parasperm 1 respectively. Note that the heads of eusperm and parasperm contain the same amount of genetic content, but parasperm heads are a fraction of the size of eusperm heads. There is evidence that the parasperm in these species groups do not serve as just a 'cheap filler' in the ejaculate, but aid in sperm competition and protecting eusperm from female spermicides present in the female reproductive tract (Alpern *et al*. 2019). Specifically, parasperm 2 was seen in higher proportions when male competition was perceived by the copulating males, whereas both types of parasperm were positively correlated with eusperm survival in the presence of female reproductive tract proteins (Holman and Snook 2008; Alpern *et al.* 2019).

1.4.1 Evolutionary drives for sperm heteromorphism

Multiple studies have examined how heteromorphic sperm might enhance male fitness. One theory is that parasperm are produced to provide protection to eusperm against the female reproductive tract proteins (Holman and Snook 2008). The female reproductive tract produces spermicides, as is the case for *D. pseudoobscura* females. More eusperm survive *D. pseudoobscura* female reproductive tract proteins when more parasperm are present, with parasperm 1 and 2 being equal contributors (Holman and Snook 2008;

Alpern *et al.* 2019). This could suggest that of parasperm allow eusperm to have better chances of surviving and fertilizing eggs.

Another purpose of parasperm may be to provide an advantage in male-male sperm competition. Parasperm could be used to physically block or displace the sperm of other males, so the eusperm has greater odds of fertilizing the available eggs. After comparing *D. pseudoobscura* parasperm and eusperm quantities after exposure to other males, parasperm 2 proportion was altered based on male competitive environment as well as eusperm proportion (Alpern 2013; Alpern *et al.* 2019). Thus, it seems likely that parasperm can provide multiple advantages to males, but this question needs further investigation for *D. pseudooscura* and other species with heteromorphic sperm.

1.5 Overview of Thesis

The focus of this thesis was to identify the genetic basis of both sperm heteromorphism and hybrid sterility in *D. persimilis* and *D. pseudoobscura*. Comparisons of whole genome sequencing data of introgressed lines were used to find single nucleotide polymorphisms (SNPs, single base change in DNA) associated with sperm heteromorphism and sterility.

Two hypotheses, and predictions, were formed, one for sterility and one for heteromorphism:

1. Alleles associated with sterility are more commonly found in sterile than in fertile introgressed hybrid males. If introgressed parental alleles (single nucleotide polymorphisms) are more frequent in the sterile hybrid males than in fertile hybrid males, then these SNPs are in genomic regions associated with sterility.

2. Loci associated with sperm morph determination are more commonly found in introgressed males with a specifc sperm morph (e.g. eusperm) having a needle-eye phenotype than males with a different sperm morph (ex. parasperm). If introgressed parental alleles (single nucleotide polymorphisms) are more frequent in the males with one sperm morph having a needle-eye (NE) phenotype, then those SNPs are in loci associated with that sperm morph.

I examined the genetic basis of hybrid sterility and sperm heteromorphism between *D. pseudoobscura* and *D. persimilis* by isolating SNPs associated with either the sterile NE phenotype or the wild-type phenotype. I repeatedly backcrossed hybrids of each species in order to introgress genomic regions from one species into the genetic background of the other species. I first backcrossed (BC) female hybrids into two genetic backgrounds, *D. pseudoobscura* or *D. persimilis*, for eleven generations using females whose brothers had both sterile needle-eye eusperm and parasperm, needle-eye eusperm but wild-type (WT) parasperm, or WT eusperm but needle-eye parasperm. Each generation, males were scored for sperm phenotype, while their hybrid sisters were collected and used in each subsequent BC (Figure 4). By selecting for these sperm traits and BCing repeatedly into one parental genome, the genetic background of the species not used in the BC would diminish except for loci linked to sterility and sperm phenotype. Each generation, DNA was pooled from related hybrid males with the same sperm phenotype. To identify the genetic basis of the sterility phenotype and sperm heteromorphism, I used whole genome next-generation sequencing. Data analysis of the sequences was performed by Dr. Katharine Korunes from Duke University, who compared differences in loci for each sterile and fertile sperm morphology. A comparison

of sterile *vs.* fertile genotypes resulted in many SNPs that could be pursued for candidate loci for both male hybrid sterility and sperm heteromorphism.

Chapter 2

2 Methods

2.1 Drosophila strains and stock maintenance

Drosophila pseudoobscura AFC 57 (*pse57*) and *AFC 60* (*pse60*) and *Drosophila persimilis MSH* (*per*) were provided by the Dr. M. Noor Lab (Duke University). Both species lines were previously sequenced (McGaugh and Noor, 2012). All flies were kept on standard cornmeal molasses media (Bloomington Drosophila Stock Center Recipe) in 30 mL polystyrene vials plugged with cotton. These stocks were maintained at room temperature 21-22°C with a 14:10 hour light:dark cycle.

2.2 Introgression crosses

For all crosses, the parents were removed from the vial after approximately two weeks, which is when larvae were readily visible. Offspring eclosed three to four weeks after the initial mating of the parents. Single pairs of virgin *per* males were crossed with *pse60* females to produce F1 hybrids. The initial cross for *per* was one *per* male and one *pse* female. The initial cross for *pse* was two separate crosses of one *pse* male and one *per* female. Female virgin F1 hybrids were aged 5-7 days and paired with either *pse60* (backcross *pse*) or *per* (backcross *per*) males, aged 5-7 days and allowed to mate and lay eggs until larvae were present. Once backcross larvae were present (approximately two weeks), the parents were removed from the vial. Male hybrid backcross offspring (backcross 1: BC1) from this cross were scored for sperm phenotype (see below). Virgin female BC1 hybrids, sisters to these males, were again mated in single pairs with males of the backcross parental species. This same protocol was repeated each generation of

backcross until BC10 or BC11, leaving approximately 0.02% of the maternal species genetic background behind.

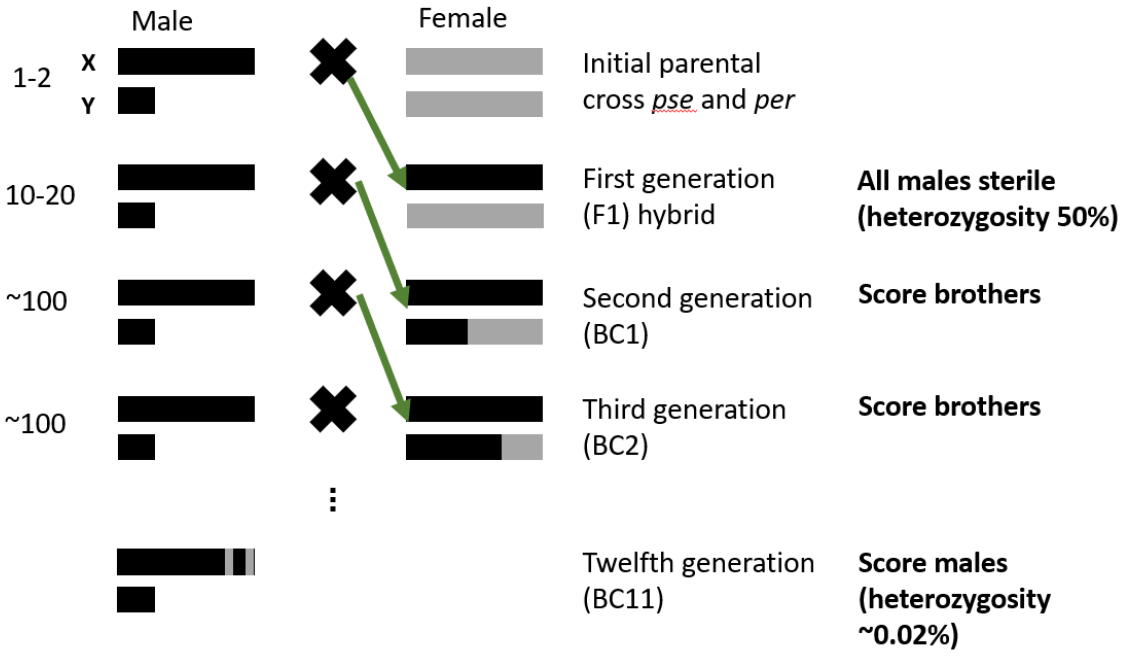


Figure 4 BC diagram illustrating how the loci of interest introgresses with each generation. Long bars represent X chromosomes and short bars represent Y chromosomes. BC females are chosen based on brother's sperm sterility phenotype. Loci associated with the selected trait remain in the next BC. Loci not associated with the phenotype may be lost through recombination with each generation. By BC11, remaining loci from the original female parental background should be potentially associated with the selected trait, having only one sperm type present. One case is shown but recombination locations differ over the population. Approximately 100 females were used for crosses each generation. Black represents the paternal species DNA and residual F1 male DNA; gray represents maternal species DNA. Green arrows show the progression from one BC to the next generation.

2.3 Hybrid male testes dissections and imaging

Within 36 hours of eclosion, virgin hybrid males were anesthetized by CO_2 and decapitated. Testes were removed from the male hybrids in testes buffer (185 m*M* KCl, 47 m*M* NaCl, 10 m*M* Tris-HCl) using ultra fine dissecting tweezers on a glass dissecting plate. Testes were moved to siliconized cover slips where each testes pair was placed in their own 30 μl drop of testes buffer. Nicks were made in each testis using the ultra fine dissecting tweezers to release sperm. Twenty microliters of testes buffer were drawn of with a pipette, taking care not to remove the sperm mass, and 20 μl of 0.5 μl/ml of 4',6-diamidino-2-phenylindole (DAPI) nuclear stain was added. Samples were left for two minutes to allow sperm to be stained by DAPI, and then samples were washed three times with testes buffer. Sperm samples were examined and imaged using an Upright Zeiss AxioImager Z1 Compound Fluorescent Microscope with the fluorescent Zeiss MRc5 camera.

Hybrid male sperm samples were scored for the presence of eusperm, parasperm, needle-eye (NE) eusperm, or NE parasperm. Parasperm and eusperm are easily distinguishable from each other by sperm head size (Alpern *et al.* 2019), and wildtype and needle-eye can be distinguished by the presence of the hole shape in the sperm head (Figure 3).

2.4 DNA isolation and sequencing

DNA was isolated from the carcasses of dissected frozen hybrid males, whose sperm phenotypes (NE, WT, parasperm or eusperm) were examined, using a modified Phenol/Chloroform method (Sambrook *et al.* 1989). 500 µl of squishing buffer (100 µl Tris HCL pH 8.0, 20 µl EDTA, 50 µl 5M NaCl, sterile H_20) was mixed with 8.7 µl of proteinase K (20mg/ml solution). The squishing buffer/proteinase K mix was added to one tube of pooled flies with the number of males listed in Table 1, with the final/last generation used for that sample. A motorized pestle was used to homogenize and squish the flies. Each sample was incubated for 30 minutes at 37°C. Five-hundred microliters of Invitrogen Ultra-Pure phenol:chloroform:isoamyl alcohol (25:24:1, v/v) was added each sample and inverted to mix. Samples were centrifuged at 10000 rpm for 12 minutes. Approximately 390 µl of the top layer was drawn off and added to a tube containing 1 ml of ice-cold 95% ethanol and 20 µl of 3M sodium acetate. Samples were then placed at -20°C for one hour and then centrifuged for 20 minutes at 10000 rpm. After decanting the liquid from the tubes, 250 µl of 70% ethanol was added and centrifuged for three minutes at 10000 rpm. The ethanol was then carefully removed with a pipettor as to not disturb the DNA pellet and allowed to dry. The DNA pellet was resuspended in 100µl of elution buffer (10 mM Tris·Cl; 0.5 mM EDTA; pH 9.0).

Fifty microliters of each DNA sample were sent for Illumina's NovaSeq 6000 S4 PE150 whole genome sequencing at Genome Quebec Innovation Centre with paired-end reads. The adapter 1 sequence was AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC, and the adapter 2 sequence was AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT. Raw genomic sequences

with an average sequencing depth of 70 million reads and 30x coverage were received for 20 samples. Each sample came with three FASTQ files.

Table 1. Twenty DNA samples sent for whole genome sequencing.

Comparison group[1]	Sample name[2]	# male flies[3]	BC parent[4]	Final BC[5]	Eusperm[6]	Parasperm[6]
1	10	6	*Pse*	10	WT	WT; NE
1	11	11	*Pse*	10	WT; NE	WT; NE
2	12	6	*Pse*	10	WT	WT; NE
2	13	23	*Pse*	10	WT; NE	WT; NE
3	17	9	*Pse*	9	WT	WT; NE
4	18	6	*Pse*	10	WT	WT; NE
4	19	19	*Pse*	10	WT; NE	WT; NE
5	C2_724	11	*Pse*	11	WT; NE	WT; NE
5	F1_719	4	*Pse*	11	WT	WT; NE
6	20	5	*Pse*	11	WT	WT; NE
6	D1_719	12	*Pse*	11	WT; NE	WT; NE
7	A1_719	14	*Pse*	11	WT; NE	WT; NE
7	B1_719	5	*Pse*	11	WT	WT; NE
8	28	5	*Per*	10	WT	WT; NE
8	29	8	*Per*	10	WT; NE	WT; NE
9	30	11	*Per*	11	WT	WT; NE
9	31	11	*Per*	11	WT; NE	WT; NE
9	35	6	*Per*	9	NE	WT; NE
10	34	11	*Per*	9	WT; NE	WT; NE
11	25	11	*Per*	11	WT	WT; NE

[1] Comparison group have the same parental lineage through the backcross generations. Each comparison group (1-11) is represented by a number. Samples with the same number are in the same comparison group.

[2] Sample name indicates the label used by the genome sequencing facility for that DNA sample.

[3] # male flies indicate how many flies were pooled into one sample.

[4] BC parent indicates the paternal line the sample came from. Males pooled in a sample had the same sperm types present in the testes. Pse: *D. pseudoobscura*, Per: *D. persimilis*.

[5] Final BC indicates the last BC the sample came from, for example, sample 10 consists of 6 males with pse as the paternal background, and the samples came from BC10 or earlier. BC lineage indicates the lineage all males from the sample came from.

[6] Eusperm and parasperm phenotypes were scored as either wild-type (WT), needle-eye (NE), or both (WT; NE).

2.5 Whole genome sequence analysis

Raw genomic sequence was sent to Dr. Katharine Korunes for whole genome sequence assembly. Assembly was conducted using the known sequences for *D. pseudoobscura* and *D. persimilis* and aligning the FastQC sequences from the backcross samples in the Burrows Wheeler Aligner (BWA) software (Li and Durbin, 2010). Once the whole genome sequences were assembled for the hybrid DNA samples (Table 1), Dr. Korunes identified single nucleotide polymorphisms (SNP) in these hybrid sequences when compared to the parental species sequences through the use of the SNP calling software Genome Analysis Toolkit (GATK) and then highlighting SNPs specific to one species found in the hybrid sequences (McKenna *et al*. 2010). Loci that were found in hybrids with a specific phenotype (NE or no NE), but not in other hybrids with a different phenotype were investigated to assess whether that locus was introgressed. For example, if a SNP was found in samples where the male had NE parasperm and that SNP was not found in samples lacking that sperm phenotype, that SNP would be considered biologically relevant. Two types of comparison where performed: SNPs were compared between hybrids with different sperm phenotypes, and between each hybrid and parent species sequences obtained from the Noor lab at Duke University. Dr. Korunes ran a principle component analysis (PCA) analysis using PLINK and R studio in order to see

if there were patterns in the SNP data, in order to potentially identify cluster of SNPs on specific chromosomes. Specifically, the PCA was used to see if there was variance in the data between the two genetic backgrounds (*pse* paternal or *per* paternal lineage) and if there was variance between the males with NE eusperm and those lacking NE eusperm.

Chapter 3

3 Results

The goal of this study was to find candidate loci involved in male hybrid sterility for *D. pseudoobscura* and *D. persimilis* crosses and to find candidate loci involved in sperm hetermorphism in these two species. The approach was selecting for specific sperm phenotypes [eusperm, parasperm, needle-eye (NE) eusperm, NE parasperm] observed in hybrid males, and repeatedly backcrossing the sisters of the hybrid males with each of the two parental species in order to isolate SNPs associated with the various sperm types from the maternal genetic background. The hybrid male offspring from late backcross generations (BC 9, 10 and 11) of both cross directions were used for whole genome sequencing.

3.1 Sperm Phenotypes

Hybrid male sperm phenotypes were examined through dissection and fluorescent microscopy for each BC generation. The details of how many males from each generation possessed the different sperm type combinations is summarized in Table 2 and further discussed below.

3.1.1 First generation males

The first males analyzed for sperm morphology were from the F1 generation of hybrids. These males were all sterile and presented only with the NE sperm. All males in this generation had the same phenotype, unlike what occurs in subsequent BC generations. The NE sperm in these hybrid males was qualitatively assessed as uniform in size. Individuals from the F1 generation present with a NE sperm phenotype that is an

approximately intermediate length between the parasperm and eusperm as previously reported (Kanippayoor *et al.* 2020). There were no wild-type sperm present in the testes of these males.

3.1.2 Subsequent BC generations

Starting at backcross 1, male hybrids showed both wild type sperm and NE sperm in their testes at the same time. There was also a reappearance of heteromorphism in the sperm, with pboth parasperm and eusperm showing wild-type and NE morphologies (Figure 5, Figure 6). There was a distinct separation of the NE phenotype into the shorter parasperm and longer eusperm morph (~8μm and ~22μm head lengths), with no morphs of intermediate head length between these two. Individual hybrid males differed by which sperm morph combinations were present in their testes, and presented with different combinations of wild-type eusperm, wild-type parasperm, NE eusperm, and/or NE parasperm (Table 2).

3.1.3 Backcross 10 and 11 individuals

It was expected that by BC10, the loci controlling sperm heteromorphism and sterility would be isolated from the maternal genome and the phenotypes caused by these loci would be presented singly in male hybrids, where one male would only present with one sperm type, like only WT eusperm. This prediction was made based on the previous work done by Kanippayoor (2017).

In these two generations, male hybrids still presented with both wild-type and NE sperm types of both parasperm and eusperm within their testes (Table 2). For the BC10*pse* and BC11*pse*, the most common phenotype for male hybrids was all four types

of sperm present. The same is not true for BC10*per* and BC11*per*. The proportion of individuals from these two BC generations who present all four sperm types is similar to the proportion of individuals lacking in NE eusperm.

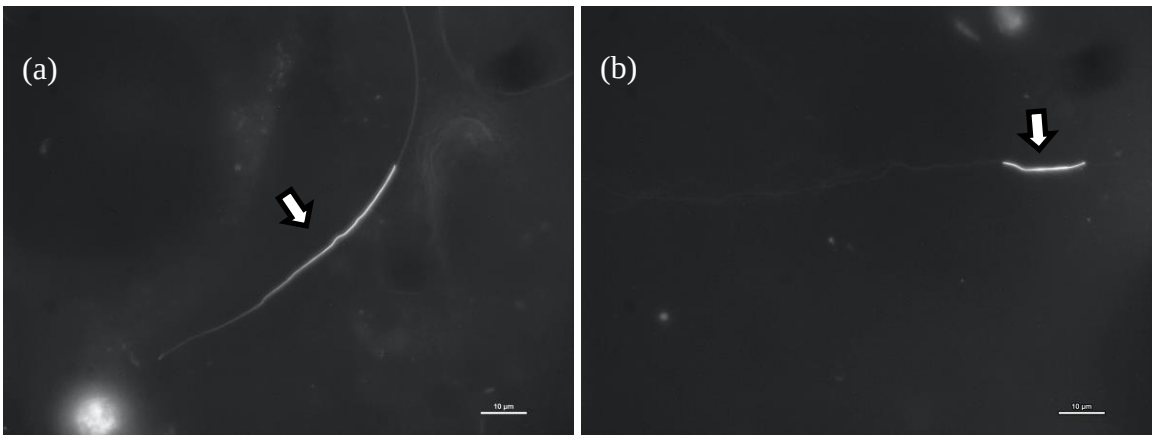


Figure 5 Wild type sperm present in hybrid male. (a) Wild type eusperm head in BC4 male hybrid with *D. persimilis* father. (b) Wild-type parasperm head in BC4 male hybrid with *D. persimilis* father. DAPI stained sperm 100x objective magnification using a fluorescent microscope. White arrows point to the heads of the sperm, which is long and thin. Note the distinct difference in size of the heads between the two sperm types. Scale bar is 10 μm**.**

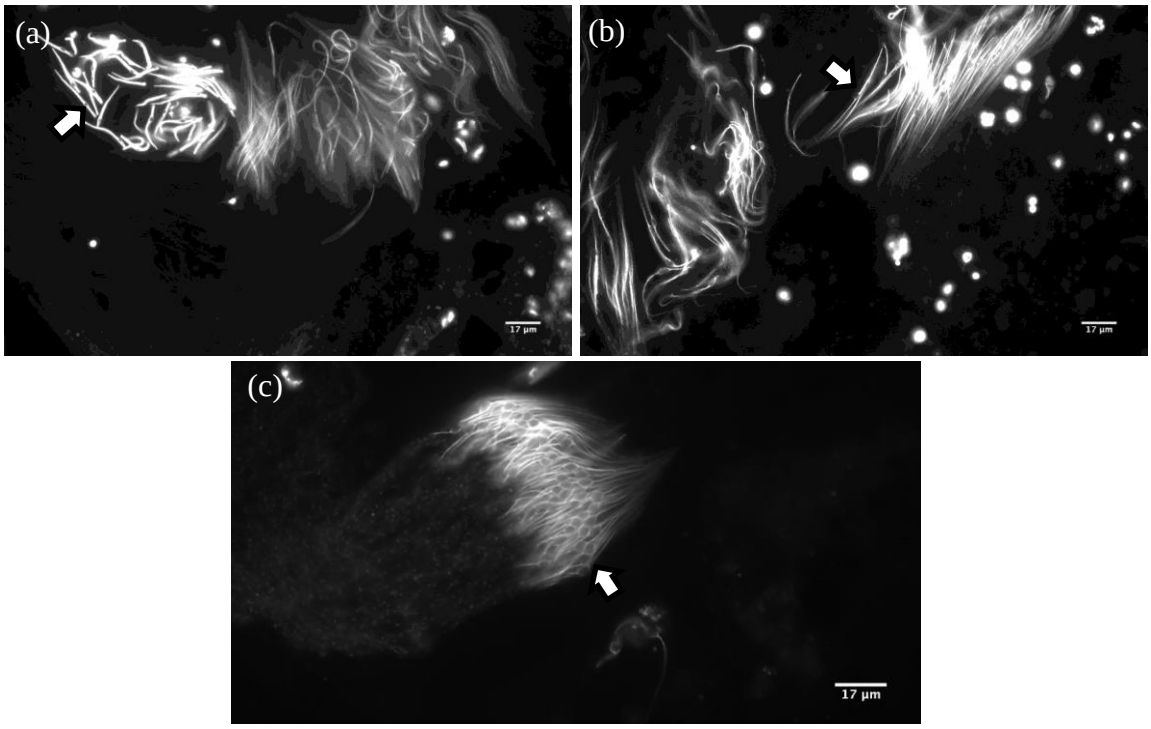
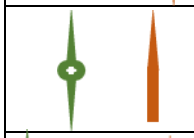
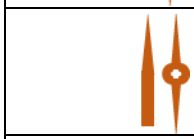
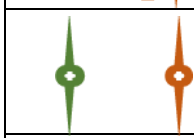
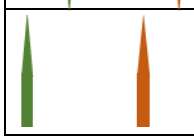


Figure 6 Male hybrids present with two types of NE sperm. (a) Parasperm needle-eye (NE) sperm heads are present on the left while eusperm NE is present on the right from BC9 hybrid male with *D. pseudoobscura* as the father. (b) Eusperm NE sperm heads from same male as (a). (c) Eusperm NE sperm heads from a BC11 hybrid male with *D. persimilis* paternal parent. The sperm were stained with DAPI and imaged with a fluorescent microscope with an objective magnification of 63x. White arrows point to NE sperm. Scare bar is 17 µm.

Table 2 Number of hybrid males with each sperm phenotype in each BC generation for BC*pse*. Sperm phenotypes are represented by symbols. Green is eusperm (eu) and orange is parasperm (para). The straight images are WT sperm and the line with a hole represents NE sperm. Sperm phenotype was separated based on the presence of the NE phenotype and size of the sperm. The categories observed were: all sperm phenotypes, WT eu and WT para and NE para, WT eu and WT para and NE eu, WT para and NE eu and NE para, WT eu and NE eu and NE para, only WT sperm, only NE sperm, only NE para and WT para, only NE para and WT eu, only NE eu and WT para, only NE para. The following categories were not observed: only eu, only para, only NE eu, only NE eu and WT eu. NE: needle-eye, WT: wild type, para: parasperm, and eu: eusperm.

Total Males	57	74	63	48	86	99	94	156	268	260
	0	0	1	0	0	0	0	0	0	0
	0	0	1	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	1	0	0
	7	0	3	0	1	0	0	1	0	0
	3	6	4	0	0	0	1	0	1	1
	0	0	3	0	4	1	0	1	2	0

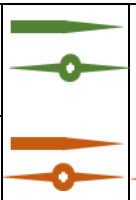
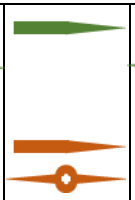
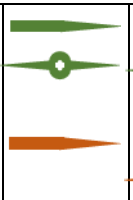
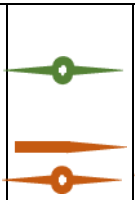
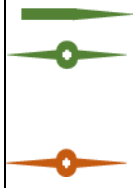
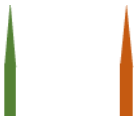
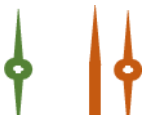
EUSPERM *PARASPERM* *BC Generation* ↓					
BC2	21	11	5	10	0
BC3	34	7	2	25	0
BC4	15	16	2	16	2
BC5	33	4	6	3	2
BC6	47	24	8	2	0
BC7	75	21	1	1	0
BC8	46	39	3	4	1
BC9	83	67	2	1	0
BC10	150	110	5	0	0
BC11	189	67	1	0	2

Table 3 Number of hybrid males with each sperm phenotype in each BC generation for BC*per*. Sperm phenotypes are represented by symbols. Green is eusperm (eu) and orange is parasperm (para). The straight images are wild type sperm and the line with a hole represents needle-eye sperm. “All” represents individuals who presented with all four sperm phenotypes. The categories observed were: all sperm phenotypes, WT eu and WT para and NE para, WT eu and WT para and NE eu, WT para and NE eu and NE para, WT eu and NE eu and NE para, only WT sperm, only NE sperm, only NE para and WT para, only NE eu and WT para. We did not observe the following categories: only eu, only para, only NE eu, only NE para and WT eu, only NE eu and WT eu, only NE para. NE: needle-eye, WT: wild-type, para: parasperm, and eu: eusperm.

Total Males	45	54	74	63	48	49	69	111
	0	0	2	1	0	1	0	0
	2	0	2	0	1	0	0	0
	5	4	3	1	0	2	0	0
	0	0	4	1	8	0	1	3
	0	0	0	2	0	1	2	0
	20	8	7	1	1	10	0	0

	0	0	4	10	7	1	2	1
	3	4	20	15	18	10	28	55
	15	38	32	32	13	24	36	52
EUSPERM *PARASPERM*	*BC Generation* →	*BC4*	*BC5*	*BC6*	*BC7*	*BC8*	*BC9*	*BC10*	*BC11*

3.2 SNP analysis

3.2.1 Genome sequencing

The raw sequences from the whole genome sequences had a quality score of 33, which was a sufficient score for this analysis. The coverage for the sequences was 30x. Read depth was an average of 70 million reads for each sample. See Appendix A for further detail. Sequences for the parent species were obtained from the Noor lab (Machado *et al*. 2007; Noor *et al.* 2001) and are also available on http://pseudobase.biology.duke.edu/.

3.2.2 SNP analysis

SNPs were called using GATK. The number of SNPs differed between chromosomes that were assessed for all hybrid samples: the left (chrXL) and right (chrXR) arms of the X chromosome, and the 2nd, 3rd, and 4th chromosomes (Table 4). The 5th or "dot" chromosome was not included due to the highly condensed nature of that chromosome having low levels of crossover and is previously reported to have little divergence between *D. pseudoobscura* and *D. persimilis* (Noor *et al*. 2007).

Table 4 Number of single nucleotide polymorphisms from each chromosome after variant calling from all samples**.**

	chrXL	chrXR	chr2	chr3	chr4
# of SNPs	571,664	743,422	932,822	410,485	895,034

3.2.3 SNPs and QTLs

Previous work identified inversions on the XL, XR, and 2nd chromosome are candidate regions for divergence between *D. pseudoobscura* and *D. persimilis* due to the strongly reduced gene flow in these areas (Machado *et al.* 2007; Noor *et al.* 2001). There are SNPs from the current study found in these regions (Table 5). The inversion with the largest number of SNPs inside the inversion was the 2nd chromosome inversion. The percentages of SNPs within the inversions were calculated using the total number of SNPs reported for each chromosome in Table 4.

Table 5 SNPs found within inverted regions of the XL, XR, and 2nd chromosome. Each chromosome listed has one inversion region.

Inversion Chromosome	XL	XR	2
# SNPs in the inversion	2,175	28,626	262,819
% of chromosome inversion covers	41.67%	48.27%	25.80%
% of total SNPs on that chromosome	0.38%	3.85%	28.17%

3.2.4 Linking SNPs to phenotype

Principle component analysis (PCA) is used to visualize variation in a data set (Wold *et al.* 1987; Lever *et al*. 2017). In this case, the PCA was used to observe variation in all SNPs between sequenced samples and give a broad visualization of the behavior of the SNP data in reference to phenotype (sperm morphology) and species (*pse* or *per*). Samples were grouped based on relatedness and phenotype. PC1 shows the greatest variance in data with each PC after that explaining less and less variability (Wold *et al.* 1987; Lever *et al*. 2017). The points plotted on the PCA represent samples. If the samples from different groups are mixed, it means the samples across all groups are similar to each other. For the present samples, if the samples from the two different species were mixed, it means the SNPs associated with those samples' phenotypes are similar.

The PCA shows clustering for all five chromosomes for samples from the *D. persimilis* paternal lineage, but less so for the *D. pseudoobscura* genetic background (Figure 7). This is also true when the data from the four autosomes were pooled and compared to the X chromosome (Figure 8). The same pattern emerges where *D. persimilis* background samples have few differences from each other but are separate from samples with *D. pseudoobscura* as the genetic background. Although samples are separated by paternal species, as expected, they are not separated by phenotype (sterile or fertile). The PCA therefore shows that the variants called from the samples are species specific but does not show phenotype specific variability between samples.

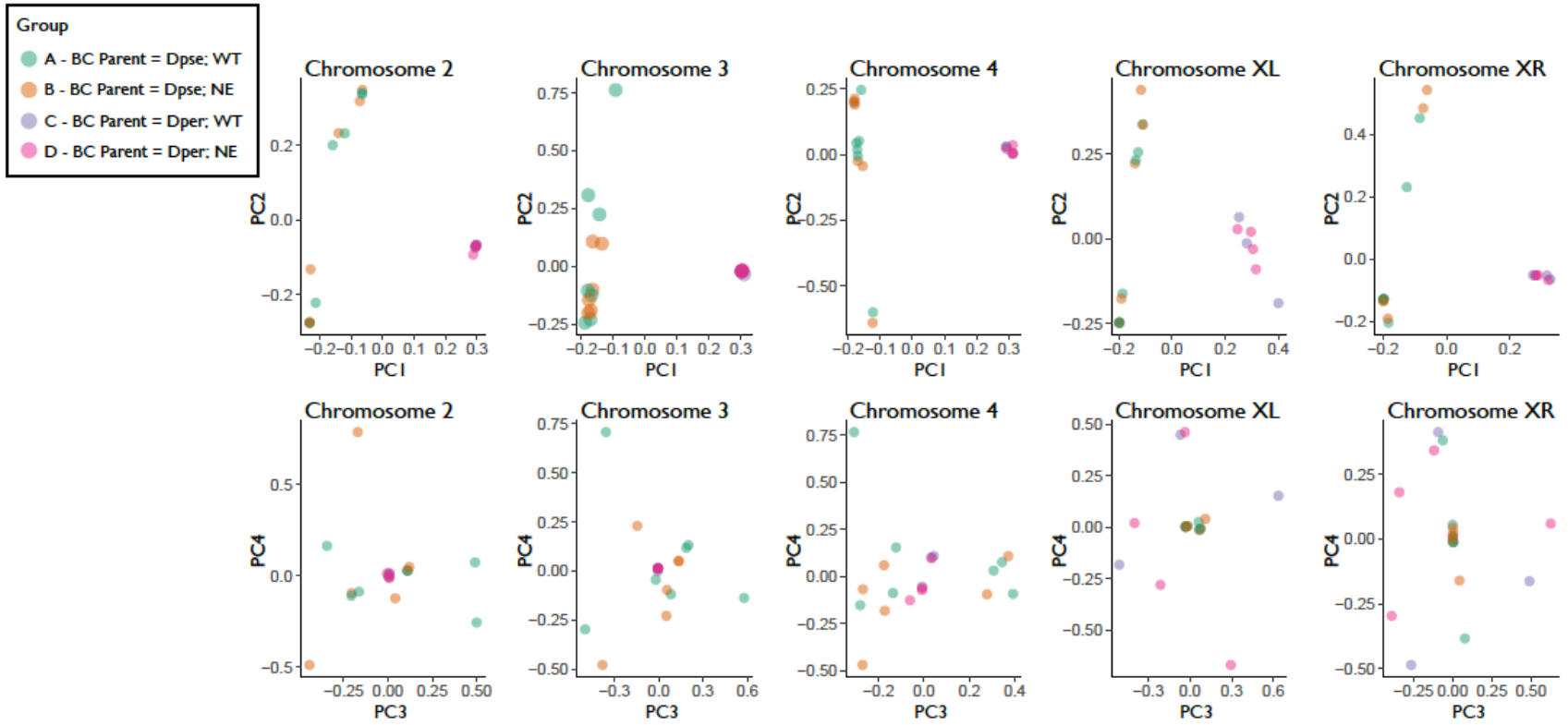

Group
A - BC Parent = Dpse; WT
B - BC Parent = Dpse; NE
C - BC Parent = Dper; WT
D - BC Parent = Dper; NE
Chromosome 2
Chromosome 3
Chromosome 4
Chromosome XL
Chromosome XR
PC1
PC2
PC3
PC4

Figure 8 Principle component analysis of SNPs for males with either WT or NE sperm from *D. pseudoobscura* paternal lineage or *D. persimilis* paternal lineage for each chromosome. Each dot represents a sample from sequencing. PC1 (X axis) for each chromosome is plotted against PC2 (Y axis) in the left column of graphs. PC3 (X axis) for each chromosome is plotted against PC4 (Y axis) in the right column of graphs.

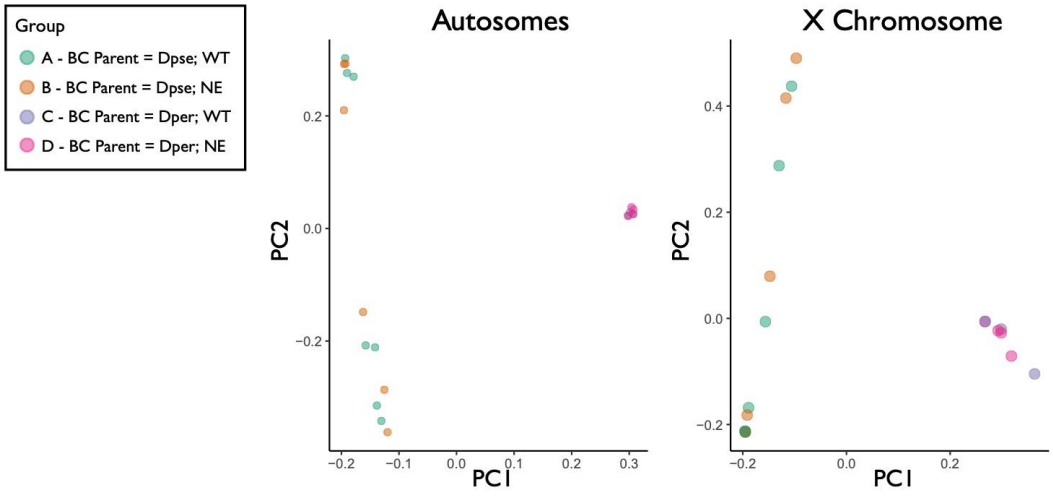


Figure 9 Principle component analysis of SNPs for males with either WT or NE sperm from *D. pseudoobscura* paternal lineage or *D. persimilis* paternal lineage for autosomes vs the X chromosome. Each dot represents a sample from the whole genome sequencing. PC1 is on the X axis and is plotted against PC2 on the Y axis.

Chapter 4

4 Discussion

Through the use of backcrossing with phenotypic selection, I examined the heritability of different sperm phenotypes in interspecies hybrids. I expected that loci controlling these different phenotypes (eusperm, parasperm, NE eusperm, and NE parasperm) would separate such that one hybrid male would only present one sperm phenotype based on which loci were introgressed into that male's genetic background. This prediction was made with the assumption that few loci control these sperm phenotypes and that these phenotypes would separate over backcross generations as they did in a study on hybrids formed from a different species pair of Drosophila (Kanippayoor 2017). The phenotypes in hybrids of my species pair, however, did not isolate as expected. Most hybrid individuals presented with multiple sperm types or all four sperm types.

4.1 Phenotypic separation

The incomplete separation of sperm type was consistent through each generation of backcross. The majority of samples had more than one morphology of each sperm type, unlike what was found by Kanippayoor (2017). There was a difference in proportion of individuals with all sperm types between the two BC lineages, with BC*pse* having larger proportions of males with all sperm types, whereas BC*per* males had similar proportions between having all sperm phenotypes or missing NE eusperm.

4.1.1 Persistent Sperm Phenotypes

Previous work on the needle-eye phenotype in the Moehring lab focused on two Drosophila species that do not have heteromorphic sperm, unlike *D. pseudoobscura* and

D. persimilis. Hybrid individuals from this study showed a 50/50 separation, where half of the males presented with the needle-eye phenotype, and the other had WT sperm. No clear and consistent separation of phenotype occurred during my examination of hybrids between *D. pseudoobscura* and *D. persimilis* (Table 3, Table 4). Multiple combinations of sperm phenotypes were seen in hybrid males offspring produced from the same hybrid female parent. Very few individuals presented only WT sperm. The high prevalence of the sterile NE sperm could be the result of multiple loci with SNP differences between the two species that results in the improper separation of spermatids during spermatogenesis. When observing NE sperm from dissected testes, they were often still joined in the sperm bundles that form during spermatogenesis; the sperm failed to complete the last step of individualization. In these bundles, the phenotype of the sperm was all one sperm type, as seen in Figure 6c for NE eusperm. The failure to separate into individual sperm could be the result of an error during meiosis II. Evidence from previous work on hybrids formed in another species pair (Kanippayoor 2017: Kanippayoor *et al*. 2020) supports the NE phenotype being the result of two sperm failing to separate, giving half the amount of total sperm. It is possible that eusperm could be vulnerable to spermatogenic errors caused by pertubations in genetic pathways, resulting in the production of NE eusperm. It is also possible that the loci controlling sperm heteromorphism are linked, and therefore not easily separated during introgression.

4.2 Loci analysis

4.2.1 Multiple loci

The lack of separation of sperm type within male testes may indicate multiple loci controlling these sperm traits. If few loci control a sperm phenotype, then they have a

higher probability of separating through recombination over multiple generations. If there are many loci that are potentially each, or in small groups, sufficient to induce the NE phenotype, the likelihood of separating a consistent single locus contributing to the phenotype is low. Having multiple loci controlling these sperm morphs makes it more difficult to separate these sperm types into single phenotypes, even with reducing heterozygosity to ~0.02% through 11 generations of backcrossing.

Even though the two species differ by 2 inversions on chromosome X, and those inversions have been linked to the divergence of the two species (Noor *et al.* 2007), very few SNPs on the X were located within those inversions compared to outside of the inversions. Indeed, over 95% of the SNPs on the X chromosome that were associated with sterility were found outside of the inversions. In contrast, approximately the same number of SNPs were found inside the inversion on the 2nd chromosome as expected based on the inversion's size. Regions on the X chromosome outside of the inversions can therefore be given greater focus to see if they are linked to hybrid male sterility in *D. pseudoobscura* and *D. persimilis*.

4.2.2 Results from PCA analysis

Variability between two data points or samples in a PCA is indicated by distance between data points. The closer two samples are to each other on a PCA, the less variance there is between the data of those two samples. The further away two points are, the greater the variance (Jolliffe and Cadima 2016).

I used whole genome sequencing and sequence comparison to identify SNPS in the backcross hybrids. The samples with *D. persimilis* as the genetic background have

little variance in genotype in terms of what SNPs were found for heteromorphism and needle-eye (Figure 7, Figure 8). This could mean the slight differences in SNPs in these BC*per* males are those controlling the difference in sperm morphology. Samples from this paternal line have similar SNP calls. The samples do not show variance in phenotype in the PCA. Samples from the *D. persimilis* paternal line do not separate based on phenotype in the PCA, therefore, no conclusions can be drawn from the PCA concerning sterility vs fertility associated SNPs or for eusperm vs. parasperm associated SNPs.

The same cannot be said for samples from the *D. pseudoobscura* background. Across all the chromosomes, samples with *D. pseudoobscura* as the background differ from each other in terms of SNPs associated with either NE or WT sperm (Figure 7, Figure 8). Because these samples also did not separate by phenotype, this means that samples from this paternal line were less genotypically similar to each other compared to the *D. persimilis* paternal line.

4.2.3 Future analyses

More detailed analyses of the sequence data are needed to identify candidate genes linked to the SNPs found from the hybrid DNA sequences. Based on the results of this study, the next steps would be to group SNPs based on parental background and phenotype and annotate the SNPs. The comparison SNPs that are different between phenotypes within one background, and then differences between species would allow isolation of candidate regions for sterility based on the location of SNPs that are specific to sterile hybrid male samples and specific to one species.

In order to assess which parts of the genome are involved in hybrid sterility, the SNPs that are only found in the samples that had the NE phenotype need to be identified. To narrow down the initial search, the SNPs that are within the same regions where there were no shared polymorphisms between *D. persimilis* and *D. pseudoobscura* (Machado *et al.* 2007) should be focused on. SNPs can then be assessed for whether they are in coding or non-coding regions. Genes for male hybrid sterility are likely those involved in spermatogenesis and testicular development due to the association of male sterility and the NE sperm phenotype with reduced testes size (Kanippayoor 2017; Kanippayoor *et al.* 2020). Due to sperm production still occurring in sterile males, the gene(s) implicated would likely be involved in the meiosis stages of spermatogenesis. Further analysis of the SNPs within the inverted regions (Table 5) could be done by separating the SNPs within those regions based on phenotype and parental origin and assessing their location relative to the QTL markers used by Machado *et al.* (2007). Finally, an analysis of the SNPs located within known sterility genes, such as *overdrive* (Orr and Irving 2001; Orr and Irving 2005; Phadnis and Orr 2009; Phadnis 2011), will need to be conducted.

4.3 Limitations

4.3.1 Species pair with needle-eye but no heteromorphism

Another species pair that shows the NE sperm phenotype in hybrid males is *Drosophila mojavensis* and *D. arizonae* crosses and *D. mauritiana* and *D. simulans* crosses (Kanippayoor 2017). Both of these species pairs do not present with sperm heteromorphism, and only makes a single, fertilizing sperm morph (Kanippayoor 2017). Using these species pair to study NE sperm would allow for a focused selection of a sterility phenotype during introgression. Whole genome sequencing and SNP analysis of

male hybrids after 10 generations of backcrossing could be done for males with NE sperm and males with WT sperm. These SNPs could be compared with those found in this study as well as those found by Kanippayoor (2017) for hybrids of *D. simulans* and *D. mauritiana*.

4.3.2 Choice of species

Drosophila persimilis and *D. pseudoobscura* present with both sperm heteromorphism and the NE phenotype in hybrid males. Attempting to isolate a single sperm phenotype (example: NE eusperm) was not successful. If sterility was controlled by a single locus that affects both eusperm and parasperm, and one were to focus on singling out wild type individuals from NE individuals, the process would have resulted in a 50/50 separation, where half the males have WT eusperm and parasperm, and the other half have NE eusperm and NE parasperm. Due to the lack of hybrid males who were WT, it is unlikely that only one locus controls sterility within this species pair.

There are other species of Drosophila with phenotypic tools that could aid in phenotype scoring, such as sperm with GFP-fluorescing heads. Two such species are *D. simulans* and *D. mauritiana*, both of which have a strain containing GFP-tagged protamine B, which is a protein that replaces histones in DNA packing within sperm heads (Kanippayoor and Moehring 2012). Backcrossing with these two species using the GFP-tagged sperm showed whether the NE phenotype is affected by protamine B (Kanippayoor *et al*. 2020). Hybrid sperm heads in this case increased in length as the male aged, whereas pure species male sperm heads do not change in length over time. Even the shorter sperm heads found in young hybrid males is longer than pure species

sperm heads. These hybrid males expressed protamines from both parents (Kanippayoor *et al*. 2020).

4.3.3 Use of CRISPR/Cas9 in Sterility studies

CRISPR/Cas9 has been used to study sterility in mice through targeted knockout of testis specific genes (Lu *et al.* 2019). After candidate genes have been identified, a similar method could be used to study the NE phenotype or sperm heteromorphism in *D. persimilis* and *D. pseudoobscura*. A study using *D. suzukii* was able to use CRISPR/Cas9 to improve on a sterile insect technique for the invasive species by using a tissue specific promoter implicated in spermatogenesis (Ahmed *et al.* 2019). Through the use of CRISPR/cas9, candidate genes for sperm sterility and sperm heteromorphism could be found through targeted knockout of spermatogenesis genes by using a mixed model of the mouse and Drosophila studies. In the case of sperm sterility, the knockout could be performed in a parental species instead of a hybrid. Knock outs of the spermatogenesis genes would be done until the presence of the NE phenotype occurs. Based on the evidence from the current study, this may involve the knockout of multiple spermatogenesis genes within one individual. Because the NE phenotype could be cause by multiple loci, altering different loci in separate *D. pseudoobscura* males, and then crossing these males to contain different combinations of the *D. persimilis* allele would allow us to see which combination of alleles result in NE phenotype. The same could be done for sperm heteromorphism, except the goal would be to eliminate either eusperm or parasperm through the knockout or by using CRISPR to edit the spermatogenesis genes at multiple loci.

4.4 Conclusions

Genes responsible for hybrid sterility remain elusive for most species. Studying a sterility phenotype can encounter challenges, such as the inability to create a stable ‘sterility line’ or inability to separate sterility phenotypes. There appear to be many loci controlling NE sperm in *Drosophila pseudoobscura/D. persimilis* hybrids, causing a lack of separation or disappearance of the NE sperm phenotype in introgression male hybrids.

References

Ahmed, H.M.M., Hildebrand, L., and Wimmer, E.A. (2019). Improvement and use of CRISPR/Cas9 to engineer a sperm-making strain for the invasive fruit pest *Drosophila suzukii*. BMC Biotechnology *19*: 10.1186/s12896-019-0588-5

Alpern, J.H.M. (2013). Reproductive biology and speciation in *Drosophila pseudoobscura*. University of Western Ontario Electronic Thesis and Dissertation Repository. 1731 http://ir.lib.uwo.ca/etd/1731

Alpern, J.H.M., Asselin, M.M., and Moehring, A.J. (2019). Identification of a novel sperm class and its role in fertilization in Drosophila. Journal of Evolutionary Biology *32*: 259-266.

Ayala, F.J., Tracey, M.L., Hedgecock, D., and Richmond, R.C. (1974). Genetic differentiation during the speciation process in Drosophila. Evolution *28*: 576-592.

Bateson, W. (1909). Heredity and variation in modern lights. In: Seward, A.C. editor. Darwin and Modern Science. Cambridge: Cambridge University Press. p. 85-101.

Bayes, J.J., and Malik, H.S. (2009). Altered heterochromatic binding by a hybrid sterility protein in Drosophila sibling species. Science *326*: 1538-1541.

Brideau, N.J., Flores, H.A., Wang, J., Maheshwari, S., Wang, X., and Barbash, D.A. (2006). Two Dobzhansky-Muller genes interact to cause hybrid lethality in Drosophila. Science *314*: 1292-1295.

Civetta, A. (2016). Misregulation of gene expression and sterility in interspecies hybrids: casual links and alternative hypothesis. Journal of Molecular Evolution *82*:176-182.

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

Coyne, J.A. (1985). The genetic basis of Haldane's rule. Nature *314:* 736-738.

Coyne, J.A. (1992). Genetics and speciation. Nature *355*: 511-515.

Coyne, J.A., and Orr, H.A. (1989). Patterns of speciation in Drosophila. Evolution *43*: 362-381.

Coyne, J.A., and Orr, H.A. (1997). "Patterns of speciation in Drosophila" revisited. Evolution *51*: 295-303.

Davis, B.W., Seabury, C.M., Brashear, W.A., Li, G., Roelke-Parker, M., and Murphy, W.J. (2015). Mechanisms underlying mammalian hybrid sterility in two feline interspecies models. Molecular Biology and Evolution *32*: 2534-2546.

Dobzanhsky, T.G. (1934). Studies on hybrid sterility. I. Spermatogenesis in pure and hybrid *Drosophila pseudoobscura*. Zeitschrift für Zellforschung und Mikroskopische Anatomie *21*: 169-223.

Dobzanhsky, T.G. (1937). Genetics and the origin of species. New York: Columbia University Press.

Dobzanhsky, T.G. (1973). Is there gene exchange between *Drosophila pseudoobsura* and *Drosophila persimilis* in their natural habitats? The American Naturalist *107*: 312-314.

Ehrman, L. (2963). The transitional races of *Drosophila paulistorum*: a study of hybrid sterility. PNAS *48*: 157-159.

Erdelyi, M., and Szabad, J. (1989). Isolation and characterization of dominant female sterile mutations of *Drosophila melanogaster*. I. Mutations on the third chromosome. Genetics *122*: 111-127.

Fuller, M.T. (1993). Spermatogenesis. In: Bate, M. and Arias, A.M., editors. The development of *Drosophila melanogaster,* Vol I. New York: Cold Spring Harbor Laboratory Press. p. 71-147.

Good, J.M., Dean, M.D. and Nachman, M.W. (2008). A complex genetic basis to X-linked hybrid male sterility between two species of house mice. Genetics *179*: 2213-2228.

Hahn, M.W., Han, M.V., and Han, S.G. (2007). Gene family evolution across 12 Drosophila genomes. Plos Genetics *3*: 2135-2146.

Hales, K.G., Korey, C.A., Larracuente, A.M., and Roberts, D.M. (2015). Genetics on the fly: A primer on the Drosophila model system. Genetics *201*: 815-842.

Haldane, J.B. (1922). Sex-ratio and unidirectional sterility in hybrid animals. Journal of Genetics *12*: 101-109.

Harrison, R.G., and Larson, E.L. (2014). Hybridization, introgression, and the nature of species boundaries. Journal of Heredity *105*: 795-809.

Holman, L., Freckleton, R.P., and Snook, R.R. (2008). What use is an infertile sperm? A comparative study of sperm-heteromorphic Drosophila. Evolution *62*:374-385.

Holman, L., and Snook, R.R. (2008). A sterile sperm caste protects brother fertile sperm from female-mediated death in *Drosophila pseudoobscura*. Current Biology *18*:292-296.

Huang, J., Zhao, Y., Bai, d., Shiraigol, W., Bi, B., Yang, L., Wu, J., Bao, W., Ren, X., Jin, B., Zhao, Q., Li, A, Bao, S., Bao, W., Xing, Z., An, A., Gao, Y., Wei, R., Bao, Y., Bao, T., Han, H., Bai, H., Bao, Y., Zhang, Y., Daidiikhuu, D., Zhao, W., Liu, S., Ding, J., Ye, W., Ding, F., Sun, Z., Shi, Y., Zhang, Y., Meng, H., and Dugarjaviin, M. (2015). Donkey genome and insight into the imprinting of fast karyotype evolution. Scientific Reports *5*: 14106

Hutter, P., Roote, J., and Ashburner, M. (1990). A genetic basis for the inviability of hybrids between sibling species of Drosophila. Genetics *124*: 909-920.

Jezovit, J.A., Levine, J.L., and Schneider, J. (2017). Phylogeny, environment and sexual communication across the Drosophila genus. Journal of Experimental Biology *220*: 42-52.

Jolliffe, I.T., and Cadima, J. (2016). Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A *374*: 20150202.

Joly, D., and Lachaise, D. (1994), Polymorphism in the sperm heteromorphic species of the *Drosophila obscura* group. Journal of Insect Physiology *40*: 933-938.

Kanippayoor, R.L. (2017). Cellular and molecular analysis of interspecies sterile male hybrids in *Drosophila*. University of Western Ontario Electronic Thesis and Dissertation Repository. 4647. http://ir.lib.uwo.ca/etd/4647

Kanippayoor, R.L., Alpern, J.H.M., and Moehring, A.J. (2013). Protamines and spermatogenesis in Drosophila and *Homo sapiens*: A comparative analysis. Spermatogenesis e24376.

Kanippayoor, R.L., Alpern, J.H.M., and Moehring, A.J. (2020). A common suite of cellular abnormalities and spermatogenetic errors in sterile hybrid males in Drosophila. Proceedings of the Royal Society B *287:* 20192291.

Kanippayoor, R.L., and Moehring, A.J. (2012). Allelic expression of Drosophila protamines during spermatogenesis. International Journal of Evolutionary Biology *2012*: doi:10.1155/2012/947381

Koevoets, T., and Beukeboom, L.W. (2009). Genetics of postzygotic isolation and Haldane's rule in haplodiploids. Heredity *102*: 16-23.

Koopman, K.F. (1950). Natural selection for reproductive isolation between *Drosophila pseudoobscura* and *Drosophila persimilis*. Evolution *4*: 135-148.

Lachaise, D., David, J.R., Lemeunier, F., Tsacas, L., and Ashburner, M. (1986). The reproductive relationships of *Drosophila sechellia* with *D. mauritiana*, *D. simulans*, and *D. melanogaster* from the afrotropical region. Evolution *40:*262-271.

Lai-Fook, J. (1982). Structural comparison between eupyrene and apyrene spermiogenesis in CaZpodes ethlius (Hesperiidae, Lepidoptera). Canadian Journal of Zoology *60*: 1216-1230.

Larracuente, A.M., Noor, M.A.F., and Clark, A.G. (2010). Translocation of Y-linked genes to the dot chromosome in *Drosophila pseudoobscura*. Molecular Biology and Evolution *27*: 1612-1620.

Laurie, C.C. (1997). The weaker sex is heterogametic: 75 years of Haldane's Rule. Genetics *147*: 937-951.

Lever, J., Krzywinski, M., and Altman, N. (2017). Principal component analysis. Nature Methods *14*: 641-642.

Li, H., and Durbin, R. (2009). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*: 589-595.

Liang, J., and Sharakhov, I.V. (2019). Premeiotic and meiotic failures lead to hybrid male sterility in *Anopheles gambiae* complex. Proceedings of the Royal Society B *286*: 20191080.

Lu, X., Shapiro, J.A., Ting, C., Li, Y., Li, C., Xu, J., Huang, H., Cheng, Y., Greenberg, A.J., Li, S., Wu, M., Shen, Y., and Wu, C. (2010). Genome-wide misexpression of

X-linked versus autosomal genes associated with hybrid male sterility. Genome Research *20*: 1097-1102.

Lu, Y., Oura, S., Matsumura, T., Oji, A., Sakurai, N., Fujihara, Y., Shimada, K., Miyata, H., Tobita, T., Noda, T., Castaneda, J. M., Kiyozumi, D., Zhang, Q., Larasati, T., Young, S., Kodani, M., Huddleston, C. A., Robertson, M. J., Coarfa, C., Isotani, A., Aitken, R.J., Okabe, M., Matzuk, M.M., Garcia, T.X., and Ikawa, M. (2019). CRISPR/Cas9-mediated genome editing reveals 30 testis-enriched genes dispensable for male fertility in mice. Biology of Reproduction *101*: 501–511.

Machado, C.A., Haselkorn, T.S., and Noor, M.A.F. (2007). Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in *Drosophila pseudoobscura* and *D. persimilis*. Genetics *175*: 1289-1306.

Mayr, E. (1982). Speciation and macroevolution. Evolution *36*: 1119-1132.

McDaniel, S.T., Willis, J.H., and Shaw, A.J. (2008). The genetic basis of developmental abnormalities in interpopulation hybrids of moss *Ceratodon purpureus*. Genetics *179*: 1425-1435.

McGaugh, S.E., and Noor, MA. (2012). Genomic impact of chromosomal inversions in parapatric Drosophila species. Philosophical Transactions of the Royal Society B *367*: 422-429.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A.)2010). The

genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research *20*: 1297-1303.

Mihola, O., Trachtulec, Z., Vleck, C., Schimenti, J.C., and Forejt, J. (2009). A mouse speciation gene encodes a meiotic histone H3 methyltransferase. Science *323*: 373-375.

Moore, A.J., Bacigalupa, L.D., and Snook, R.R. (2013). Integrated and independent evolution of heteromorphic sperm types. Proceedings: Biological Sciences *280*: 1-5.

Muller, H.J. (1940). Bearing of the Drosophila works on Systematics In: Huxely, J. editor. The New Systematics. Oxford: Oxford University Press. p 185-268.

Muller, H.J. (1942). Isolating mechanisms, evolution, and temperature. Biology Symposium *6*: 71-125.

Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M., Reinert, K.H.J., Remington, K.A., Anson, E.L., Bolanos, R.A., Chou, H.H., Jordan, C.M., Halpern, A.L., Lonardi, S., Beasley, E.M., Brandon, R.C., Chen, L., Dunn, P.J., Lai, Z., Liang, Y., Nusskern, D.R., Zhan, M., Zhang, Q., Zheng, X., Rubin, G.M., Adams, M.D., Venter, J.C. (2000). A whole-genome assembly of Drosophila. Science *287*: 2196-2204.

Nishino, R., Petri, S., Handel, M.A., Kuneida, T., and Fujiware, Y. (2019). Hybrid sterility with meiotic metaphase arrest in intersubspecific mouse crosses. Journal of Heredity *110*: 183-193.

Njogu, M., Ricketts, P., and Klaus, A.V. (2010). Spermatogenic cyst and organ culture in *Drosophila pseudoobscura*. Cell and Tissue Research *341*: 453-464.

Noor, M.A., Garfield, D.A., Schaeffer, S.W., and Machado, C.A. (2007). Divergence between the *Drosophila pseudoobscura* and *D. persimilis* genome sequence in relation to chromosomal inversions. Genetics *177*: 1417-1428.

Noor, M.A., Grams, K.L, Bertucci, L.A., Almendarez, Y., Reiland, J., and Smith, K.R. (2001). The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. Evolution *55*: 512-521.

Orr, H.A. (1993). A mathematical model of Haldane's rule. Evolution *47*: 1606-1611.

Orr, H.A. (1995). The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics *139*: 1805-1813.

Orr, H.A., and Irving, S. (2001). Complex epistasis and the genetic basis of hybrid sterility in the *Drosophila pseudoobscura* Bogota-USA hybridization. Genetics *158*: 1089-1100.

Orr, H.A., and Irving, S. (2005). Segregation distortion in hybrids between the Bogota and USA subspecies of *Drosophila pseudoobscura*. Genetics *169*: 671-682.

Orr, H.A., and Presgraves, D.C. (2000). Speciation by postzygotic isolation: forces, genes and molecules. BioEssays *22*: 1085-1094.

Palopoli, M.F., and Wu, C. (1994). Genetics of hybrid male sterility between *Drosophila* sibling species: a complex web of epistasis is revealed in interspecific studies. Genetics *138*: 329-41.

Pasini, M.E., Redi, C.A., Caviglia, O., and Perotti, M.E. (1996). Ultrastructural and cytochemical analysis of sperm dimorphism in *Drosophila subobscura*. Tissue & Cell *28*: 165-175.

Phadnis, N. (2011). Genetic architecture of male sterility and segregation distortion in *Drosophila pseudoobscura* Bogota-USA hybrids. Genetics *189*: 1001-1009.

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

Pilder, S.H., Olds-Clarke, P., Orth, J.M., Jester, W.F., and Dugan, L. (1997). Hst7: a male sterility mutation perturbing sperm motility, flagellar assembly, and mitochondrial sheath differentiation. Journal of Andrology *18*: 663-671.

Pitnick, S., Hosken, S.J., and Birkhead, T.R. (2008). Sperm morphological diversity. In: Pitnick, S., Hosken, S.J., and Birkhead, T.R., editors. Sperm biology: An evolutionary perspective. Massachusetts: Elsevier. p. 76-78.

Presgraves, D.C. (2010). The molecular evolutionary basis of species formation. Nature Reviews: Genetics *11*: 175-180.

Rizki, M.T.M. (1951). Morphological differences between two sibling species, *Drosophila pseudoobscura* and *Drosophila persimilis*. Proceedings of the National Academy of Sciences *37*: 156-159.

Sambrook, J., Fritch, E.F., and Maniatis, T. (1989). Molecular cloning: A laboratory manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Savory, T.H. (1970). The mule. Scientific American *223*: 102-109.

Snook, R.R. (1997). Is the production of multiple sperm types adaptive? Evolution *51*: 797-808.

Snook, R.R., Markow, T.A., and Karr, T.L. (1994). Function nonequivalence of sperm in *Drosophila pseudoobscura*. Proceedings of the National Academy of Sciences *91*: 11222-11226.

Stevens, N.M. (1905). A study of the germ cells of *Aphis rosæ* and *Aphis ænotheræ*. Journal of Experimental Zoology *2*: 313-333.

Storchova, R., Gregorova, S., Buckiova, D., Kyselova, V., Divina, P., and Forejt, J. (2004). Genetic analysis of X-linked hybrid sterility in the house mouse. Mammalian Genome *15*: 515-524.

Sturtevant, A.H. (1920). Genetic studies on *Drosophila simulans*. I. Introduction. Hybrids with *Drosophila melanogaster*. Genetics *5*: 488-500.

Sun, J., and Spradling, A.C. (2013). Ovulation in *Drosophila* is controlled by secretory cells of the female reproductive tract. eLife 2: e00415.

Thomae, A.W., Schade, G.O.M., Padeken, J., Borath, M., Vetter, I., Kremmer, E., Heun, P., and Imhof, A. (2013). A pair of centromeric proteins mediates reproductive isolation in Drosophila species. Developmental Cell *27*: 412-424.

Till-Bottraud, I., Joly, D., Lachaise, D., and Snook, R.R. (2005). Pollen and sperm heteromorphism: convergence across kingdoms? Journal of Evolutionary Biology *18*: 1-18.

Ting, C., Tsaur, S., Wu, M., Wu, C. (1998). A rapidly evolving homeobox at the site of a hybrid sterility gene. Science *282*: 1501-1504.

Tokuyasu, K.T. (1974). Dynamics of spermiogenesis in *Drosophila melanogaster*. IV. Nuclear Transformation. Journal of Ultrastructure Research *48*: 284–303.

Turelli, M., Barton, N.H., Coyne, J.A. (2001). Theory and speciation. Trends in Ecology and Evolution *16*: 330-343.

Turelli, M., and Orr, H.A. (1995). The dominance theory of Haldane's rule. Genetics *140*: 389-402.

Volpi, S., Bongiorni, S., Fabbretti, F., Wakimoto, B.T., and Prantera, G. (2013). *Drosophila rae1* is required for male meiosis and spermatogenesis. Journal of Cell Science *126*: 3541-3551.

Wang, R.L., Wakeley, J., and Hey, J. (1997). Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relative. Genetics *147*: 1091-1106.

Wold, S., Esbensen, K., and Geladi, P. (1987). Principle component analysis. Chemometrics and Intelligent Laboratory Systems *2*: 37-52.

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

Yuan, X., Zheng, H., Yang, S., Guo, P., Zhang, X, Zhao Q., Ge, W., Li, C., Xi, Y., and Yang, X. (2019). *Drosophila* Pif1A is essential for spermatogenesis and is the homolog of human CCDC157, a gene associated with idiopathic NOA. Cell Death & Disease *10*: 125.

Appendices

Appendix A: Whole genome sequence information and specifics about the DNA sequencing output performed by Genome Quebec Innovation Center. Name, Library Name, and Alias are the labels given to the DNA sample.

Name	Library Name	Alias	Run Concentration (pM)	Run	Quality Offset	Region	Total Region	Number of Reads	Number of Bases	Number of Cycles	Average Quality	% Duplicate
C2_724	C2_724	C2_724	225	1251	33	1	4	55,312,600	16,704,405,200	318	36	29.299
35	35	35	225	1251	33	1	4	74,584,518	22,524,524,436	318	36	29.122
34	34	34	225	1251	33	1	4	63,101,673	19,056,705,246	318	36	21.506
A1_719	A1_719	A1_719	225	1251	33	1	4	28,351,517	8,562,158,134	318	36	25.76
10	10	10	225	1251	33	1	4	73,383,888	22,161,934,176	318	36	20.606
11	11	11	225	1251	33	1	4	79,671,757	24,060,870,614	318	36	28.943
19	19	19	225	1251	33	1	4	62,891,165	18,993,131,830	318	36	22.811
29	29	29	225	1251	33	1	4	76,996,455	23,252,929,410	318	36	25.684
28	28	28	225	1251	33	1	4	75,144,869	22,693,750,438	318	36	25.615
20	20	20	225	1251	33	1	4	77,746,970	23,479,584,940	318	36	23.867
31	31	31	225	1251	33	1	4	81,426,064	24,590,671,328	318	36	21.669
17	17	17	225	1251	33	1	4	72,658,618	21,942,902,636	318	36	22.729
18	18	18	225	1251	33	1	4	65,665,812	19,831,075,224	318	36	28.061
30	30	30	225	1251	33	1	4	90,123,888	27,217,414,176	318	36	21.775
D1_719	D1_719	D1_719	225	1251	33	1	4	81,871,184	24,725,097,568	318	36	25.675
12	12	12	225	1251	33	1	4	48,195,393	14,555,008,686	318	36	20.384
13	13	13	225	1251	33	1	4	71,235,449	21,513,105,598	318	36	20.698
B1_719	B1_719	B1_719	225	1251	33	1	4	65,819,823	19,877,586,546	318	36	23.409
F1_719	F1_719	F1_719	225	1251	33	1	4	75,611,688	22,834,729,776	318	36	20.388
25	25	25	225	1251	33	1	4	96,897,696	29,263,104,192	318	36	22.373

Curriculum Vitae

Name: Alannah Mattice

Post-secondary Education and Degrees:
The University of Western Ontario
London, Ontario, Canada
2013-2017 B.Sc.

Honours and Awards:
Western Scholarship of Distinction
2013

Dean's Honour List
2017

Related Work Experience
Teaching Assistant
The University of Western Ontario
2017-2019

Conference Presentations:

Mattice, A., Karagiannis, J. Genetic Interaction Analysis of Site-Directed Mutants of the *Schizosaccharomyes pombe tsc1* gene. Ontario Biology Day. Laurentian University, Sarnia, Ontario. March 2017. (Poster Presentation)

Mattice, A., Moehring, A. Genetic basis of hybrid sterility between *Drosophila pseudoobscura* and *D. persimilis*. OE3C. Western Ontario University, London, ON, Canada. May 2018. (Oral Presentation)

Mattice, A., Moehring, A. Genetic basis of hybrid sterility between *Drosophila pseudoobscura* and *D. persimilis*. Biology Graduate Research Forum. Western Ontario University, London, ON, Canada. October 2018. (Oral Presentation)

Mattice, A., Moehring, A. Genetic basis of hybrid sterility between *Drosophila pseudoobscura* and *D. persimilis*. Evolution 2019. Providence, RI, United States. June 2019. (Oral Presentation)

Publications:

Rayan, A., Faller, A., Chevalier, R., Mattice, A., Karagiannis, J. (2018) Using genetic buffering relationships identified in fission yeast to reveal susceptibilities in cells lacking hamartin or tuberin function. Biology Open *7*: 1-10.