UNIVERSITY OF CALIFORNIA
RIVERSIDE

Finding Genomic Hybrid Incompatibilities Between Members of the *Anopheles gambiae* Complex

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Genetics, Genomics and Bioinformatics

by

Raissa Genevieve-Green Kay

June 2017

Dissertation Committee:
Dr. Bradley White, Chairperson
Dr. Anupama Dahanukar
Dr. Jason Stajich

The Dissertation of Raissa Genevieve-Green Kay is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

I would like to thank all those who encouraged, supported, and assisted me through my time in graduate school. First I would like to thank Dr. Bradley White, my advisor, for guiding me through my studies and helping me find my own path to success. I would also like to thank my dissertation committee: Dr. Anupama Dahanukar and Jason Stajich for offering their advice and support. Dr. Dahanukar and Dr. Stajich, along with Dr. Nunney, Dr. Wessler, and Dr. Stouthammer I would like to thank for being part of my qualifying exam committee. For giving me the opportunity to rotate and learn from their labs I would like to thank Dr. Martins-Green and Dr. Nugent.

I would also like to acknowledge my lab-mates, colleagues and friends that offered advice and support, without them I wouldn't have made it through. A special thanks to my fellow graduate students Eric Smith and James Ricci for offering a helping hand and a critical ear. Thank you to Dr. David Turissini for his patience in teaching me programming. A special thanks to Michelle Bui for always going above and beyond to help me with lab work. And to Stephanie Gamez, Colince Kamdem, Caroline Fouet, Ming Li, and Sean Prager, thank you for making the White Lab a great place to research. I would also like to acknowledge Bryan Cassone who was an invaluable collaborator.

For always being helpful and giving excellent advice I would like to thank the core facility staff, especially Dr. Glenn Hicks and Clay Clark. For making sure I always signed up for credits in time and got all the appropriate documents

signed I’d like to thank Deidra Kornfeld and Julio Sosa, my graduate student service advisors.

Last, but not least, I want to thank my friends and family. Without their support and love I could not have made it this far. Thank you to my best friend Kayla Wong, who always knew I could do it. Thank you to my Parents, Grant and Heather Green, and to my sisters, Ariana Green and Cambia Rome (and their husbands) for willingly listening to my struggles and successes. And finally, thank you to my husband, Jarren Kay, who pushed me to keep going and keep trying even when I didn’t know I could.

ABSTRACT OF THE DISSERTATION

Finding Genomic Hybrid Incompatibilities Between Members of the *Anopheles gambiae* Complex

by

Raissa Genevieve-Green Kay

Doctor of Philosophy, Graduate Program in Genetics, Genomics, and Bioinformatics
University of California, Riverside, June 2017
Dr. Bradley White, Chairperson

The Batson-Dobzhasnky-Muller model predicts that as species diverge over time they will accumulate genetic differences, which may be incompatible with each other when combined into the same genetic background. Despite the central importance of this process to diversification and speciation, identification of genes causing hybrid dysfunction have been limited primarily to model species such as mice and *Drosophila*. *Anopheles gambiae*, the principal mosquito vector of malaria in Africa, belongs to a complex of at least nine isomorphic species. In accordance with Haldane's rule for speciation, reciprocal crosses between members of the *Anopheles gambiae* complex resulted in completely sterile males, a fact that was used to classify different mosquito populations into separate species. Some of these hybrid crosses were also found to have sex-ratio distortion implying some form of female hybrid inviability, refuting Haldane's rule. Those females that were viable were also found to be fertile. Using reciprocal crosses between *An. coluzzii* and *An. merus* I found that when F1 hybrid females were backcrossed, the resulting male progeny displayed a range of phenotypes from completely sterile to completely fertile providing the foundation for quantitative trait locus (QTL) mapping. We performed reciprocal backcrosses, phenotyped ~2500 males from those backcrosses, and then

genotyped each individual male at an average of >10,000 markers across the genome using a reduced representation sequencing approach. Using RNA-seq to measure gene expression differences in F1 hybrid testes, I also found evidence of disruption of sex chromosome inactivation in the *An. coluzzii*♀ by *An. merus*♂ crosses, an under-appreciated mechanism of Haldane's rule.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

CHAPTER II

LIST OF TABLES

CHAPTER I

CHAPTER II

CHAPTER III

Introduction

Introduction to Speciation

Evolutionary processes have given rise to the diversity of organisms on Earth. Through billions of years the process of evolution has influenced changes that manifested in life forms from simple single-celled organisms to complex multi-cellular organisms. Speciation, or the process of one population dividing into two or more separate species that are distinct form one another. Speciation results in organisms differently adapted to their environment, and this, in turn, increases the biodiversity of Earth. This biodiversity is not only important for the health of our planet, but human health in general. Biodiversity has given humans the variety of plants and animals we eat, the animals that can recycle waste or remove pest species, and some animals that lead to unexpected medical advances (gila monster and wasp paper). Even given the direct importance of speciation to humans and our need to classify different groups of living things to better our understanding of them, the definition of what a species is can vary.

Since before Darwin's "On the Origin of Species", biologists have attempted to define what makes up a species (Coyne and Orr, 2004; Linnaeus, 1735; Ray, 1686). Were species an actual phenomenon or were they abstract concepts with no real definition except for that which humans give them? Darwin supported the idea that species were not real biological phenomena, but only human concepts, however subsequent research and experimentally determined definitions of species have been developed

(Coyne and Orr, 2004; Darwin, 1859). One of these phenomena is a lack of interbreeding occurring between populations when they are undergoing speciation. This is a primary premise in one definition of species: the "Biological Species Concept" (Mayr, 1963). This theory states that: "Species are groups of interbreeding natural populations that are reproductively isolated from other such groups" (Mayr, 1963). This isolation can be either pre-zygotic or post-zygotic. In sexually reproducing species, pre-zygotic isolation is either behavioral (e.g. courtship differences), ecological (e.g. populations are divided by physical barriers or only occur in specific niches), or mechanical (e.g. the organs or other mechanisms used for copulation are not compatible) (Coyne and Orr, 2004). Post-zygotic isolation occurs when a hybrid zygote is formed, but that hybrid either does not make it to sexual maturity or cannot reproduce and is sterile. Post-zygotic isolation is often the first measure that two populations are becoming different species as it is easier to measure inviability and/or sterility than to measure behavioral differences.

Rules of Speciation

In 1922, J.B.S. Haldane first described an interesting phenomenon: in populations that were thought to be speciating, the first sign of post-zygotic isolation in the form of inviability or sterility is seen in the heterogametic sex only (Haldane, 1922). Haldane's rule holds true throughout a diverse range of animals where either the male or female is the heterogametic sex including:

Mammalia (mammals), Diptera (flies, mosquitoes), Orthoptera (grasshoppers and crickets), Teleostei (bony fish), Amphibia (frogs and newts), Aves (birds), Lepidoptera (butterflies), and Reptilia (reptiles) (Coyne, 1992; Coyne and Orr, 2004; Laurie, 1997; Orr and Presgraves, 2000; Presgraves, 2010; Schilthuizen et al., 2011; Volff, 2005). Schilthuizen et al. (2011) compiled a list of the cases in animals that both conform or don't conform to Haldane's rule by examining hybrid sterility or hybrid inviability found that 213 out of 223 examples of hybrid sterility and 381 out of 452 of hybrid inviability adhered to Haldane's rule. While Haldane's rule is upheld throughout the animal kingdom, very little is known about the biological mechanisms that make it true for so many diverse organisms.

Along with Haldane's rule, there are two other findings about post-zygotic reproductive isolation that also occur frequently: the Large-X effect and asymmetric hybrid incompatibility ("Darwin's corollary") (Bolnick et al., 2008; Wu and Davis, 1993). The large X-effect is the observation that the X chromosome plays a prominent role in hybrid incompatibilities (HI) (Coyne, 1985; Coyne and Orr, 1989; Dobzhansky, 1970; Zouros et al., 1988). The large X-effect has been confirmed by studies that show introgression of the X chromosome in *Drosophila* hybrids results in more sterile males than introgression into autosomes (Coyne and Orr, 1997; Curtis, 1982; Dobzhansky, 1936; Masly and Presgraves, 2007). Asymmetric hybrid inviability occurs when hybrid reciprocal crosses vary in the proportion of inviable offspring depending on which species is the mother or father (Bolnick and Near, 2005; Bolnick et al., 2008; Tiffin et al., 2001; Turelli and Moyle,

2006). The frequent occurrence of asymmetric hybrid inviability has lead Bolnick et al. (Bolnick et al., 2008) to place it with Haldane's rule and the large X-effect as a rule of speciation.

Mechanism of Hybrid Incompatibilities

The mode-of-action of how HI occurs was hypothesized in 1909 by William Bateson, and in the late 1930's independently described by Theodosius Dobzhansky and Herman Muller. The models of hybrid inviability and sterility were thought to arise through the accumulation of genomic incompatibilities, also called Bateson-Dobzahnsky-Muller (BDM) incompatibilities (Bateson, 1909; Dobzhansky, 1937; Muller, 1942). The theory proposed that pre-zygotic isolation, geographical or otherwise, would occur to separate two populations of the same species. Those two populations evolve due to selection and/or drift and when members of the two populations come into contact and hybridize the differently evolved loci would no longer be able to interact properly causing post-zygotic incompatibilities in the hybrid. BDM incompatibilities are related back to Haldane's rule through theories which try to explain whether incompatibility is caused by few genes of large effect, as in the *Prdm9* gene in mammals, or many genes of smaller effect, such as the genes found around the *OdsH* locus in *Drosophila* (Oliver et al., 2009; Perez and Wu, 1995). There is a debate as to whether incompatibility genes are responsible for speciation, or are just a side effect of positive selection at a linked locus (Wright et al., 2013). This theory of sterility

genes as speciation genes is incredibly hard to prove without studying actively diverging populations, and even then the route of speciation may be different for different populations. Sterility genes are normally fast evolving genes, meaning positive selection is acting or has acted on these genes (Wang et al., 2011). The fast evolving nature of these genes contradicts Dobzhansky's original theory that these incompatibilities arose over long periods of evolutionary time and more likely to be polygenic traits. Goldschmidt in 1940, and later Stanley in 1979 (Goldschmidt, 1940; Stanley, 1979), disagreed and found that reproductive isolation could happen in a relatively short period of time and could involve very few genes. As there are many evolutionary paths to reproductive isolation no unifying theory can yet explain its occurrence across all different forms of life. Of the theories that attempt to explain reproductive isolation and the rules of speciation, dominance theory, the faster-male theory, and the faster-X theory are more supported than the rest.

Hybrid Incompatibility Theories

The first theory suggested to explain Haldane’s rule was posited by Muller in 1942. The “dominance theory” assumed that sterility or inviability were caused by recessive alleles on the homogametic sex chromosome. Assuming an XX-XY system where the female is homogametic, this would mean that the female hybrid would have another X to compensate or dominate the incompatible allele, but male hybrids would express the incompatible allele causing sterility or inviability (Muller, 1942). This theory has support based on

observations of inviability in *Drosophila* species where female F1 hybrids forced to be homozygous for one species's X were inviable just like the F1 males (Orr, 1993; Wu and Davis, 1993). The dominance theory also applied to sterility where homozygous X introgression into F1 hybrids have been found to result in sterility (Hollocher and Wu, 1996; True et al., 1996). The position of these recessive incompatible alleles on the X chromosome can also account the large-X effect (Coyne and Charlesworth, 1989; Turelli and Orr, 1995; Wu and Davis, 1993).

This large-X effect is caused by recessive alleles that can only be expressed and selected for in the heterogametic sex thus increasing the rate of evolution for those genes involved (Meisel and Connallon, 2013). This increased rate of evolution generates a snowball-like effect after the initial speciation event that greatly increases the rate of incompatibilities accumulated between two species (Oliver et al., 2009; Städler et al., 2012).

The faster-male theory posits that Haldane's rule is caused by faster evolution of male incompatibility factors than female (Wu and Davis, 1993; Wu et al., 1996). This could be because spermatogenesis has been found to be much more prone to dysfunction than oogenesis in multiple studies of *Drosophila* hybrids (Malone and Michalak, 2008; Moran et al., 2017; Wu and Davis, 1993). Another explanation of faster-male evolution could be due to sexual selection (Coyne and Orr, 1997; Wu and Davis, 1993). In *Drosophila* hybrids alleles causing male sterility evolved faster than those that would cause female sterility providing supporting evidence for faster-male evolution

(Hollocher and Wu, 1996; True et al., 1996). One weakness of the faster-male theory is that Haldane's rule holds true even in species where the female is the heterogametic sex, as is the case in butterflies and birds. A second weakness in the theory is that it assumes that male genes are evolving faster due to selection and ignores the possibility that female genes may be co-evolving with the males' and so the rates of evolution will not differ between sexes (Coyne and Orr, 1997).

An additional popular theory as to how Haldane's rule functions is the "faster-X theory". First proposed by Brian Charlesworth et al. (Charlesworth et al., 1987), the faster-X theory finds that loci on the X chromosome evolve faster than autosomal loci as long as they are partially or fully recessive in the parental species. These partially or fully recessive loci could confer an evolutionary advantage that can only be selected for in the heterogametic sex until those loci become more prevalent in the population. This theory is different from dominance theory as it examines dominance in the parental species, whereas dominance theory is about dominance in the hybrid only (Coyne and Orr, 1997). Recent genome-wide analysis searching for faster-X evolution has produced mixed results. Some species are consistent with the faster-X model, while others have no significant difference between X and autosomal evolution (Ávila et al., 2014; Meisel and Connallon, 2013). Faster X chromosome evolution can account for both Haldane's rule and the large-X effect.

Causes of Hybrid Incompatibilities

While some BDM incompatibilities cause HI through interacting pairs of genes in the hybrid, others are thought to arise through epigenetic interactions between parental factors and hybrid genomes, especially in cases of asymmetric HI. While only a few genes have been strongly implicated in HI, the specific action of these genes that may cause HI is still relatively unknown. The epigenetic factors that might be involved in HI are also hypothetical, though some evidence is available for hypotheses that include genetic imprinting, maternal effects, and cytoplasmic incompatibilities. Together both gene interactions and epigenetic factors have been found to result in HI.

The genes causing partial or complete HI have been found mainly in *Drosophila*, but recent studies in mice have also found a hybrid sterility causing gene (Mihola et al., 2009; Nosil and Schluter, 2011). In *Drosophila* genes have been found to influence both hybrid sterility and hybrid inviability. *Odysseus* (*OdsH*) was the first HI gene identified. It is a gene that usually functions to slightly enhance sperm production in young *Drosophila* males but can cause sterility in hybrid males (Sun et al., 2004). Additional HI genes found in *Drosophila* include hybrid male rescue (*Hmr*), lethal hybrid rescue (*Lhr*), nucleoporin 96 and nucleoporin 160 which cause inviability rather than sterility (Barbash et al., 2003, 2004; Brideau et al., 2006; Nosil and Schluter, 2011; Presgraves and et al., 2003; Tang and Presgraves, 2015). The first mammalian HI gene, Prdm9*,* was identified in mice (Mihola et al., 2009) and

encodes a DNA-binding protein that “Directs the positioning of double-stranded breaks” that start recombination during meiosis (Davies et al., 2016). The discovery of these HI genes and research into their normal functions provides a window into the molecular mechanisms of HI and can suggest additional candidate genes with similar functions.

Gene-gene interactions are not the only causes of HI. Additional factors can work epigenetically on the hybrid’s genome while it is still developing in the mother. Imprinting, or the mono-allelic parental-dependent expression of certain genes, has been implicated in asymmetric HI a number of times in both plants and animals (Ishikawa and Kinoshita, 2009; Johnson, 2010; Maheshwari and Barbash, 2011; Wu and Ting, 2004), and can cause gross phenotypic differences along with HI. The directionality of these differences has been well described in horse and donkey crosses, where if the female parent is a horse and the male a donkey you get a mule, whereas if the donkey is the female and the horse the male you get a hinny (Allen, 1969), and in crosses between two mouse species *P.polionotus* and *P.maniculatus*: when *P.maniculatus* is the female, the hybrid offspring are 40% smaller than either parent, and when *P.polionotus* is the female the offspring and placenta are oversized and often die before birth (Duselis and Vrana, 2010; Laschiavo et al., 2007; Vrana et al., 1998). However, imprinting can act like a selfish element by biasing the expression of the allele from one sex over the allele of the other sex. Moore and Haig in their 1991 paper proposed that competition over maternally provided resources between

offspring fathered by different males can favor imprinted expression (Moore and Haig, 1991).

The other likely mechanism for asymmetric HI is the incompatibilities between maternal effects and the hybrid genome. Maternal factors control the development of a growing embryo, and also suppress elements of the genome that may disrupt the growth process, until the “maternal-zygotic transition” where the zygote's genome takes over in these processes (Wang and Dey, 2006). Sawamura and Yamamoto in 1993 found a maternal acting gene on chromosome 2 in *D. simulans* that was incompatible with a zygotic acting gene on the X chromosome of *D. melanogaster*. Female hybrids with *D. simulans* mothers are inviable giving all male F1s. In the reciprocal cross hybrid males were inviable giving all female F1s (Sawamura, 1996; Sawamura et al., 1993).

Tools for Studying Hybrid Incompatibilities

Approaches to studying HI have evolved with new technologies. Initial work focused on theoretical studies and simple crossing experiments that produced Haldane’s Rule and BDM incompatibilities (Bateson, 1909; Dobzhansky, 1937; Haldane, 1922; Mayr, 1963; Muller, 1942). As methods to manipulate *Drosophila* improved, it became a vital tool with the ability to track rearrangements of known loci that cause visible phenotypic differences. Using these markers, it was possible to identify loci near sterility and inviability loci by careful introgression of genetic material of one species into another species genome (Perez et al., 1993). With the advent of genome sequencing

and the ability to obtain full genome sequences of any organism, these introgression studies could be used for other taxa, outside of *Drosophila*, as phenotypic markers were replaced with identifying the new sequences transferred as part of an introgression that caused the HI. However, the volume and resolution of information from genome sequencing requires a new way to apply significance to these introgression, for this we can use Quantitative Trait Loci (QTL) mapping (Seaton et al., 2002).

QTL approaches to map loci involved in speciation work by scoring sterility or a measurable phenotype related to sterility (sperm morphology, number of offspring, etc.) of backcrossed or introgressed individuals as a quantitative trait and identifying genotypes that are consistent with the direction of the phenotype (Seaton et al., 2002). The genotype of an individual can be determined in a variety of ways, including traditional marker association (as with *Drosophila*), genotyping by sequencing that entails sequencing the same small part of the genome for each individual, and Genome Wide Association Studies (GWAS) which require full genome sequences of each individual in the study (Davis et al., 2015; Moyle, 2006; Perez et al., 1993). While GWAS studies are usually the best approach for looking at the rate of HI in wild populations, traditional QTL studies are still very powerful if using lab-reared colonies (Slotman et al., 2004).

Additionally, the genotype can be values gleaned from the transcriptome which may be influenced by epigenetic factors that also underlie HI differences. Starting with microarrays and now current technology

RNA-seq, the transcriptome of hybrids can show patterns of gene expression either similar or different from the parental species. The expression of a gene resemble one parent, may be the average of both parents, or can be completely outside the range of parental expression (Bi et al., 2014; Stupar et al., 2007, 2008). It is thought that those genes whose expression falls outside of normal parental expression are either direct cause or side-effect of HI (Renaut et al., 2009; Tulchinsky et al., 2014; Zufall and Rausher, 2004). By combining multiple tools, strong candidate genes underlying HI can be identified.

Anopheles gambiae complex

My dissertation examines members of the *Anopheles gambiae* complex that are thought to have separated around 1.8 million years ago, ideal evolutionary divergence to study speciation and sterility genes (Fontaine et al., 2015). Mosquitoes are disease vectors and have been intensely studied with hopes of breaking the cycle of infectious diseases. All but one of the six members of the *Anopheles gambiae* complex are vectors of malaria and contribute to over 214 million cases estimated worldwide in 2015, with 188 million cases just in Africa where the species of this species complex resides (2015). Of those 188 million cases in Africa, the 295,000 cases resulted in death. Though this number has decreased from 670,000 deaths in 2005, the fight against malaria is still an on-going one (2005). Any study into the biology of these disease

vectors could have a major impact on those living in areas where the vector and disease are present, and suggest better control strategies.

Historical studies of this species complex have been re-assessed in this dissertation. The two species that were used for more intensive analysis, *An. coluzzii* and *An. merus*, are suitable species for finding incompatibility genes because they also have the highest sequence similarity on the X chromosome of any members in the *Anopheles gambiae* complex (White et al., 2011). *An. coluzzii* and *An.merus* are also important to public health as they both transmit the human malaria parasite *P. falciparum.* However, *An. coluzzii* is considered a more successful vector due to its extensive geographical range and increased tendency to bite humans. Finding sterility factors in the *Anopheles* genus gives us the potential to develop more effective control strategies.

Though sterile insect technology (SIT) control strategies usually focus on sterilization of the male mosquito there is little biological information about the males mainly because they are the non-disease vectoring sex. In fact only a few genes have been found that are male specific (Hall et al., 2013, 2016; Krzywinska et al., 2016; Slotman et al., 2004).

Dissertation Objectives and Aims

Studying how speciation and reproductive isolation comes about is both complicated and important for understanding evolutionary processes and

the trajectory that created their modern genotypes. The *An. gamibae* complex is recently diverged, has heteromorphic sex chromosomes, and includes vectors of deadly diseases, making it an important non-model organism to add to the growing literature on HI. Hybrid crosses between mosquitoes of this complex tend to follow all three rules of speciation, though the few exceptions may lead to explanations for when Haldane's rule fails. Each chapter of this dissertation uses one of the most common tools for uncovering and explaining HI. Chapter I uses simple crossing experiments that give an interesting picture of Darwin's corollary and some small rejection of Haldane's rule. We see that the asymmetric sex-ratio distortion that was very prevalent in past hybridizations is now only significant in a rare few crosses. Chapter II utilizes reciprocal crosses and backcrosses between *An. coluzzii* and *An. merus* to find Quantitative Trait Loci (QTL) related to hybrid sterility. The backcrosses between *An. coluzzii* and *An. merus* form males that lie on a spectrum from fertile to sterile giving us a quantitative trait to use. Using the same reciprocal cross that Chapter II used, Chapter III looks at the expression differences between the F1 hybrids and parental testes. Each of these chapters not only adds new insight and information to HI studies, but also adds to our understanding of male mosquito genetics and mosquito speciation in general.

References

Allen, W.R. (1969). Factors influencing pregnant mare serum gonadotrophin production. Nature *223*, 64–65.

Ávila, V., Marion de Procé, S., Campos, J.L., Borthwick, H., Charlesworth, B., and Betancourt, A.J. (2014). Faster-X Effects in Two Drosophila Lineages. Genome Biol. Evol. *6*, 2968–2982.

Barbash, D.A., Siino, D.F., Tarone, A.M., and Roote, J. (2003). A rapidly evolving MYB-related protein causes species isolation in Drosophila. Proc. Natl. Acad. Sci. *100*, 5302–5307.

Barbash, D.A., Awadalla, P., and Tarone, A.M. (2004). Functional Divergence Caused by Ancient Positive Selection of a Drosophila Hybrid Incompatibility Locus. PLoS Biol. *2*, e142.

Bateson, W. (1909). Heredity and variation in modern lights. In Darwin and Modern Science, (Cambridge: Cambridge University Press), pp. 85–101.

Bi, Y.-M., Meyer, A., Downs, G.S., Shi, X., El-Kereamy, A., Lukens, L., and Rothstein, S.J. (2014). High throughput RNA sequencing of a hybrid maize and its parents shows different mechanisms responsive to nitrogen limitation. BMC Genomics *15*, 77.

Bolnick, D.I., and Near, T.J. (2005). Tempo of hybrid inviability in centrarchid fishes (Teleostei: Centrarchidae). Evolution *59*, 1754–1767.

Bolnick, D.I., Turelli, M., Lopez-Fernandez, H., Wainwright, P.C., and Near, T.J. (2008). Accelerated Mitochondrial Evolution and “Darwin’s Corollary”: Asymmetric Viability of Reciprocal F1 Hybrids in Centrarchid Fishes. Genetics *178*, 1037–1048.

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.

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

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 Charlesworth, B. (1989). Genetic analysis of X-linked sterility in hybrids between three sibling species of Drosophila. Heredity *62*, 97–106.

Coyne, J.A., and Orr, H.A. (1989). Two rules of speciation. Speciat. Its Consequences 180–207.

Coyne, J.A., and Orr, H.A. (1997). “Patterns of Speciation in Drosophila” Revisited. Evolution *51*, 295.

Coyne, J.A., and Orr, H.A. (2004). Speciation (Sinauer Associates, Inc.).

Curtis, C.F. (1982). the mechanism of hybrid male sterility from crosses in the Anopheles gambiae and Glossina morsitans complexes. In Recent Developments of Insect Disease Vectors, (New York: Stipes Publishing), pp. 290–312.

Darwin, C. (1859). On the Origin of Species by Means of Natural Selection, Or, the Preservation of Favoured Races in the Struggle for Life (London: J. Murray).

Davies, B., Hatton, E., Altemose, N., Hussin, J.G., Pratto, F., Zhang, G., Hinch, A.G., Moralli, D., Biggs, D., Diaz, R., et al. (2016). Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. Nature *530*, 171–176.

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. Mol. Biol. Evol. *32*, 2534–2546.

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

Dobzhansky, T. (1937). Genetics and the Origin of Species (New York: Columbia University Press).

Dobzhansky, T. (1970). Genetics of the evolutionary process (Columbia University Press).

Duselis, A.R., and Vrana, P.B. (2010). Aberrant growth and pattern formation in Peromyscus hybrid placental development. Biol Reprod *83*, 988–996.

Fontaine, M.C., Pease, J.B., Steele, A., Waterhouse, R.M., Neafsey, D.E., Sharakhov, I.V., Jiang, X., Hall, A.B., Catteruccia, F., Kakani, E., et al. (2015). Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science *347*, 1258524–1258524.

Goldschmidt, R. (1940). The material basis of evolution (New Haven, Conn.: Yale University Press).

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

Hall, A.B., Qi, Y., Timoshevskiy, V., Sharakhova, M.V., Sharakhov, I.V., and Tu, Z. (2013). Six novel Y chromosome genes in Anopheles mosquitoes discovered by independently sequencing males and females. BMC Genomics *14*, 1–13.

Hall, A.B., Papathanos, P.-A., Sharma, A., Cheng, C., Akbari, O.S., Assour, L., Bergman, N.H., Cagnetti, A., Crisanti, A., Dottorini, T., et al. (2016). Radical remodeling of the Y chromosome in a recent radiation of malaria mosquitoes. Proc. Natl. Acad. Sci. *113*, E2114–E2123.

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

Ishikawa, R., and Kinoshita, T. (2009). Epigenetic programming: the challenge to species hybridization. Mol Plant *2*, 589–599.

Johnson, N.A. (2010). Hybrid incompatibility genes: remnants of a genomic battlefield? Trends Genet. *26*, 317–325.

Krzywinska, E., Dennison, N.J., Lycett, G.J., and Krzywinski, J. (2016). A maleness gene in the malaria mosquito Anopheles gambiae. Science *353*, 67–69.

Laschiavo, M., Nguyen, Q.K., Duelis, A.R., and Vrana, P.B. (2007). Maping and identification of candidate loci responsible for Peromyscus hybrid overgrowth. Mamm. Genome *18*, 75–85.

Laurie, C.C. (1997). the weaker sex is heterogametic: 75 years of Haldane's rule. Genetics *147*, 937–951.

Linnaeus, C. (1735). Systema Naturae (Netherlands: Johann Friedrich Gmelin).

Maheshwari, S., and Barbash, D.A. (2011). The Genetics of Hybrid Incompatibilities. Annu. Rev. Genet. *45*, 331–355.

Malone, J.H., and Michalak, P. (2008). Physiological Sex Predicts Hybrid Sterility Regardless of Genotype. Science *319*, 59.

Masly, J.P., and Presgraves, D.C. (2007). High-resolution genome-wide dissection of the two rule of speciation in Drosophila. PLOS Biol. *5*, e243.

Mayr, E. (1963). Animal species and Evolution (Cambridge: Belknap Press of Harvard University Press).

Meisel, R.P., and Connallon, T. (2013). The faster-X effect: integrating theory and data. Trends Genet. *29*, 537–544.

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

Moore, T., and Haig, D. (1991). Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet. *7*, 45–49.

Moran, P.A., Ritchie, M.G., and Bailey, N.W. (2017). A rare exception to Haldane's rule: Are X chromosomes key to hybrid incompatibilities? Heredity.

Moyle, L.C. (2006). Genome-Wide Associations Between Hybrid Sterility QTL and Marker Transmission Ratio Distortion. Mol. Biol. Evol. *23*, 973–980.

Muller, H. (1942). Isolating mechanisms, evolution and temperature. Biol Symp 71–125.

Nosil, P., and Schluter, D. (2011). The genes underlying the process of speciation. Trends Ecol. Evol. *26*, 160–167.

Oliver, P.L., Goodstadt, L., Bayes, J.J., Birtle, Z., Roach, K.C., Phadnis, N., Beatson, S.A., Lunter, G., Malik, H.S., and Ponting, C.P. (2009). Accelerated Evolution of the Prdm9 Speciation Gene across Diverse Metazoan Taxa. PLoS Genet. *5*, e1000753.

Orr, H.A. (1993). Haldane's rule has multiple genetic causes. Nature *361*, 532–533.

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

Perez, D.E., and Wu, C.-I. (1995). Further Characterization of the Odysseus Locus of Hybrid Sterility in Drosophila: One Gene is Not Enough. Genetics *140*, 201–206.

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

Presgraves, D.C. (2010). The molecular evolutionary basis of species formation. Nat. Rev. Genet. *11*, 175–180.

Presgraves, D.C., and et al. (2003). Adaptive evolution drives divergence of a hybrid inviability gene between two species of Drosophila. Nature *423*, 715–719.

Ray, J. (1686). Historia Plantarum (London: Clark).

Renaut, S., Nolte, A.W., and Bernatchez, L. (2009). Gene Expression Divergence and Hybrid Misexpression between Lake Whitefish Species Pairs (Coregonus spp. Salmonidae). Mol. Biol. Evol. *26*, 925–936.

Sawamura, K. (1996). Maternal effect as a cause of exceptions for Haldane's rule. Genetics *143*, 609.

Sawamura, K., Taira, T., and Watanabe, T.K. (1993). Hybrid lethal systems in the Drosophila melanogaster species complex. I. the maternal hybrid rescue (mhr) gene of Drosophlia simulans. Genetics *133*, 299–305.

Schilthuizen, M., Giesbers, M., and Beukeboom, L.W. (2011). Haldane's rule in the 21st century. Heredity *107*, 95–102.

Seaton, G., Haley, C.S., Knott, S.A., Kearsey, M., and Visscher, P.M. (2002). QTL Express: mapping quantitative trait loci in simple and complex pedigrees. Bioinformatics *18*, 339–340.

Slotman, M., Della Torre, A., and Powell, J.R. (2004). The genetics of inviability and male sterility in hybrids between Anopheles gambiae and An. arabiensis. Genetics *167*, 275–287.

Städler, T., Florez-Rueda, A.M., and Paris, M. (2012). Testing for "Snowballing" Hybrid Incompatibilities in Solanum: Impact of Ancestral Polymorphism and Divergence Estimates. Mol. Biol. Evol. *29*, 31–34.

Stanley, S.M. (1979). Macroevolution, Pattern and Process (San Francisco: W.H. Freeman).

Stupar, R.M., Hermanson, P.J., and Springer, N.M. (2007). Nonadditive Expression and Parent-of-Origin Effects Identified by Microarray and Allele-Specific Expression Profiling of Maize Endosperm. PLANT Physiol. *145*, 411–425.

Stupar, R.M., Gardiner, J.M., Oldre, A.G., Haun, W.J., Chandler, V.L., and Springer, N.M. (2008). Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. BMC Plant Biol. *8*, 33.

Sun, S., Ting, C.-T., and Wu, C.-I. (2004). the normal function of a speciation gene, Odysseus, and its hybrid sterility effect. Science *305*, 81–83.

Tang, S., and Presgraves, D.C. (2015). Lineage-Specific Evolution of the Complex Nup160 Hybrid Incompatibility Between Drosophila melanogaster and Its Sister Species. Genetics *200*, 1245–1254.

Tiffin, P., Olson, M.S., and Moyle, L.C. (2001). Asymmetric crossing barriers in angiosperms. Proc. R. Soc. B *268*, 861–867.

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

Tulchinsky, A.Y., Johnson, N.A., and Porter, A.H. (2014). Hybrid incompatibility despite pleiotropic constraint in a sequence-based bioenergetic model of transcription factor binding. Genetics *198*, 1645–1654.

Turelli, M., and Moyle, L.C. (2006). Asymmetric Postmating Isolation: Darwin's Corollary to Haldane's Rule. Genetics *176*, 1059–1088.

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

Volff, J.N. (2005). Genome evolution and biodiversity in teleost fish. Heredity *94*, 280–294.

Vrana, P.B., Guan, X.-J., Ingram, R.S., and Tilghman, S.M. (1998). Genomic imprinting is disrupted in interspecific Peromyscus hybrids. Nat. Genet. *20*, 362–365.

Wang, H., and Dey, S.K. (2006). Roadmap to embryo implantation: clues from mouse models. Nat. Rev. Genet. *7*, 185–199.

Wang, D., Liu, F., Wang, L., Huang, S., and Yu, J. (2011). Nonsynonymous substitution rate (Ka) is a relatively consistent parameter for the defining fast-evolving and slow-evolving protein-coding genes. Biol. Direct *6*, 1–17.

White, B.J., Collins, F.H., and Besansky, N.J. (2011). Evolution of Anopheles gambiae in Relation to Humans and Malaria. Annu. Rev. Ecol. Syst. *42*, 111–132.

Wright, K.M., Lloyd, D., Lowry, D.B., Macnair, M.R., and Willis, J.H. (2013). Indirect evolution of hybrid lethality due to linkage with selected locus in Mimulus guttatus. PLOS Biol. *11*, e1001497.

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

Wu, C.-I., and Ting, C.-T. (2004). Genes and speciation. Nat. Rev. Genet. *5*, 114–122.

Wu, C.-I., Johnson, N.A., and Palopoli, M.F. (1996). Haldane's rule and its legacy: Why are there so many sterile males? Trends Ecol. Evol. *11*, 281–284.

Zouros, E., Lofdahl, K., and Martin, P.A. (1988). Male Hybrid Sterility in Drosophila: Interactions between Autosomes and Sex Chromosomes in Crosses of D. mojavensis and D. arizonensis. Evolution *42*, 1321.

Zufall, R.A., and Rausher, M.D. (2004). Genetic changes associated with floral adaptation restrict future evolutionary potential. Nature *428*, 847–850.

(2005). World Malaria Report.

(2015). World Malaria Report.

Chapter I

Re-evaluation of 50 year old hybridizations between members of the *Anopheles gambiae* complex

Abstract

Hybridization experiments to test for reproductively compatible populations are used define morphologically identical species. Defining species has been especially important in mosquito research because some morphologically identical species can have vastly different impacts on transmitting human diseases. In this study I examine a 1964 hybridization study to both see how our current colonies have changed with respect to sex-ratio distortion and what new information we can extract from these hybridization experiments. By exploring current experimental evidence of possible causes of incompatibilities like those observed in hybrids between species of the *Anopheles gambiae* complex we can gain a better understanding of how these hybrid incompatibilities arose.

Introduction

In the 1960's George Davidson made many contributions to the field of mosquito speciation including, in his 1964 paper, evidence that there was a third freshwater species in the *Anopheles gambiae* complex (Davidson, 1964a). Since the mosquitoes of this complex are all morphologically identical (e.g. showing no obvious phenotype to separate populations systematically), a biological cross is needed to test if the populations are reproductively

isolated and thus different species (Coluzzi et al., 1979). For the *An. gambiae* complex, Davidson defined separate species as those populations that resulted in sterile hybrid males (Davidson 1964).These mosquitoes are relatively young species and only 1.85 million years diverged contributing to the observed pattern that only male hybrids were sterile while the hybrid females remained fertile. However, just in the past few years *An. gambiae* S and M form have been found to have recently speciated and only show pre-zygotic incompatibilities in the form of mating behavior differences in wild populations (Coetzee et al., 2013). The two newly christened species, *An. gambiae* and *An. coluzzii,* can mate readily in colony and form fertile male and female hybrids. In hybrids of closely related species the heterogametic sex, in this case the male, is the first to show signs of hybrid incompatibilities (HI) in the form of sterility or inviability of the F1. This observation of sex-specific HI is called Haldane's rule (Haldane, 1922). Davidson found both sterility and inviability between these *Anopheles* crosses that would be expected to follow Haldane's rule; however, in some of his crosses the F1 females were partially or completely inviable. This observation is inconsistent with Haldane's rule.

Exceptions to Haldane’s rule are rare, though they are more commonly found when the species’ hybrids exhibit inviability rather than sterility (Schilthuizen et al., 2011). Of the 461 (452 from Schilthuizen et al., and 9 from Wu and Davis) species crosses that resulted in hybrid inviability two Aves (bird) crosses, 28 Amphibia crosses, 37 Teleostei (fish) crosses, 9 *Drosophila* crosses, one Hemiptera cross, one Coleoptera cross, and one Lepidoptera cross are known to not follow Haldane’s rule in terms of inviability

(Schilthuizen et al., 2011; Wu and Davis, 1993). These taxa utilize either male or female heterogametry, meaning Haldane's rule can be rejected regardless of which sex is heterogametic. In the literature to date, rejection of Haldane's rule that results in female inviability when the male is the heterogametic sex is caused by maternal effects (as is the case in some *Drosophila* hybrid crosses) or, potentially, homomorphic sex chromosomes where the X and Y can still recombine (in the case of *Bufonidae* toad hybrid crosses) (Fontenot, 2009; Sawamura, 1996).

While the majority of in-depth HI studies have utilized model organisms, we chose to look at the *Anopheles gambiae* complex not only because of its recent radiation and previously documented strong HI, but because of the fact that *Anopheles* mosquitoes have a direct impact on human health. Anopheles mosquitoes are the most efficient vectors of malaria where they reside in Africa. All but one of the six members of the *Anopheles gambiae* complex are considered effiecient vectors of the malaria parasite which causesover 214 million casesworldwide , 188 million in Africa,in 2015 alone (2015). Of those 188 million cases in Africa, 295,000 cases resulted in death, and while this number has decreased from 670,000 deaths reported in 2005, the fight against malaria is still an on-going one (2005). Any study into the biology of these disease vectors could help improve control strategies and looking back to previous studies is a good place to start.

In this paper I re-examined data from Davidson's crosses and repeated Davidson's original crosses, to the best of my ability, in order to

assess any changes that have occurred over the years of colonization and to ask if there has been any insights that can better explain the female and asymmetric inviability that Davidson found.

Methods

Mosquito rearing and sterility testing

Mosquitoes used in this study were obtained through BEI Resources, NIAID, NIH: *An. gambiae* (Kisumu and Pimperena colonies), *An. coluzzii* (Mali-NIH and Akron colonies), *An. quadriannulatus* (Sangqua colony), *An.arabiensis* (Dongola1 and KGB colonies), and *An. merus* (MAF colony). All mosquitoes were reared and crossed in accordance with standard protocols (White et al., 2013). In brief, mosquitoes were maintained in insectaries under controlled conditions of 27°C, 65% relative humidity, and a 12h:12h light:dark cycle with 1 h dawn and dusk transitions. All parental and F1 larvae were reared in freshwater (dH2O) at a density of 200 larvae/L of water. Each larval tray was fed ~100 mg per day of a 4:1 mixture of finely ground fish pellets to baker's yeast. Reciprocal F1 crosses were performed with 200-400 virgins from each parental colony, parental crosses were performed in the same fashion. All eggs of F1 crosses were hatched, resulting in 200-1600 first instar larvae per cross. At least 800 eggs are hatched (when possible) to keep Parental colonies alive. F1 males were tested for sterility at 5-7 days old by visual inspection of their reproductive organs and considered sterile if no

mature sperm was found in the testes or vas deferens, picture of these testes were taken using an iPhone camera as it offered better lighting visualization than a microscope camera. Only reciprocal F1 crosses between *An. gambiae* and *An. coluzzii* resulted in fertile males. All parental crosses had fertile offspring that had sex-ratios that did not significantly differ from 50% female: 50% male.

KGB and MAF cross survivorship

To see if the cause of extreme sex-ratio distortion seen in the KGB (*An.arabiensis)* by MAF (*An. merus)* reciprocal crosses was death in the embryonic stage or one of the later stages of development I took 200 just hatched first instar larvae and counted the larvae each day until adulthood. Four different trials of 200 larvae each, for each cross, were counted, resulting in 800 larvae counted per cross. The resulting pupae were pooled for each cross and allowed to emerge in one cage. Adults were counted each day and moved to a cage separate from the remaining pupae.

Statistical analysis

Statistical analysis of sex-ratio distortion of the lab-reared colony crosses was carried out by means of a Chi-squared test. All crosses were expected to have 50-50 sex-ratios. Comparison to the original wild crosses made by Davidson in 1964 was done using a two-tailed Fisher's exact test.

Results

Colony crosses and Hybrid Sterility

Five different species of the *Anopheles gambiae* complex (using eight different colonies: two each for *An. gambiae s.s.* (Kisumu and Pimperena), *An*. *coluzzii* (Akron and Mali-NIH), and *An*. *arabiensis* (KGB and Dongola 1), and one per *An*. *quadriannulatus* (Sangqua) and *An*. *merus* (MAF)) were crossed in order to form 12 pairs of reciprocal F1 hybrids. For the species where I had access to multiple colonies, I selected the colony with the most larvae available for the F1 cross. These hybrids were immediately sexed and counted upon emergence in order to test for sex-ratio distortion All male hybrids, except for the *An*. *gambiae* by *An*. *coluzzii* reciprocal crosses, were dissected at 3-5 days old and found to be lacking mature sperm and were thus considered sterile. All pure-species parental colonies were also tested for sex-ratio distortion and all were found to not significantly deviate from a 1:1 ratio of females:males. While the original crosses from Davidson (1964) contained an additional species, I was unable to include *An*. *melas* in the crosses, as no colony is currently available.

Most of the hybrid testes show the same phenotype originally described in Davidson's paper: very reduced testes size, and somewhat reduced accessory gland size, though the *An. merus* by *An. arabiensis* cross gave testes of normal size (Figure 1.1). The reduced testes are very thin and have no mature sperm though they do contain large cells, possibly primary spermatocytes that have arrested in mitosis or before meiosis I. In the *An*.

merus by *An*. *arabiensis* cross, the testes contain some cells that can have multiple tails and one or more bulbous “heads” (Figure 1.1).

Davidson crosses versus current colony crosses

Data from twelve different hybridizations between four different species were taken from Davidson's 1964 paper where he described male hybrid sterility and sex-ratio distortion in the F1 generations. The mosquito species used were as follows: A form (now called *An*. *gambiae*), B form (now called *An*. *arabiensis*), C form (now called *An*. *quadriannulatus*), and *An*. *merus* (Bryan, 1979; Mattingly, 1977; White, 1973). As the original crosses Davidson used included *An*. *gambiae* from regions where there is now known to be overlap between *An*. *gambiae* and the new species *An*. *coluzzii*, the original crosses containing *An*. *gambiae* were compared to the recent crosses that contained *An*. *gambiae* or *An*. *coluzzii* colonies in order to account for the possibility that the original crosses had some ad-mixture of both species, or were entirely *An*. *coluzzii*.

Of the crosses that have significantly different sex-ratios between the old and new crosses, three show a switch in sex-ratio, two becoming more male-skewed (*An. gambiae*♀ by *An. quadriannulatus*♂ and *An. gambiae*♀ by *An. arabiensis*♂) where the original was female-skewed and one becoming more female-skewed (*An. arabiensis*♀ by *An. coluzzii*♂) where the original cross was male-skewed (Table 1.1). One of the new crosses, *An. arabiensis*♀ by *An. merus*♂*,* shows an increase in the sex-ratio distortion, becoming even more female skewed (67.51% female in the new cross versus 56% female in

the old cross). The remainder of the crosses that show a significant difference in sex-ratios are comprised of four new crosses that are now not significantly different from a 50:50 sex-ratio and three new crosses that have a milder male-skew than the original crosses.

Of the non-significantly different crosses *An. merus*♀ by *An. arabiensis*♂ is the only cross to have a highly significant skew towards male offspring, the original cross not having any adult female F1s and our new crosses only having 6 female F1s out of 1026 total offspring. It is also interesting to note that while the *An. arabiensis*♀ by *An. coluzzii*♂ cross that uses the Dongola1 colony has become significantly more female-skewed than the original cross, the *An. arabiensis*♀ by *An. coluzzii*♂ cross that uses the KGB colony sex-ratio is not significantly different from the original cross.

Sex-ratio distortion

Of the 24 reciprocal crosses performed (12 different pairs of colonies), eleven were found to have sex-ratios significantly different from 50% female: 50% male (Table 1.2). Of those eleven, only three show significant skew towards female offspring, the male skew in eight of crosses seems to dispute Haldane's rule. These data suggest that the female F1 hybrids have the more extreme phenotype of being less viable, while the males are only sterile (Haldane, 1922). Five of the skewed crosses, one female biased and the other four male biased, had normal sex ratios in their corresponding reciprocal cross, and the remaining three sets of the reciprocal crosses had sex-ratio distortion in both crosses: *An*. *gambiae* (Pimperena) by *An*. *arabiensis* (KGB),

An. *arabiensis* (KGB) by *An*. *merus* (MAF), and *An*. *coluzzii* (Akron) by *An*. *coluzzii* (Mali-NIH).

The Pimperena female by KGB male and KGB female by Pimperena male crosses both resulted in male-biased sex ratios (35% female, 65% male, $p = 0.001$; 36.07% female, 63.93% male, $p = 0.0002$). The KGB by MAF and Akron by Mali-NIH crosses show opposite sex-ratio distortion in their reciprocal crosses. KGB by MAF is female skewed (67.65% female, 32.35% male, $p < 0.0001$) whereas MAF by KGB F1's are almost exclusively male (0.58% female, 99.42% male, $p < 0.0001$), and the males have sperm that seems to have arrested further into spermatogenesis (Figure 1.1). Akron by Mali-NIH and Mali-NIH by Akron show mild sex-ratio distortion, which is odd since they are both *An*. *coluzzii*, of 45.47% female and 54.53% male ($p = 0.0485$) and 54.6% female and 45.40% male, respectively. This apparent difference between the Akron and Mali-NIH colonies of *An*. *coluzzii* is also seen when crossing them with *An*. *quadriannulatus* (Sangqua). When Mali-NIH is crossed with Sangqua, no sex-ratio distortion is seen, however, when Sangqua females are crossed with Akron males, a significant, but slight, male-biased sex ratio occurs (41.21% female, 58.79% male, $p = 0.0131$).

To take a closer look at the current colonies with the most extreme sex-ratio distortion, KGB by Mali-Nih and its reciprocal, I performed a larval survivorship trial by measuring how many larvae survived from hatching to adulthood. By counting the number of larvae each day we can determine if the sex-ratio distortion is occurring at any specific larval stage, and if we don't see

enough deaths to account for the missing sex then we can assume that the distortion is occurring during the embryonic stage or prior. Figure 1.2 shows the average survivorship from hatching to pupal stage of each cross. Figure 1.2, and Table 1.3 demonstrate that the KGB by MAF cross survivorship decreases at a relatively steady rate that is enough to account for the female-skewed sex-ratios if only male larvae are dying off (Hatch-Adult survival: 67.4%). For the MAF by KGB cross, we see relatively high survivorship of the larvae to pupation and adulthood, which cannot account for the highly male-skewed sex-ratios that we see in this cross (Hatch-Adult survival: 76.9%). This means that the mechanism behind the male-skewed sex ratios in this cross must be due to events that occur during embryonic development or fertilization.

Discussion

In re-evaluating of Davidson's crosses in 1964 I found that many of the current colonies do not show similar signs of sex-ratio distortion in the F1 generation compared the original crosses. I observed interesting patterns in the current crosses and can hypothesize that genomic interactions may be occurring to induce “Darwin's Corollary”, or isolation asymmetry, and female inviability (Turelli and Moyle, 2006). While Davidson's work predates theories for the mechanisms that underpin Darwin’s corollary, he did report that female-skewed sex-ratios in hybrids may be due to cytoplasmic incompatibilities as was found by Dobzhansky while crossing *Dros. texana* to *Dros. montana* (Dobzhansky 1951). While this is still a valid hypothesis more is known about

the likely causes of this incompatibility, namely maternal effects (Sawamura, 1996; Turelli and Orr, 2000), chromosome imprinting (Kelsey and Feil, 2013; Wolf et al., 2014), or simple chromosomal incompatibilities (Bateson, 1909; Dobzhansky, 1937; Muller, 1942). These incompatibilities may also interfere with normal sperm development contributing to hybrid male sterility as all our crosses show complete sterility, save for *An. gambiae* by *An. coluzzii*.

Davidson Cross comparison

Although some of the cross comparisons did not show significant change in sex ratio, the majority of the new crosses showed either an increase in distortion, the opposite of the original distortion direction, or a reduction of the distortion towards a 50:50 ratio than seen in the original crosses. This suggests that these colonies are constantly evolving and the changes seen are most likely caused by genetic drift and/or a founder effect when the colonies where started (Mason et al., 1987). Davidson also reported on mostly single-pair matings, and mentioned that that may be the reason for some of the more extreme sex-ratio distorted crosses as he had found more normal (e.g. 50:50) sex ratios between saltwater females (*An. merus* and *An. melas*) and freshwater males (*An. arabiensis, An. gambiae,* and *An. quadriannulatus*) when mated en-mass (Davidson 1962). This reasoning can account for the lack of sex-ratio distortion I found in the current *An. merus* by *An. gambiae* and *An. merus* by *An. quadriannulatus.* However, the crosses between the current colonies of *An. merus* and *An. arabiensis* had sex-ratio distortion

towards overwhelmingly male F1 offspring that he saw in the original cross even though these crosses where carried out en-mass.

Only three of the eight crosses with an *An. coluzzii* parent show a difference in sex-ratio as compared to the original Davidson cross. This suggests that Davidson had some contamination of *An. coluzzii* in his *An. gambiae* populations. The three distorted crosses could be explained a lack of *An. coluzzii* or that my crosses, being mass-matings, lacked enough diversity so that alleles causing distortion are balanced by enough individuals that have non-distorting alleles. This same difference in diversity between mass-mating versus single-pair matings could also be true of all those crosses that are significantly different in sex-ratio from their original cross.

Strangely I also see differences in our comparisons depending on which colony of *An. arabiensis* used in crosses with *An. coluzzii*. When the cross is KGB♀ by *An. coluzzii*♂ I observe no difference in sex-ratio compared to the original cross, whereas when I crossed Dongola1♀ by *An. Coluzzii* ♂ I see a significant switch in distortion to female-biased from the original male-biased observations. This case could be due to the different colonies being made from different wild populations: KGB was colonized from Kanyemba, Zimbabwe in 1975 and Dongola1 from the Northern State of Sudan in 2005 (Benedict et al., 2009).

Maternal Effects

In hybrids there can be incompatibilities between the maternal factors and paternal alleles before the maternal-zygotic transition that can lead to

sterility or inviability of the hybrids. These incompatibilities are what is thought to underlie the exceptions to Haldane's rule, those cases where the homogametic sex is seemingly more affected that the heterogametic sex (Sawamura, 1996; Turelli and Orr, 2000). This time scale tells us that there must be signs of inviability before the zygotic genes have taken over in the larval stage.

This maternal effect is thought to work through piRNA and siRNA, and when there are incompatibilities between the maternal piRNA or siRNA and the paternal transposons it can cause a failure of transposon suppression (Ferree and Barbash, 2007; Labrador et al., 1999; Tao et al., 2003). Increased transposon activity leads to increased mutation rates, chromosomal re-arrangements, and sterility(Bregliano et al., 1980; O'Neil et al., 1998). It is possible these TE accumulations could also lead to inviability due to growth defects caused by those mutations and chromosomal re-arrangements. This could explain the asymmetrical nature of some of our crosses. When we see sex-ratio bias in one reciprocal cross but not the other this could be due to the maternal effects in species one's female not being able to suppress transposons on X or Y chromosome from species two's male due to lack of or incompatible siRNAs. If the transposon is on the X there is female inviability, if it is on the Y there is male inviability. The inviability does not occur in the other cross because species one's males do not contain those transposons and species two's females contain the right suppressors for any transposons that they might contribute from their X chromosome.

While more complicated, it could also explain the one cross where the reciprocal crosses have the opposite biases as in the *An. merus* by *An. arabiensis* crosses, or the same bias as in the *An. gambiae* by *An. arabiensis* crosses. In the case of opposite biases there would have to be different incompatible maternal effects in both species that affect different sexes. There would be a maternal siRNA in *An. merus* that failed to suppress a transposon on the *An. arabiensis* X chromosome and a different maternal siRNA in *An. arabiensis* that half the time failed to suppress a transposon on the *An. merus* Y as sex ratios for that cross are 67.51% female to 32.49% male suggesting that either the siRNA or the recognition site for the siRNA is polymorphic and segregating evenly in either species populations. To get the same biases for both reciprocal crosses the X chromosomes of both species would have to contain transposons that the other species could not suppress, as they are both have sex-ratios of about 35% female to 65% male, the elements involved most likely have two alleles segregating in the population, one that can still suppress, and one that cannot.

Recently piRNAs have been found in *An. gambiae* though as yet it is unclear if they contribute to tranposable element suppression in reproductive or developmental processes as in other insects (Biryukova and Ye, 2015; Castellano et al., 2015; George et al., 2015; Macias et al., 2014).

Paternal Chromosomal Imprinting

It is possible that some of the asymmetrical differences in our crosses are due to chromosomal imprinting of the paternal chromosomes that can either not be maintained in the zygote by maternal factors, or by genomic factors in the hybrids (Turelli and Moyle, 2006; Wolf et al., 2014). Genes that contribute to reproductive isolation are typically fast evolving, and paternally methylated imprinting control regions are thought to evolve faster than maternal ones (Schulz et al., 2010). Imprinting signals are also known to be rapidly evolving and are usually growth factors in animals (Vrana et al., 1998). Insects, plants, and mammals all show imprinting, though they differ proportion of the genome that is imprinted. Imprinting was first discovered in the insect *Sciara* (Crouse, 1960). *Sciara* was found to have large imprinted domains that usually encompass whole chromosomes and are the mechanism for *Sciara*'s sex-determination (Bongiorni and Prantera, 2003; de la Casa-Esperon and Sapienza, 2003; Goday and Esteban, 2001). While whole-chromosome imprinting has also been observed in *Drosophila,* there is also a mammalian-specific form of imprinting that only controls the expression of single genes(Fitch et al., 1998; Lloyd, 2000; Loppin et al., 2005). While imprinting is a possible mechanism of reproductive isolation in our hybrids, there is usually an aspect of sexual competition associated with imprinting that these mosquitoes are not known to practice.

In our crosses that show female inviability in one cross and normal sex-ratios in the reciprocal , we can speculate that the X chromosome delivered

by the sperm is imprinted in a way that cannot be maintained by the female's imprinting maintenance mechanism, or that the female lacks the mechanistic elements necessary for imprinting maintenance (Kelsey and Feil, 2013; Wolf et al., 2014). If this occurs there can be a loss of imprinting of the paternal X which leads to a misregulation of chromatin states and lethality, as seen in *Arabidopsis* hybrids (Josefsson et al., 2006). This could also be the case in our crosses that show male inviability, where the Y chromosome is imprinted and improperly maintained in the zygote. The *An. gambiae* Y chromosome has recently been sequenced and found to have a multitude of repetitive and transposon elements, which may be transcriptionally suppressed through imprinting (Hall et al., 2016).

Genomic Incompatibilities

Genomic incompatibilities are though to underlie most B-D-M incompatibilities. These factors are separate from imprinting and maternal effects because they are independent of the parent-of-origin and are not necessarily epigenetically regulated (Oliver et al., 2009; Perez and Wu, 1995). These gene incompatibilities may also lead to inviability of hybrids if cell-cycle regulation genes are involved (Phadnis et al., 2015).

B-D-M factors may be underlying the mechanisms of inviability observed in the *An. gambiae* by *An. arabiensis* crosses, where in both reciprocal crosses half the females are inviable. Interactions within the females between X chromosomes from the two species are likely to be causing inviability of the female hybrids. The male hybrids, having only the X

from the maternal species, are still viable, though sterile. Having incompatibilities between the X chromosomes could be explained by the large-X effect, where speciation genes are disproportionately found on the X chromosomes (Coyne, 1992; Coyne and Orr, 1989; Presgraves, 2008). It may be that the combination of both X chromosomes with the rest of the hybrid genome causes too incompatibilities that it results in inviability, where the one X chromosome in the males only leads to sterility.

In our asymmetrically biased crosses we can explain the asymmetry with the idea that there is an allele segregating in one of the species where one copy is lethal to the hybrid and the other is not. If this allele is on the X chromosome of the male parent, we would see half inviable females, half viable females, and all viable males as they would get the Y chromosome instead. This would give us a 1:2 ratio of viable females to viable males, which we see in our *An. quadriannulatus* ♀ by *An. arabiensis* ♂ cross, and the reciprocal cross would have an even ratio because half the males and females would be inviable do to the even chance of getting the lethal allele from the maternal parent. The other cases of male-skew could be due to the allele being less abundant in the population.

Hybrid Male Sterility versus Inviability

In most of the new crosses we see the same signature of hybrid sterility as Davidson saw in the original crosses. These sterility phenotypes remain unchanged since the 1960's even after many lab generations, even though their inviability phenotypes have changed (Figure 1.1). This suggests

that hybrid sterility may have evolved first and reach a stable point while the interactions causing inviability are still evolving. The cross between *An. coluzzii* Mali-NIH and Akron colonies have slight, but significant, asymmetrical inviability , but no sterility, meaning that the genic or epigenetic interactions that cause inviability have more readily evolved in the *An. coluzzii* species. As the Mali-NIH colony was started with only 80 wild females it is likely to carry different alleles than the Akron colony due to founder-like effects. Inviability is considered a more extreme phenotype than sterility and the genes involved are thought to effect intra-specific breeding and growth as they involve cell cycle regulators (Phadnis et al. 2015) or chromosomal defects due to transposons (Labrador et al., 1999; Metcalfe et al., 2007; O'Neil et al., 1998).

Future Directions

To narrow down the possible causes of asymmetric inviability several experiments could be performed. One would be to record when during the life cycle the inviable F1s were dying. While Davidson did record this, there is no clear stage that the deaths were occurring as his pure *An. gambiae* and *An. merus* control colonies still had fairly low hatch rates, 64% and 68% respectively, and the pure *An. arabiensis* colony had only a 37% survival rate from egg to adult. The differences in numbers could be from the fact that these were very new colonies that were not used to lab conditions, but our current colonies may have a better survival rate. If we see that the F1s are dying during the embryonic stage, then we know that maternal factors are most likely.

To uncover genomic incompatibilities the most thorough technique would be to use Quantitative Trait Loci (QTL) mapping in backcrossed or introgressed lines of hybridized mosquitoes, though this would only be possible where one of the sexes was still present and fertile.

Conclusions

The comparisons between old and new crosses of *Anopheles* species gave us a unique look at how these mosquitoes may have evolved since the 1960s. To repeat and re-affirm the most intriguing finding of almost complete female inviability in one species cross reminds the research community that organisms other than model organism can, and should, be used in inviability and sterility research.

Given the importance of *Anopheles* mosquitoes, especially those in the *Anopheles gambiae* complex, in transmitting malaria, any knowledge on how to disrupt their life cycle is important when designing new control strategies. While the focus has been mainly on inducing sterility or releasing sterile males into the wild to suppress wild populations, males containing genes that induce inviability in females can be extremely useful in this effort. The genes or interactions that cause female inviability between *An. merus* and *An. arabiensis* could either be used in place of sterility genes in a form of sterile insect technique (SIT) that instead of releasing sterile males into the wild would release males with this inviability gene under a drive mechanism that would then produce all male offspring that would also only be able to produce

male offspring and so on until there are no females left to propagate the species. Or these genes could be used to produce just males when making enough sterile males to release for SIT without having to filter out females.

Acknowledgements

I'd like to acknowledge Mikkal Blick and Michelle Bui for rearing and counting mosquitoes (both hybrid and parental crosses).

References

Anaka, M., Lynn, A., McGinn, P., and Lloyd, V.K. (2009). Genomic Imprinting in Drosophila has properties of both mammalian and insect imprinting. Dev. Genes Evol. *219*, 59–66.

Bateson, W. (1909). Heredity and variation in modern lights. In Darwin and Modern Science, (Cambridge: Cambridge University Press), pp. 85–101.

Benedict, M., Knols, B., Bossin, H., Howell, P., Mialhe, E., Caceres, C., and Robinson, A. (2009). Colonisation and mass rearing: learning from others. Malar. J. *8*, S2–S4.

Biryukova, I., and Ye, T. (2015). Endogenous siRNAs and piRNAs derived from transposable elements and genes in the malaria vector mosquito Anopheles gambiae. BMC Genomics *16*.

Bongiorni, S., and Prantera, G. (2003). Imprinted facultative heterochromatization in mealybugs. Genetica *117*, 271–279.

Bregliano, J., Picard, G., Bucheton, A., Pelisson, A., Lavige, J., and L'Heritier, P. (1980). Hybrid dysgenesis in Drosophila melanogaster. Science *207*, 606–611.

Bryan, J.H. (1979). Observations on the member species of the Anopheles gambiae complex in The Gambia, West Africa. Trans. R. Soc. Trop. Med. Hyg. *73*, 463–466.

de la Casa-Esperon, E., and Sapienza, C. (2003). Natural selection and the evolution of genome imprinting. Annu. Rev. Genet. *37*, 349–370.

Castellano, L., Rizzi, E., Krell, J., Di Cristina, M., Galizi, R., Mori, A., Tam, J., De Bellis, G., Stebbing, J., Crisanti, A., et al. (2015). The germline of the malaria mosquito produces abundant miRNAs, endo-siRNAs, piRNAs and 29-nt small RNAs. BMC Genomics *16*, 100.

Coetzee, M., Hunt, R.H., Wilkerson, R., Della Torre, A., Coulibaly, M.B., and Besansky, N.J. (2013). Anopheles coluzzii and Anopheles amharicus, new members of the Anopheles gambiae complex. Zootaxa *3619*, 246–274.

Coluzzi, M., Sabatini, A., Petrarca, V., and Di Deco, M.A. (1979). Chromosomal differentiation and adaptation to human environments in the Anopheles gambiae complex. Trans. R. Soc. Trop. Med. Hyg. *73*, 483–497.

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

Coyne, J.A., and Orr, H.A. (1989). Two rules of speciation. Speciat. Its Consequences 180–207.

Crouse, H.V. (1960). The controlling element in sex chromosome behavior in Sciara. Genetics *45*, 1429–1443.

Davidson, G. (1964). Anopheles gambiae, a complex of species. Bull. World Health Organ. *31*, 625.

Dobzhansky, T. (1937). Genetics and the Origin of Species (New York: Columbia University Press).

Ferree, P.M., and Barbash, D.A. (2007). Distorted Sex Ratios: A Window into RNAi-Mediated Silencing. PLoS Biol. *5*, e303.

Fitch, K.R., Yasuda, G.K., Owens, K.N., and Wakimoto, B.T. (1998). Paternal effects in Drosophila: implications for mechanisms of early development. Curr Top. Dev Biol *38*, 1–34.

Fontenot, B.E. (2009). Natural Hybridization And Speciation In Toads Of The Anaxyrus americanus Group.

George, P., Jensen, S., Pogorelcnik, R., Lee, J., Xing, Y., Brasset, E., Vaury, C., and Sharakhov, I.V. (2015). Increased production of piRNAs from euchromatic clusters and genes in Anopheles gambiae compared with Drosophila melanogaster. Epigenetics Chromatin *8*.

Goday, C., and Esteban, M.R. (2001). Chromosome elimination in sciarid flies. Bioessays *23*, 242–250.

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

Hall, A.B., Papathanos, P.-A., Sharma, A., Cheng, C., Akbari, O.S., Assour, L., Bergman, N.H., Cagnetti, A., Crisanti, A., Dottorini, T., et al. (2016). Radical remodeling of the Y chromosome in a recent radiation of malaria mosquitoes. Proc. Natl. Acad. Sci. *113*, E2114–E2123.

Josefsson, C., Dilkes, B., and Comai, L. (2006). Parent-Dependent Loss of Gene Silencing during Interspecies Hybridization. Curr. Biol. *16*, 1322–1328.

Kelsey, G., and Feil, R. (2013). New insights into establishment and maintenance of DNA methylation imprints in mammals. Philos. Trans. R. Soc. B Biol. Sci. *368*, 20110336.

Labrador, M., Farré, M., Utzet, F., and Fontdevila, A. (1999). Interspecific hybridization increases transposition rates of Osvaldo. Mol. Biol. Evol. *16*, 931–937.

Lloyd, V. (2000). Parental imprinting in Drosophila. Genetica *109*, 35–44.

Loppin, B., Bonnefoy, E., Anselme, C., Laurencon, A., Karr, T.L., and Couble, P. (2005). The histone H3.3 chaperone HIRA is essential for chromatin assembly in the male pronucleus. Nature *437*, 1386–1390.

Macias, V., Coleman, J., Bonizzoni, M., and James, A.A. (2014). piRNA pathway gene expression in the malaria vector mosquito *Anopheles stephensi*: piRNA genes in *Anopheles stephensi*. Insect Mol. Biol. *23*, 579–586.

Mason, L.J., Pashley, D.P., and Johnson, S.J. (1987). The Laboratory as an Altered Habitat: Phenotypic and Genetic Consequences of Colonization. Fla. Entomol. *70*, 49.

Mattingly, P.F. (1977). Names for the Anopheles gambiae complex. Mosq. Syst. *9*, 323–328.

Metcalfe, C.J., Bulazel, K.V., Ferreri, G.C., Schroeder-Reiter, E., Wanner, G., Rens, W., Obergfell, C., Eldridge, M.D.B., and O'Neill, R.J. (2007). Genomic instability within centromeres of interspecific marsupial hybrids. Genetics *177*, 2507–2517.

Muller, H. (1942). Isolating mechanisms, evolution and temperature. Biol Symp 71–125.

Oliver, P.L., Goodstadt, L., Bayes, J.J., Birtle, Z., Roach, K.C., Phadnis, N., Beatson, S.A., Lunter, G., Malik, H.S., and Ponting, C.P. (2009). Accelerated Evolution of the Prdm9 Speciation Gene across Diverse Metazoan Taxa. PLoS Genet. *5*, e1000753.

O'Neil, R.J.W., O'Neil, M.J., and Graves, J.A.M. (1998). Undermethylation associated with retroelement activation and chromosome remodeling in a interspecific mammalian hybrid. Nature *393*, 68–72.

Perez, D.E., and Wu, C.-I. (1995). Further Characterization of the Odysseus Locus of Hybrid Sterility in Drosophila: One Gene is Not Enough. Genetics *140*, 201–206.

Phadnis, N., Baker, E.P., Cooper, J.C., Frizzell, K.A., Hsieh, E., De La Cruz, A.F.A., Shendure, J., Kitzman, J.O., and Malik, H.S. (2015). An essential cell cycle regulation gene causes hybrid inviability in Drosophila. Science *350*, 1552–1555.

Presgraves, D.C. (2008). Sex chromosomes and speciation in Drosophila. Trends Genet. *24*, 336–343.

Sawamura, K. (1996). Maternal effect as a cause of exceptions for Haldane's rule. Genetics *143*, 609.

Schilthuizen, M., Giesbers, M., and Beukeboom, L.W. (2011). Haldane's rule in the 21st century. Heredity *107*, 95–102.

Schulz, R., Proudhon, C., Bestor, T.H., Woodfine, K., Lin, C.-S., Lin, S.-P., Prissette, M., Oakey, R.J., and Bourc'his, D. (2010). The Parental Non-Equivalence of Imprinting Control Regions during Mammalian Development and Evolution. PLoS Genet. *6*, e1001214.

Tao, Y., Chen, S., Hartl, D.L., and Laurie, C.C. (2003). Genetic dissection of hybrid incompatibilities between Drosophila simulans and D. mauritiana. I. Differential accumulation of hybrid male sterility effects on the X and autosomes. Genetics *164*, 1383–1398.

Turelli, M., and Moyle, L.C. (2006). Asymmetric Postmating Isolation: Darwin's Corollary to Haldane's Rule. Genetics *176*, 1059–1088.

Turelli, M., and Orr, H.A. (2000). Dominance, Epistasis, and the Genetics of Postzygotic Isolation. Genetics *154*, 1663–1679.

Vrana, P.B., Guan, X.-J., Ingram, R.S., and Tilghman, S.M. (1998). Genomic imprinting is disrupted in interspecific Peromyscus hybrids. Nat. Genet. *20*, 362–365.

White, G.B. (1973). Comparative Studies on Sibling Species of Anopheles-Gambiae Giles Complex (Dipt Culicidae) 3. Distribution, Ecology, Behavior and Vectorial Importance of Species-D in Bwamba-County, Uganda, with an Analysis of Biological, Ecological, Morphological and Cytogenetical Relationships of Ugandan Species-D. Bull. Entomol. Res. *63*, 65–97.

White, B.J., Kundert, P.N., Turissini, D.A., Van Ekeris, L., Linser, P.J., and Besansky, N.J. (2013). Dose and developmental responses of Anopheles merus larvae to salinity. J. Exp. Biol. *216*, 3433–3441.

Wolf, J.B., Oakey, R.J., and Feil, R. (2014). Imprinted gene expression in hybrids: perturbed mechanisms and evolutionary implications. Heredity *113*, 167–175.

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

(2005). World Malaria Report.

(2015). World Malaria Report.

FIGURES

Figure Legends:

Figure 1.1 Testes and male accessory glands (MAGs) of fertile *An. coluzzii* parental (A) and fertile *An. coluzzii* testes close-up (B). Sterile testes and MAGs of *An. merus* by *An. coluzzii* hybrids (the same phenotype for all sterile hybrids except for *An. merus* by *An. arabiensis* hybrids) (C) and close up of sterile hybrid testes (D). *An. merus* by *An. arabiensis* testes with MAG (E) and close up of testes (F). Close up of deformed *An. merus* by *An. arabiensis* sperm, green arrows point to bulges in normally thin, hair-like sperm (G). Red arrows denote MAGs, blue arrows denote testes (A, C, E).

Figure 1.2 Survivorship of *An. arabiensis* by *An. merus* reciprocal crosses Survivorship from first instar larvae to pupae of *An. arabiensis* ♀ by *An. merus* ♂ F1 hybrids (A.) and *An. merus* ♀ by *An. arabiensis* ♂ F1 hybrids (B.)

Figure 1.1

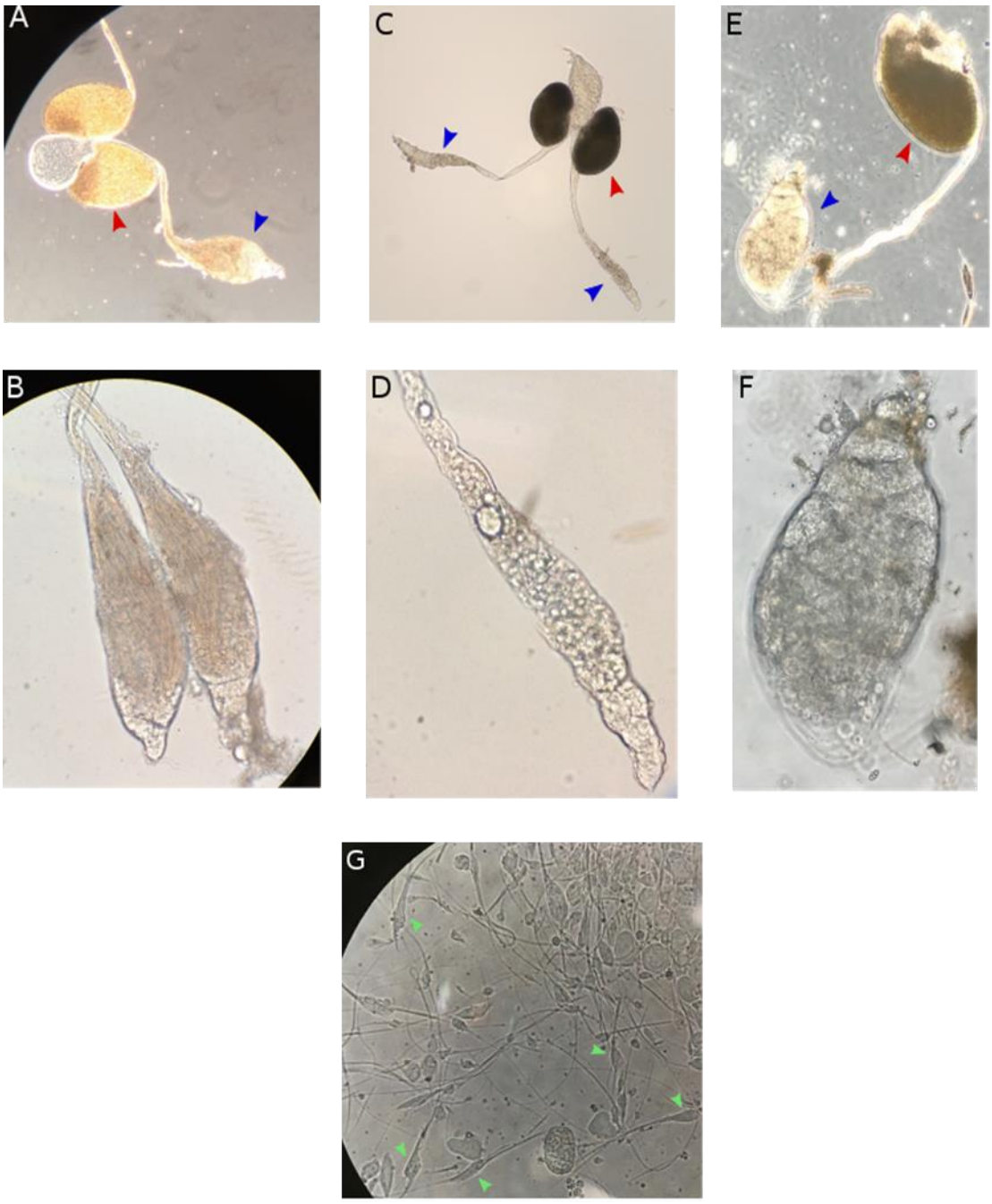
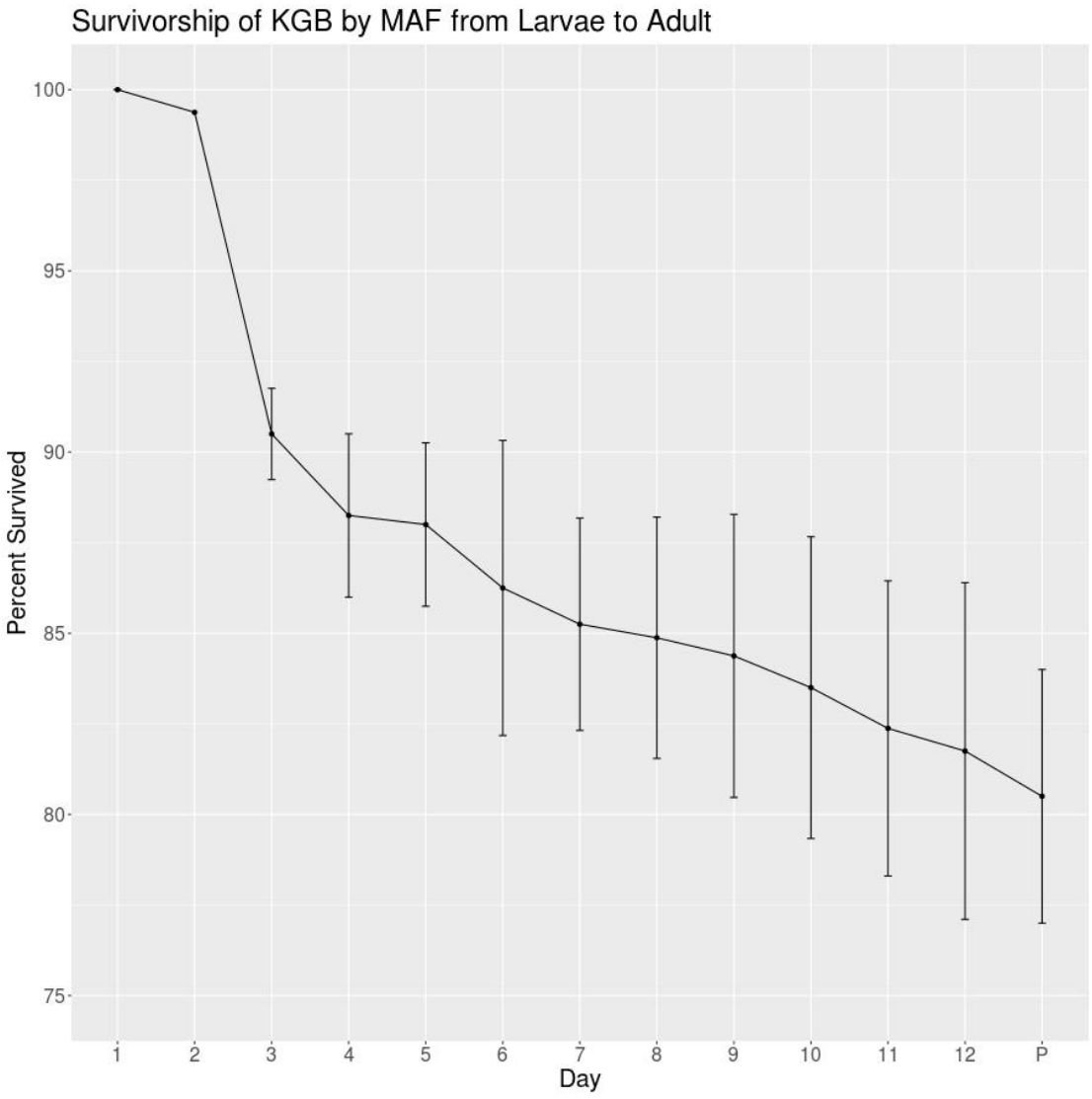
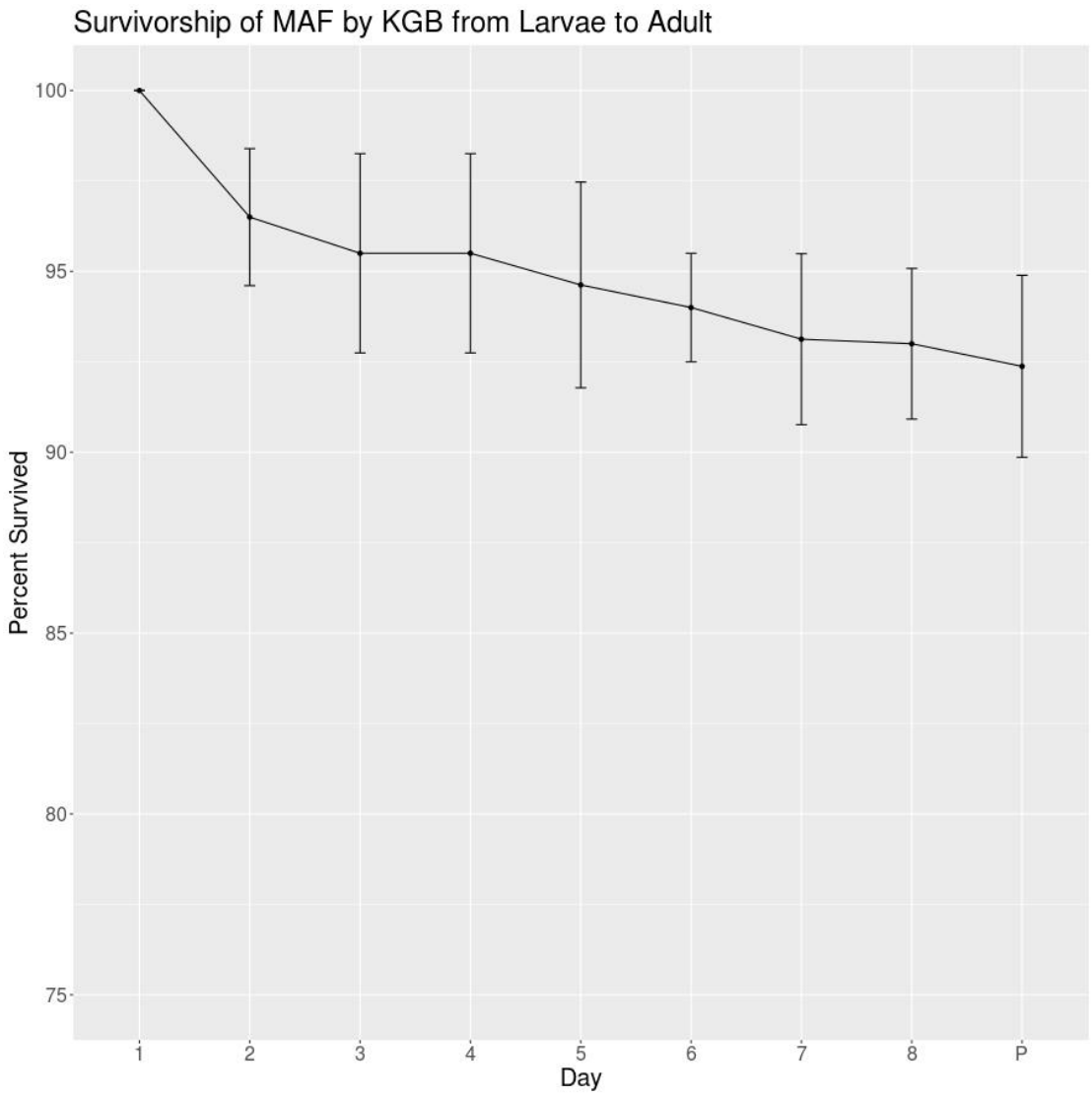


Figure 1.2

A.


Survivorship of KGB by MAF from Larvae to Adult
Percent Survived
100
95
90
85
80
75
1
2
3
4
5
6
7
8
9
10
11
12
P
Day

B.


Survivorship of MAF by KGB from Larvae to Adult
Percent Survived
100
95
90
85
80
75
1
2
3
4
5
6
7
8
P
Day

TABLES

Table 1.1 New crosses versus Davidson's crosses.

New Crosses					Davidson's Crosses					P-value
Female Parent	**Male Parent**	**F1 %female**	**F1 %male**	**Total F1**	**Female Parent**	**Male Parent**	**F1 %female**	**F1 %male**	**Total F1**	
An.arabiensis (KGB)	*An.gambiae* (Pimperena)	36.07	63.93	183	*An. arabiensis* (B)	*An. gambiae/coluzzii* (A)	45	55	383	0.0558
An. gambiae(Pimperena)	*An.arabiensis* (KGB)	35.00	65.00	120	*An. gambiae/coluzzii* (A)	*An. arabiensis* (B)	51	49	1445	0.0008*
An.coluzzii (Akron)	*An.arabiensis* (Dongola1)	51.76	48.24	255	*An. gambiae/coluzzii* (A)	*An. arabiensis* (B)	51	49	1445	0.8387
An.arabiensis (Dongola1)	*An.coluzzii* (Akron)	58.71	41.29	201	*An. arabiensis* (B)	*An. gambiae/coluzzii* (A)	45	55	383	0.0017*
An.coluzzii (Akron)	*An.arabiensis* (KGB)	47.06	52.94	255	*An. gambiae/coluzzii* (A)	*An. arabiensis* (B)	51	49	1445	0.249
An.arabiensis (KGB)	*An.coluzzii* (Akron)	46.84	53.16	348	*An. arabiensis* (B)	*An. gambiae/coluzzii* (A)	45	55	383	0.6039
An.merus(MAF)	*An.gambiae* (Kisumu)	43.79	56.21	507	*An. merus*	*An. gambiae/coluzzii* (A)	2	98	326	0.0001*
An.gambiae (Kisumu)	*An.merus(* MAF)	49.68	50.32	473	*An. gambiae/coluzzii* (A)	*An. merus*	57	43	564	0.0208
An.merus(MAF)	*An.arabiensis* (KGB)	0.58	99.42	1026	*An. merus*	*An. arabiensis* (B)	0	100	284	0.3502
An.arabiensis (KGB)	*An.merus(* MAF)	67.65	32.35	1490	*An. arabiensis* (B)	*An. merus*	56	44	382	0.0001*
An.gambiae (Kisumu)	*An.quadriannulatus* (Sangqua)	44.67	55.33	347	*An. gambiae/coluzzii* (A)	*An. quadriannulatus* (C)	53	47	1087	0.008*
An.quadriannulatus (Sangqua)	*An.gambiae* (Kisumu)	45.08	54.92	193	*An. quadriannulatus* (C)	*An. gambiae/coluzzii* (A)	13	87	477	0.0001*
An.coluzzii (Akron)	*An.quadriannulatus* (Sangqua)	51.61	48.39	434	*An. gambiae/coluzzii* (A)	*An. quadriannulatus* (C)	53	47	1087	0.6493
An.quadriannulatus (Sangqua)	*An.coluzzii* (Akron)	41.21	58.79	199	*An. quadriannulatus* (C)	*An. gambiae/coluzzii* (A)	13	87	477	0.0001*
An.coluzzii (Mali-Nih)	*An.quadriannulatus* (Sangqua)	49.44	50.56	629	*An. gambiae/coluzzii* (A)	*An. quadriannulatus* (C)	53	47	1087	0.1607
An.quadriannulatus (Sangqua)	*An.coluzzii* (Mali-Nih)	49.39	50.61	164	*An. quadriannulatus* (C)	*An. gambiae/coluzzii* (A)	13	87	477	0.0001*
An.arabiensis (KGB)	*An.quadriannulatus* (Sangqua)	50.53	49.47	374	*An. arabiensis* (B)	*An. quadriannulatus* (C)	48	52	334	0.4987
An.quadriannulatus (Sangqua)	*An.arabiensis* (KGB)	38.50	61.50	361	*An. quadriannulatus* (C)	*An. arabiensis* (B)	25	75	334	0.0001*
An.merus(MAF)	*An.quadriannulatus* (Sangqua)	50.00	50.00	1006	*An. merus*	*An. quadriannulatus* (C)	57	43	345	0.0246*
An.quadriannulatus (Sangqua)	*An.merus(* MAF)	50.19	49.81	261	*An. quadriannulatus* (C)	*An. merus*	61	39	131	0.053

Total F1 are the total number of F1 progeny that were sex-sorted. P-values of <0.05 are considered significant and are marked with *.

Table 1.2 Reciprocal crosses using modern-day colonies.

Cross		F1 generation		Chi-squared	P-value
Female Parent	Male Parent	Female	Male		
An.arabiensis(KGB)	An.gambiae(Pimperena)	42 (35%)	78 (65%)	10.8	0.001*
An. gambiae(Pimperena)	An.arabiensis(KGB)	66 (36.07%)	117 (63.93%)	14.213	0.0002*
An.coluzzii(Akron)	An.arabiensis(Dongola1)	132 (51.76%)	123 (48.24%)	0.318	0.573
An.arabiensis(Dongola1)	An.coluzzii(Akron)	118 (58.71%)	83 (41.29%)	6.095	0.0136*
An.coluzzii(Akron)	An.arabiensis(KGB)	120 (47.06%)	135 (52.94%)	0.882	0.3476
An.arabiensis(KGB)	An.coluzzii(Akron)	163 (46.84%)	185 (53.16%)	1.391	0.2383
An.merus(MAF)	An.gambiae(Kisumu)	222 (43.79%)	285 (56.21%)	7.828	0.0051*
An.gambiae(Kisumu)	An.merus(MAF)	235 (49.68%)	238 (50.32%)	0.019	0.8903
An.merus(MAF)	An.arabiensis(KGB)	6 (0.58%)	1020 (99.42%)	1002.14	0.0001*
An.arabiensis(KGB)	An.merus(MAF)	1008 (67.65%)	482 (32.35%)	185.689	0.0001*
An.gambiae(Kisumu)	An.quadriannulatus(Sangqua)	155 (44.67%)	192 (55.33%)	3.945	0.047*
An.quadriannulatus(Sangqua)	An.gambiae(Kisumu)	87 (45.08%)	106 (54.92%)	1.87	0.1714
An.coluzzii(Akron)	An.quadriannulatus(Sangqua)	224 (51.61%)	210 (48.39%)	0.452	0.5016
An.quadriannulatus(Sangqua)	An.coluzzii(Akron)	82 (41.21%)	117 (58.79%)	6.156	0.0131*
An.coluzzii(Mali-Nih)	An.quadriannulatus(Sangqua)	311 (49.44%)	318 (50.56%)	0.078	0.7802
An.quadriannulatus(Sangqua)	An.coluzzii(Mali-Nih)	81 (49.39%)	83 (50.61%)	0.024	0.8759
An.arabiensis(KGB)	An.quadriannulatus(Sangqua)	189 (50.53%)	185 (49.47%)	0.043	0.8361
An.quadriannulatus(Sangqua)	An.arabiensis(KGB)	139 (38.5%)	222 (61.5%)	19.083	0.0001*
An.merus(MAF)	An.quadriannulatus(Sangqua)	503 (50%)	503 (50%)	0	1
An.quadriannulatus(Sangqua)	An.merus(MAF)	131 (50.19%)	130 (49.81%)	0.004	0.9506
An.coluzzii(Akron)	An.gambiae(Kisumu)	408 (47.55%)	450 (52.45%)	2.056	0.1516
An.gambiae(Kisumu)	An.coluzzii(Akron)	372 (53.3%)	326 (46.7%)	3.032	0.0817
An.coluzzii(Akron)	An.coluzzii(Mali-Nih)	216 (45.47%)	259 (54.53%)	3.893	0.0485*
An.coluzzii(Mali-Nih)	An.coluzzii(Akron)	380 (54.6%)	316 (45.4%)	5.885	0.0153*
Parental Colony					
An.gambiae(Kisumu)	An.gambiae(Kisumu)	214 (50.95%)	206 (49.05%)	0.152	0.6963
An.arabiensis(KGB)	An.arabiensis(KGB)	393 (50.38%)	387 (49.62%)	0.046	0.8299
An.coluzzii(Akron)	An.coluzzii(Akron)	342 (50.59%)	334 (49.41%)	0.095	0.7583
An.coluzzii(Mali-Nih)	An.coluzzii(Mali-Nih)	190 (52.49%)	172 (47.51%)	0.895	0.3441
An.quadriannulatus(Sangqua)	An.quadriannulatus(Sangqua)	61 (48.03%)	66 (51.97%)	0.197	0.6573
An.merus(MAF)	An.merus(MAF)	621 (49.72%)	628 (50.28%)	0.039	0.843

The female and male count of the F1 hybrid or parental colony is given along with the percentage. A chi-squared test was used to determine if the sex-ratios were different from 50-50 males to females, p-values < 0.05 are considered significant and are marked with *.

Table 1.3 *An.arabiensis* by *An.merus* cross larval survivorship.

Cross	Trial				Hatch-Adult	Pupae-Adult	%Female	%Male
	A	B	C	D				
KGBxMAF	88% (176)	75.5% (151)	82% (164)	76.5%(153)	67.4%(539)	83.7%(539)	67.9%(366)	32.1%(173)
MAFxKGB	90.5% (181)	89.5% (179)	91.5% (183)	94.5% (189)	76.9% (615)	83.2%(615)	0.81% (5)	99.19%(610)

Percent survival for each reciprocal cross per trial (out of 200 initial larvae). The total number survived is given in parentheses next to the percentage. The trials show the larvae percentage that made it to the pupal stage for each trial.

CHAPTER II

QTL mapping of hybrid sterility between two species in the *Anopheles gambiae* complex

Abstract

Over a century of work on post-zygotic reproductive isolation has revealed two empirical patterns: the heterogametic sex is generally the first to exhibit sterility or inviability (Haldane's rule) and sex chromosomes play a disproportionate role in hybrid dysfunction (the large X-effect). Despite the consistency of these two 'rules of speciation', the underlying mechanisms remain unresolved. We investigated the genetic basis of male sterility in hybrids between the sibling species *Anopheles coluzzii* and *An. merus* – two Afrotropical mosquito vectors of malaria. Reconstruction of genome-wide ancestry in ~2000 backcross progeny and QTL mapping shows an association with not only male hybrid sterility, but also the large X-effect.

Introduction

The Dobzhansky-Muller incompatibility (DMI) model posits that as species diverge over time they accumulate genetic differences, which may negatively interact when combined in the same genetic background (Dobzhansky 1937; Muller 1942), resulting in hybrid sterility or inviability (reviewed in(Presgraves, 2010)). Two empirical patterns have emerged from over a century of work on post-zygotic reproductive isolation and genic incompatibilities: Haldane's rule and the large X-effect. First postulated in 1922, Haldane's rule states that if only one

sex of hybrids is sterile or inviable it will be the heterogametic sex (Haldane, 1922); diverse taxa including birds (Gray, 1958), *Drosophila* (Bock, 1984), *Lepidoptera* (Presgraves, 2002), mammals (Craft, 1938; Gray, 1972), and plants (Brothers and Delph, 2010) with both XY and ZW sex determination systems largely obey the rule (Coyne and Orr, 2004; Schilthuizen et al., 2011). The large X-effect argues that the X (or Z) chromosome plays a disproportionate role in the expression of hybrid dysfunction and reflects data from both genetic backcrosses and natural hybrid zones (Coyne, 1992; Coyne and Orr, 1989; Presgraves, 2008). The remarkable consistency of these two 'rules of speciation' across taxa has spurred extensive theoretical and experimental work to uncover the mechanistic underpinnings of each pattern.

The dominance theory, which hypothesizes that most sterility or inviability alleles are partially recessive, is a well-accepted explanation for many cases of Haldane's rule (Coyne, 1985; Muller, 1942; Orr and Turelli; Turelli and Orr, 1995). If a partially recessive incompatible mutation arises on one of the sex chromosomes, the heterogametic sex is exposed to the full effect of the mutation due to hemizygosity, while the homogametic sex is – at least partially – protected. However, the dominance theory is an inadequate explanation for many cases of hybrid male sterility, which conforms to Haldane's Rule at a higher rate than other types of post-zygotic isolation (Presgraves and Orr, 1998; Watson and Demuth, 2012; Wu and Davis, 1993; Wu et al., 1996). Additionally, the dominance theory fails to predict the taxonomically widespread observation that male sterility

evolves faster than male inviability (Coyne and Orr, 2004; Schilthuizen et al., 2011; Wu and Davis, 1993; Wu et al., 1996). To explain this pattern, Wu proposed the “faster male” theory, which states that genes involved in spermatogenesis functionally diverge more rapidly than genes causing other types of hybrid dysfunction (Wu and Davis, 1993; Wu et al., 1996). “Faster-male” evolution could be due to sexual selection on male fertility factors (Hollocher and Wu, 1996; Llopart, 2012; True et al., 1996) or to an inherent fragility in the spermatogenesis process (Wu and Davis, 1993), although neither hypothesis has been extensively tested. A second explanation for the “faster-male” theory is recurrent genomic conflict between the X and Y-chromosomes. Empirical evidence suggests that meiotic drive elements arise at increased frequency on the sex chromosomes relative to the autosomes (Jaenike, 2001). When such drive occurs, it will result in unbalanced sex ratios (and often decreased fertility), providing strong selective pressure for the evolution of autosomal repressors (Atlan et al., 2003). Genes involved in this meiotic drive ‘arms-race’ will undergo accelerated divergence after lineage splitting and may have negative pleiotropic effects on hybrid sperm development (Frank, 1991; Hurst and Pomiankowski, 1991; McDermott and Noor, 2010). Indeed, certain *Drosophila* hybrid sterility genes show strong sex ratio meiotic drive when put into a heterospecific genetic background which presumably lacked the pressure to evolve drive suppressors (Orr and Irving, 2005; Phadnis and Orr, 2009; Tao et al., 2007b, 2007a). However, critics question whether sex-linked meiotic drive is actually more likely

to occur than autosomal drive, pointing to contradictory theoretical evidence (Coyne et al., 1991) and the inherent bias in empirical discovery of sex-ratio distorters over autosomal distorters (Jaenike, 2001).

In contrast to Haldane's Rule, mechanistic explanations of the large X-effect lack strong empirical support. The most comprehensive data on the genetics of the large-X effect is on hybrid male sterility. In both *Drosophila* and mouse, it appears that the large-X effect is due to a higher density of X-linked versus autosomal incompatibilities rather than differences in effect size between individual incompatibilities on the X and autosomes (Good et al., 2008; Masly and Presgraves, 2007). While the X generally experiences slightly higher rates of adaptive substitution than the autosomes (Meisel and Connallon, 2013; Singh et al., 2008), the increased rate of evolution does not appear sufficient to explain the large X-effect, especially in light of the paucity of spermatogenesis genes located on the X (Masly and Presgraves, 2007; Presgraves, 2008), although opinions differ (Counterman et al., 2004; Llopart, 2012). As with Haldane's rule, recurrent meiotic drive resulting in sex-ratio distortion and its subsequent suppression could underlie the large-X effect since each DMI would necessarily involve at least one sex-linked allele (McDermott and Noor, 2010).

The *Anopheles gambiae* complex – a group of at least eight isomorphic sibling mosquito species endemic to Sub-Saharan Africa – is an excellent system for studying post-zygotic reproduction isolation and the two 'rules of speciation' (White et al., 2011). Due to the prominent role of the sibling species in malaria

transmission, a wealth of genomic and ecological studies exists on the complex. Molecular dating based on whole genome sequencing estimates that the adaptive radiation of the group occurred ~1.85 mya (Fontaine et al., 2015). Strong assortative mating limits contemporary gene flow between species (Assogba et al., 2014; Marchand, 1984), but hybrids are found at low rates and genomic signatures of introgression are evident between many of the species (Coluzzii et al., 1979; Donnelly et al., 2004; Marchand, 1984; Weetman et al., 2014; White, 1971). In agreement with Haldane's Rule, nearly all crosses between sibling species result in sterile hybrid males, but fully fertile females (Davidson, 1964b; Hunt et al., 1998; White, 1973). Not only does the fertility of F1 hybrid females provide a conduit for gene flow in nature, it also allows for laboratory backcrossing and subsequent genetic analysis of hybrid male sterility.

Here we have used high-throughput quantitative genomics to dissect the genetic basis of hybrid male sterility between two members of the *An. gambiae* complex: *An. coluzzii* (previously *An. gambiae* M Form) and *An. merus*. Both species are synanthropic and highly efficient vectors of human malaria where present (Coetzee et al., 2013; Cuamba and Mendis, 2009; Govere et al., 2000; Mwangangi et al., 2003; Pock Tsy et al., 2003; Temu et al., 1998). Currently, the two species have allopatric distributions; the range of *An. coluzzii* extends from Northern Senegal in the west to East-Central Africa and south to Angola (Coetzee et al., 2013), while the brackish water breeding *An. merus* is primarily restricted to the coasts of east Africa and Mozambique (Sinka et al., 2010). In

addition to potential insights into the formation and maintenance of species boundaries, elucidation of genes critical to proper *Anopheles* sperm development could enable population suppression through genetic-based sterile insect technique (Nolan et al., 2011; Wang and Jacobs-Lorena, 2013). Discovery of such novel malaria control tools will be critical to curbing disease transmission as physiological and behavioral resistance to insecticidal interventions continues to spread through mosquito populations (Ranson et al., 2011; Reddy et al., 2011).

METHODS

Mosquito Rearing and crosses

All mosquitoes were reared and crossed in accordance with standard protocols (White et al., 2013). In brief, mosquitoes were maintained in insectaries under controlled conditions of 27°C, 65% relative humidity, and a 12h:12h light:dark cycle with 1 h dawn and dusk transitions. *Anopheles coluzzii* larvae were reared in freshwater (dH2O) at a density of 200 larvae/L of water, while *An. merus* larvae were reared in 25% saltwater (7.5 g NaCl/L) at the same density. Backcross individuals were reared in dH20 with no adverse effects on survival. Each larval tray was fed ~100 mg per day of a 4:1 mixture of finely ground fish pellets to baker's yeast. Backcrosses were performed with 200-400 virgins from each parental and F1 colony.

Phenotyping

We dissected the male reproductive tract, including the testes and male accessory glands (MAGs), from 5-10 day old virgin males into phosphate buffered saline. Testes were separated from the MAGs and scored for fertility under 400x or 1000x (oil) magnification on a compound microscope using a qualitative classification of 1 to 5 following (Slotman et al., 2004). Different phenotypic classes had the following characteristics: 1, limited to no morphological sperm development; 2, no mature spermatids but some abnormal sperm with bulges and multiple tails; 3, mature spermatids in vas deferens, but only abnormal sperm in the testes; 4, mature spermatids in the vas deferens and half abnormal/normal sperm in the testes; and 5, large numbers of mature spermatids in both vas deferens and testes with characteristic swirls of sperm. All F1 mosquitoes examined (50 from each reciprocal cross) were fully sterile and all parental mosquitoes were fully fertile. We attempted to correlate phenotypic classification with actual fertility, but single pair interspecific matings proved unsuccessful as is common in *An. gambiae* complex specimens due to eurygamy (Benedict et al., 2009).

ddRADseq Library Generation

We made slight modifications to the ddRAD protocol designed by Peterson and colleagues (Peterson et al., 2012). Our goal was to simultaneously sequence a reduced, yet highly reproducible and representative, portion of the

genome from individual mosquitoes. In brief, DNA was extracted from individual mosquitoes and digested with *MluC1* and *NlaIII.* After purification of the DNA with Ampure beads, a barcoded (1 of 48) and universal adapter were ligated to opposite ends of each fragment. DNA from up to 48 individuals was then pooled, cleaned, and size-selected around 400bp to make an individual sub-library. Each sub-library was PCR amplified for 12 cycles with an indexed (1 of 12) and universal primer. Up to 12 sub-libraries were pooled equimolarly and subjected to single-end 100bp sequencing on the Illumina HiSeq 2500 at the UCR Genomics core. Detailed protocols are available at mosquitogenomics.org/protocols.

Genotyping and ancestry reconstruction

We successfully assigned >98% of sequencing reads to individuals based on adapter and barcode sequence, allowing up to one mismatch in each sequence (all barcodes/adapters are unique by three mutational steps). Reads were then mapped against the *An. gambiae* PEST (AgamP4.2) reference genome using burrow-wheelers alignment (BWA) (v1.2.3) (Li and Durbin, 2009) with n=8 and otherwise default parameters. After mapping, BAM files for each group of mosquitoes (*An. coluzzii*, *An. merus*, col^{BC}, and mer^{BC}) were merged and genotypes were called in parallel for all individuals using the GATK program with default parameters (McKenna et al., 2010). A custom Perl script extracted individual genotype information from the resulting vcf file and identified SNPs that had a F_{ST} of 1 between *An. coluzzii* and *An. merus* parental colonies. Only SNPs

genotyped in at least 16 individuals from each colony were retained for analysis. To eliminate SNPs located in repetitive DNA, autosomal sites with total coverage more than 4 standard deviations from mean coverage and X-linked sites with coverage more than 1 SD from the mean were excluded from further analysis. To correct for mapping error, SNPs with any of the following characteristic were filtered out prior to analysis: 1) a frequency less than 25% or greater than 80% in backcross progeny, 2) the backcross parent allele was not present for an individual, 3) the non-backcross parent allele was homozygous in > 20% of genotypes called by GATK, and 4) X-linked SNPs where > 10% of individuals had heterozygous genotypes assigned by GATK.

Coverage varied from SNP to SNP and individual to individual leading to incorrect genotype calls using GATK. Thus, we wrote a Hidden-Markov model (HMM) to impute ancestry and infer crossover breakpoints in each individual backcross mosquito (Andolfatto et al., 2011; Elliott et al., 1994). To reduce noise, a single SNP with the highest coverage from each RAD fragment for each individual mosquito was chosen for use in the HMM analysis. The HMM had two states: homozygous and heterozygous. The HMM used a transition probability of $a = 2\times10^{-8}(x_i - x_{i-1})e^{-2\times10^{-8}(x_i - x_{i-1})}$ where $x_i - x_{i-1}$ is the distance between neighboring SNPs, and an emission probability of $e = \sum_{i=0}^{n} n!p^i(1-p)^{n-i} \sum_{j=\max(0,k-i)}^{\min(k,n-i)} \frac{p_{ab}^{2j-k+i}(1-p_{ab})^{n+k-i-2j}}{j!(n-i-j)!(j-k+i)!(k-j)!}$ where k is the read depth of the backcross parental allele, n is the total coverage, p is the expected allele

frequency for the state (p=1 for homozygous sites and p=0.5 for heterozygous sites), and p_{ab} is the sequencing error rate between the two alleles (set to 0.01).

Individuals with no markers on any chromosome or less than 30 total markers were discarded. Crossover breakpoints were called by taking the midpoint of the genomic interval over which the posterior probability of a given ancestry state dropped below 50%. Multiple 'peaks' where ancestry switched from one state to another and then back again over the scale of a few SNPs were present after application of the HMM. Due to crossover interference we do not except any crossovers to occur over such small distances (reviewed in (Hillers, 2004)). We concluded that most 'peaks' resulted from mapping errors as they were identified across multiple individuals, although a very small proportion may be true gene conversion hotspots. We filtered out SNPs associated with all such 'peaks' prior to generation of ancestry maps and downstream QTL and recombination analyses. Recombination rate maps were made by binning crossover breakpoints into 1 Mb intervals and fitting the derivative of a spline to the cumulative map using a smoothing parameter of 1.

QTL analysis

Markers from the final HMM analysis were utilized for QTL mapping with probability estimates for SNPs converted to 'hard-calls' by rounding. Generation of marker maps and initial QTL mapping was performed using both the Haley-Knott (HK) regression (Martinez and Curnow, 1992; Seaton et al., 2002) and the expectation-maximization (EM) algorithms (Liu, 1997) in the R/qtl package

(Broman et al., 2003). Use of either QTL algorithm resulted in large, chromosome-wide QTLs (Figure 2.8). To further narrow down QTL, confidence interval mapping and Bayesian interval mapping were performing using the R/qtl and R/qtlbim packages, respectively (Arends et al., 2010). Since the QTL analysis was utilized to find autosomal genomic regions that interact with a foreign X, only backcross progeny with an non-recombined X as determined by the HMM were included. Due to the data being non-normal, CIM LOD scores were found using the imputation and HK methods using window sizes of five, ten, and twenty. The resulting CIM QTL were subjected to a permutation test in order to determine the threshold of significance. Significant QTL in the BIM analysis were determined finding the best pattern of QTL using R/qtlbim and fitting it to the data using R/qtl to find the best fit. QTL intervals for CIM were calculating by identifying the entire region above the significance level, while the interval within 1.5 LPD of the peak marker was used in the BIM analysis.

All QTL intervals were searched for genes (coding and non-coding) using Biomart (vectorbase.org). If no genes were found within the QTL, the surrounding regions up to 10 kb away were searched with the idea that the QTL may be a regulatory region for a downstream gene. Overrepresented Gene Ontology terms from the resulting genes was found using PantherDB's (pantherdb.org) statistical overrepresentation test.

Results and Discussion

Crosses and Phenotyping

We used the *An. coluzzii* MALI-NIH (founded in 2005 from Niono, Mali) and *An. merus* MAF (founded in 1991 from Kruger National Park, South Africa) colonies for all experiments. These were chosen to avoid – to the extent possible – inversion differences that would suppress recombination in hybrids. The colonies only differ by the compound *2Rop* inversion, which distinguishes the species (Coluzzi et al., 2002). The two linked inversions span ~25 megabases (Mb) or slightly less than 1/10th of the genome (Kamali et al., 2012). We scored fertility of individual male mosquitoes by dissecting the testes from five to ten day old virgins and examining the quantity and morphology of sperm under a compound microscope. Our classification system closely follows Slotman *et al* (Slotman et al., 2004) who studied the genetics of sterility in *An. coluzzii* and *An. arabiensis* hybrids; briefly, males are assigned an integer of 1 (sterile) to 5 (fertile) based on the abundance of mature sperm and the ratio of normal to abnormal sperm.

We confirmed the full fertility of males from both species by phenotyping 50 mosquitoes from each colony. Next, we performed reciprocal crosses en masse and confirmed the complete sterility of F1 hybrid males (n = 50). We then backcrossed fertile F1 females to males of their maternal parent species. Figure 1A-B diagrams the chromosomal complement of each backcross, which we subsequently refer to as col^{BC} and mer^{BC}. Progeny from both backcrosses

displayed the complete spectrum of phenotypes from fully fertile to completely sterile (Figure 1C-D). High numbers of intermediate phenotypes in the backcrosses suggests that male sterility is a polygenic, complex trait. Additionally, differences in phenotypic distribution between the two backcrosses point to different DMIs causing sterility in reciprocal F1 hybrids.

Genotyping by Sequencing and Ancestry Inference

The range of phenotypes present in both backcrosses provides the foundation for quantitative trait locus (QTL) mapping of hybrid sterility. The col^{BC} males can be used to map *An. merus* loci that are incompatible in a primarily *An. coluzzii* genetic background, while the mer^{BC} males can be used to discover *An. coluzzii* loci that cause sterility in an *An. merus* genetic background. In order to link genotype with phenotype and identify genomic regions that contain hybrid sterility genes, we need to delineate crossover breakpoints in thousands of backcross progeny at extremely high resolution. Whole genome sequencing of individual mosquitoes would provide the highest resolution, but is impractical for such a large number of individuals. Instead, we utilized the double-digest restriction associated DNA sequencing (ddRADseq) methodology to genotype individual mosquitoes in parallel (Peterson et al., 2012). Briefly, DNA from individual mosquitoes was extracted, digested with two restriction enzymes, barcoded, size selected, and PCR amplified. The multiplexed RAD fragments were then sequenced on an Illumina HiSeq2500. SNPs from the fragments

provide a representative and reproducible set of genetic markers (Davey et al., 2011).

We performed deep ddRAD sequencing on 48 *An. coluzzii* and *An. merus* parental females and identified 12,408 putative fixed SNP differences between colonies. Backcross progeny were sequenced at a lower depth with a mean of 3,208 informative SNPs per individual. Due to variable coverage between individuals and among SNPs, genotype calls will often be made based on a small number of reads (McKenna et al., 2010). Such low coverage will result in true heterozygous SNPs being incorrectly genotyped as homozygous. For instance, if a SNP has 3x coverage and the backcross individual is a true heterozygote, there is a ¼ chance that one allele will be exclusively sequenced in all three reads, leading to an incorrect homozygous genotype call. To correct for erroneous genotyping resulting from low coverage, we developed a custom Hidden-Markov model (HMM) that estimates the probability that a given SNP is heterozygous or homozygous based on coverage depth, surrounding alleles, and chromosomal crossover rates (Figure 2.8). The HMM detects crossover breakpoints at extremely high resolution (Figure 2.9), producing illuminating plots (Figures 2.2 and 2.3) that allow us to directly visualize the portions of its genome a mosquito inherited from each parent.

Recombination Rates and Segregation Distortion

Transitions in ancestry as indicated by the HMM were used to construct recombination rate maps (Figure 2.10). Map distances for the three chromosomes (2, 87.0 cM; 3, 85.5 cM; X, 12.9 cM) were similar to results previously reported for intraspecific recombination within *An. gambiae/coluzzii* (Zheng et al., 1996). On all chromosomes, recombination frequency positively correlated with physical distance from heterochromatic centromeres (Pombi et al., 2006; Sharakhova and Sharakhov, 2010; Slotman et al., 2006) and virtually no crossovers occurred between the fixed inversions on 2R (Stump et al., 2007). Crossing over was less frequent on the X than the autosomes. Due to the small size of the X chromosome, mean physical distance of markers from the centromere is much shorter than on the autosomes, which alone could account for the reduction in X-linked crossovers.

Examination of ancestry graphs for individual backcross progeny, grouped by phenotypic class, reveals strong transmission distortion of the X chromosome, but not the autosomes, in both backcrosses (Figures 2.2 and 2.3). In the col^{BC}, the *An. coluzzii* homozygous or hemizygous genotype is strikingly more common on the X (78.6%), but at or near Mendelian ratios on the autosomes (2, 50.0%; 3, 48.2%). In contrast, mer^{BC} progeny exhibit X-linked transmission distortion in the opposite direction; the *An. merus* hemizygous genotype is overrepresented on the X (81.5%). Similar to the col^{BC}, we observe more normal frequencies of homozygosity on mer^{BC} autosomes (2, 43.7%; 3, 46.8%). Since distortion occurs

in opposite directions in the reciprocal crosses, it is likely attributable to inviability of males harboring certain allelic combinations. Indeed, increased distortion of the X in both crosses is consistent with incompatibilities between a hemizygous X and foreign, recessive autosomal alleles that lead to decreased viability.

Quantitative Genomics of Sterility: CxM^{F1} Hybrids

We hypothesized that sterility in *An. coluzzii x An. merus* F1 (*CxM^{F1}*) hybrids is due to incompatibilities between hemizygous *An. coluzzii* X-linked and dominant *An. merus* autosomal factors. To investigate if the *An. coluzzii* X chromosome causes sterility in a heterozygous or *An. merus* homozygous autosomal background we examined ancestry graphs of *mer^{BC}* progeny (Figure 2.3). In fact, we observe a near deterministic relationship between X chromosome genotype and sterility in *mer^{BC}* progeny. When *mer^{BC}* males inherit a fully *An. coluzzii* X they are always sterile. Additionally, individuals inheriting even a partial copy of the *An. coluzzii* X are usually sterile and no fully fertile individuals are hemizygous for any *An. coluzzii* X-linked loci. Such a striking pattern suggests that a simple incompatibility between a dominant *An. merus* autosomal factor(s) and the *An. coluzzii* X chromosome could underlie sterility in *mer^{BC}* males. Infrequent X-linked crossovers and extreme transmission distortion prevent us from mapping specific regions of the *An. coluzzii* X associated with sterility. However, since inheritance of any *An. coluzzii* X chromosomal segments

decreases fertility in *mer*^{BC} progeny, it is likely that sterility factors are dense across the chromosome.

Next, I wanted to identify dominant *An. merus* autosomal factors that may negatively interact with the *An. coluzzii* X. To find such factors, we created ancestry graphs for the subset of *col*^{BC} progeny who inherited a non-recombined *An. coluzzii* X (Figure 2.4A). The *col*^{BC} progeny are ideal for the analysis because they possess either *An. coluzzii* homozygous or heterozygous genotypes across their autosomes, allowing us to search for autosomal regions where inheritance of a single *An. merus* allele causes sterility. In this subset of *col*^{BC} individuals, percent heterozygosity across chromosome 2 is positively correlated with male sterility (r^2 = 0.38, $P < 1 \times 10^{-15}$). Chromosome 3 genotype is also associated with sterility (r^2 = 0.08, $P < 1 \times 10^{-9}$), but to a smaller degree than chromosome 2. (Figure 2.4B).

The above ancestry analysis indicates that dominant alleles on both *An. merus* autosomes may be incompatible with the *An. coluzzii* X resulting in sterility. In order to localize *An. merus* autosomal incompatibilities underlying sterility, we performed both Bayesian and Composite Interval Mapping (BIM and CIM) on the autosomes of all *col*^{BC} individuals that possess a fully *An. coluzzii* X. Both models identified the same three major QTLs (labeled as A, B, and C in Figure 5. QTLs A and B are located on 2L and 2R, respectively, while QTL C is located on 3L. These three QTL are independently responsible for 4.198%, 3.119%, and 4.145% (BIM), respectively, of the variance in male fertility. When

modeled with R/qtlbim, these three major QTL, along with three more minor QTL, make up 48.468% of the variance in male fertility (Table 2.1). Both major QTL on chromosome 2 are located in regions of reduced recombination. QTL A abuts the centromere of chromosome 2, while QTL B is coincident with the *2Rop* inversion. Genomic regions that exhibit low interspecific recombination may preferentially accumulate sterility DMIs (Kirkpatrick, 2006; Noor et al., 2001a), especially in the presence of gene flow, but are also relatively easy to map and prone to effect size overestimation (Noor et al., 2001b). In contrast, QTL C is located in the middle of 3L in a region that exhibits no significant deficit in recombination. Estimated effect sizes, genomic positions, and peaks for each QTL are listed in Tables 2.1, 2.2, and 2.3. In sum, both QTL models support the ancestry plots in finding that dominant *An. merus* autosomal factors on chromosome 2 and, to a lesser extent, chromosome 3 interact with the *An. coluzzii* X chromosome to cause sterility. The presence of multiple QTL suggests the genetic architecture of male sterility is relatively complex, though not intractable.

The 1.5 LOD interval around BIM QTL, and the threshold interval around CIM QTL, were then used to discover the number and types of genes overrepresented in each col^{BC} QTL (Table 2.4, 2.10-2.12). Only those QTL found in the BIM model were looked at for gene number in BIM and CIM. For the major QTL, QTL A only had one gene found within the 1.5 LOD interval, but 911 and 900 genes found using the CIM analysis, imputation (IMP) method and haley-knott (HK) method respectively. In QTL B there were 849 genes found in the BIM

interval, and 2052 genes found with either CIM method. The last major QTL, QTL C, had 273 genes for either CIM method, but no full genes where found in the BIM interval, though two partial genes were found in this interval.

Gene ontology (GO) analysis was then done using all genes in all significant QTL in BIM and all QTL with an LOD above threshold in CIM (six total QTL in BIM, five total QTL for CIM) using *An. gambiae* gene IDs that were orthologous to *An. coluzzii* or *An. merus* (Table 2.10 - 2.12). Using an overrepresentation test we found the most represented GO terms found in the significant QTLs. The CIM QTL from both the IMP and HK methods had only one underrepresented category for each GO family, which was “Unclassified” (Table 2.11 and 2.12). The QTL found by BIM, however, found significantly overrepresented terms in the Biological Process family, the most enriched being “intracellular cholesterol transport” (12.87 Fold enrichment (FE) *An. coluzzii* orthologs, 12.85 FE *An. merus* orthologs) (Table 2.10). The one gene found in QTL A is in the Panther Family/Subfamily “Vacuolar Protein Sorting-Associated Protein VPS13”, which are implicated in human hereditary disorders chorea acanthocytosis and Cohen syndrome and is involved in delivery of proteins to the vacuole, and other membrane-related functions, in vegetatively growing yeast (Park et al., 2012). Of the two partial genes found within the QTL C BIM interval, both have orthologs in both *An.merus* and *An. coluzzii,* though only one has a GO term associated with it. This gene, sulfakinin neuropeptide, is similar to *Drosophila* drosulfakinins and mammalian gastrin and CCK that are involved in

feeding processes (Hauser, 2006). The other gene is not well described in any *Anopheles* species, though has domains that fall within the Piezo family, whose protein are subunits of pore-forming mechanically-activated channels (Coste et al., 2012). In *Drosophila*, these channels are involved in sensing noxious mechanical stimuli (Yan et al., 2012). While GO analysis gives insight into possible functional properties of the genes in our QTL, our most intriguing QTL either have few genes or are in areas that have low amounts of recombination between species (2Rop inversion) making it difficult to determine precisely what genes are involved in causing sterility.

Quantitative Genomics of Sterility: MxC^{F1} Hybrids

Just as before, we hypothesized that sterility in *An. merus x An. coluzzii* F1 (*MxC^{F1}*) hybrids is due to X-autosomal incompatibilities. In this reciprocal cross, hemizygous *An. merus* X alleles may negatively interact with dominant *An. coluzzii* autosomal factors resulting in reduced fertility. First, we examined ancestry graphs of *col^{BC}* progeny to evaluate if the *An. merus* X chromosome causes sterility when in a heterozygous or *An. coluzzii* homozygous autosomal background (Figure 2.2). Interestingly, multiple *col^{BC}* males with an *An. merus* X are either fully or partially fertile. Since all *col^{BC}* progeny are either heterozygous or homozygous for *An. coluzzii* alleles on the autosomes, we would expect to see complete sterility in *col^{BC}* males who inherit an *An. merus* X if simple autosomal-X incompatibilities cause *MxC^{F1}* hybrid sterility. Instead, the results show that the

genetic architecture of sterility in *MxC*^{F1} hybrids is highly complex, involving dominant autosomal incompatibilities from both species interacting with each other and with the *An. merus* X.

To further evaluate the role of each autosome on *MxC*^{F1} hybrid sterility we conducted ancestry analysis on the subset of *mer*^{BC} progeny that inherited an *An. merus* X. The *mer*^{BC} progeny are either homozygous for *An. merus* alleles or heterozygous across the autosomes. Since multiple dominant *An. coluzzii* factors appear to negatively interact with each other and the *An. merus* X, we expected that the proportion of heterozygosity across the autosomes should positively correlate with sterility (Figure 2.6A). As predicted, we see a statistically significant correlation between sterility and percent heterozygosity of both autosomes (2: $r^2 = 0.28$, $P < 1 \times 10^{-15}$; 3: $r^2 = 0.14$, $P < 1 \times 10^{-15}$) with males in the two most sterile phenotypic classes heterozygous across the majority of chromosomes 2 (81.9%) and 3 (66.5%) (Figure 2.6B).

As before, low recombination rates and transmission distortion prevent us from determining which regions of the *An. merus* X are most important to sterility in *MxC*^{F1} hybrids. In order to localize dominant *An. coluzzii* autosomal alleles that negatively interact with the *An. merus* X we performed QTL mapping on all *mer*^{BC} males that inherited an *An. merus* X. BIM detected three major QTL that overlap with three largest effect QTL found using CIM. QTLs A and B are located on 2L and 2R, respectively, while QTL C is located on 3L (BIM) or 3R (CIM), though both BIM and CIM peaks are found within the BIM's 1.5 LOD interval that spans

the centromere. Together the three QTL span ~8 Mb (BIM) or ~59 Mb (CIM) of the autosomes and explain 7.99% (A), 12.18% (B), and 10.04% (C) of the variance in male fertility. As before, two of the QTL are located in regions of highly reduced recombination: QTL B is coincident with the *2Rop* inversion, while QTL C straddles the centromere of chromosome 3. However, QTL A is located on 2L in a region with no apparent reduction in recombination and, not surprisingly, was mapped to a much narrower interval than QTL B or C in CIM. In agreement with complex genetic control of sterility, BIM also identified four minor QTL, 2 of which act epistatically with major QTLs B and C (QTL 2.3 interacts with QTL C, QTL B interacts with QTL 3.4). Taken together, the three major and four minor QTL, and their interactions, can account for 60.07% of the variance in male fertility. Effect sizes, locations, and peaks for individual QTL are provided in Tables 2.5 - 2.8. Despite the apparent highly epistatic and polygenic basis of sterility in MxC^{F1} hybrids, we were able to localize genomic regions that have a major influence on sterility providing optimism that individual sterility genes can be fine mapped.

As with col^{BC}, the 1.5 LOD interval around BIM QTL, and the threshold interval around CIM QTL, were then used to discover the number and types of genes overrepresented in each mer^{BC} QTL (Table 2.9, and 2.13 - 2.14). Only those QTL found in the BIM model were looked at for gene number in BIM and CIM. For the major QTL, QTL A had 103 genes found within the 1.5 LOD interval, but 230 and 257 genes found using the CIM analysis using the IMP and HK

methods, respectively. In QTL B there were only 2 partial genes found in the BIM interval, and 2065 genes found with either CIM method. The last major QTL, QTL C, had 273 genes in the BIM interval, and 1259 genes for either CIM method.

Gene ontology (GO) analysis was then done using all genes in all significant QTL in BIM and all QTL with an LOD above threshold in CIM (seven total QTL in BIM, five total QTL for CIM) using *An. gambiae* gene IDs that were orthologous to *An. coluzzii* or *An.merus.* Using an overrepresentation test we found the most overrepresented or underrepresented GO terms found in the significant QTLs that are shown in Tables 2.13, 2.14, and 2.15. The most overrepresented GO term found in the BIM QTL was the Molecular Function term “structural constituent of the cuticle” (13.17 FE *An. coluzzii* orthologs, 9.34 FE *An.merus* orthologs).The highest FE for either of the CIM method QTL was only 2 for the Molecular Function GO term “monooxygenase activity” in *An. coluzzii* orthologs using the HK method (FE for *An. merus* orthologs was 1.97). Of the two partial genes found in QTL B, one is in the subfamily Von Willebrand Factor X Domain-Containing Protein 7, and the other is an aminoacyl-tRNA synthetase in the subfamily Methionine- tRNA ligase and is a mitochondrial protein. Interestingly, the Von Willebrand Factor X Domain-Containing Protein 7 subfamily is implicated in testicular disease and orchitis in humans, both are diseases that affect the testes (Snoek et al., 1998).

The Large X-Effect

Our data leads to mixed conclusions with regards to the large X-effect. First, sterility in MxC^{F1} does not follow the rule; the autosomes play an equal, or even more important role, than the X in sterility (Figure 2.11). This result is surprising since backcross schemes are inherently biased towards finding large X-effects since male progeny are hemizygous for foreign alleles on the X but only heterozygous on the autosomes (Hollocher and Wu, 1996; Wu and Davis, 1993). We do find evidence for a large X-effect in CxM^{F1} hybrids (Figure 2.11). A well-controlled introgression study of sterility between *Drosophila mauritiana* and *Drosophila sechellia* demonstrated that, at least in these two species, the large X-effect is due to a higher density of X-linked versus autosomal DMIs (Masly and Presgraves, 2007). A similar pattern has been inferred for mouse subspecies sterility through both experimental crosses in captivity and studies of introgression in natural hybrid zones (Good et al., 2008; Payseur et al., 2004; Teeter et al., 2007). It is possible that the large X-effect we see in CxM^{F1} hybrids is caused by the failure of X chromosome inactivation during spermatogenesis (Lifschytz and Lindsley, 1972).

Summary and Next Steps

Using high-throughput quantitative genomics we provide evidence that the Large X-effect underlies male sterility in one direction of a reciprocal cross

between *An. coluzzii* and *An. merus*. Our working hypothesis is that dominant *An. merus* autosomal factors may influence the expression of *An. coluzzii* X-linked genes in the male germline, while epistatic autosomal interactions between *An. merus* and *An. coluzzii* play a larger role in the sterility of the reciprocal cross. This will be further tested using transcriptomic techniques. Direct implication of X misregulation as the casual factor in CxM^{F1} male sterility will be challenging and likely requires a better understanding of *An. coluzzii* spermatogenesis at a molecular and cellular level. It is, however, feasible that some autosomal QTL can be further narrowed with genetic analysis of advanced backcross progeny.

Acknowledgements

I'd like to acknowledge Dr. David Turissini for the genotype analysis (including writing the scripts for finding fixed species differences and the HMM) and recombination analysis.

References

Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T.T., Mast, J., Sunayama-Morita, T., and Stern, D.L. (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res. *21*, 610–617.

Arends, D., Prins, P., Jansen, R.C., and Broman, K.W. (2010). R/qtl: high-throughput multiple QTL mapping. Bioinformatics *26*, 2990–2992.

Assogba, B.S., Djogbenou, L., Saizonou, J., Diabate, A., Dabire, A., Moiroux, N., Gilles, J.R., Makoutode, M., and Baldet, T. (2014). Characterization of swarming and mating behavior between Anopheles coluzzii and Anopheles melas in a sympatry are of Benin. Acta Trop. *132*, S53–S63.

Atlan, A., Capillon, C., Derome, N., Couvet, D., and Montchamp-Moreau, C. (2003). The evolution of autosomal suppressors of sex-ratio drive in Drosophila simulans. Genetica *117*, 47–58.

Benedict, M., Knols, B., Bossin, H., Howell, P., Mialhe, E., Caceres, C., and Robinson, A. (2009). Colonisation and mass rearing: learning from others. Malar. J. *8*, S2–S4.

Bock, I.R. (1984). Interspecific hybridization in the genus Drosophila. Evol. Biol. 41–70.

Broman, K.W., Wu, H., Sen, S., and Churchill, G.A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics *19*, 889–890.

Brothers, A.N., and Delph, L.F. (2010). Haldane's rule is extended to plants with sex chromosomes. Evolution *64*, 3643–3648.

Coetzee, M., Hunt, R.H., Wilkerson, R., Della Torre, A., Coulibaly, M.B., and Besansky, N.J. (2013). Anopheles coluzzii and Anopheles amharicus, new members of the Anopheles gambiae complex. Zootaxa *3619*, 246–274.

Coluzzi, M., Sabatini, A., Della Torre, A., Di Deco, M.A., and Petrarca, V. (2002). A polytene chromosome analysis of the Anopheles gambiae species complex. Science *298*, 1415–1418.

Coluzzii, M., Sabatini, A., Petrarca, V., and Di Deco, M.A. (1979). Chromosomal differentiation and adaptation to human environments in the Anopheles gambiae complex. Trans. R. Soc. Trop. Med. Hyg. *73*, 483–497.

Coste, B., Xiao, B., Santos, J.S., Syeda, R., Grandl, J., Spencer, K.S., Kim, S.E., Schmidt, M., Mathur, J., Dubin, A.E., et al. (2012). Piezo proteins are pore-forming subunits of mechanically activated channels. Nature *483*, 176–181.

Counterman, B.A., Ortiz-Barrientos, D., and Noor, M.A. (2004). Using comparative genomic data to test for fast-X evolution. Evolution *58*, 656–660.

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). Two rules of speciation. Speciat. Its Consequences 180–207.

Coyne, J.A., and Orr, H.A. (2004). Speciation (Sinauer Associates, Inc.).

Coyne, J.A., Charlesworth, B., and Orr, H.A. (1991). Haldane's rule revisited. Evolution 1710–1714.

Craft, W.A. (1938). The sex ratio in mules and other hybrid mammals. Q. Rev. Biol. *13*, 19–40.

Cuamba, N., and Mendis, C. (2009). The role of Anopheles merus in malaria transmission in an area of southern Mozambique. J. Vector Borne Dis. *46*, 157–159.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., and Blaxter, M.L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. *12*, 499–510.

Davidson, G. (1964). The Five Mating-Types in the Anopheles Gambiae Complex. Riv Malariol *43*, 167–183.

Donnelly, M.J., Pinto, J., Girod, R., Besansky, N.J., and Lehmann, T. (2004). Revisiting the role of introgression vs. shared ancestral polymorphisms as key processes shaping genetic diversity in the recently separated sibling species of the Anopheles gambiae complex. Heredity *92*, 61–68.

Elliott, R.J., Aggoun, L., and Moore, J.B. (1994). Hidden Markov Models (Springer).

Fontaine, M.C., Pease, J.B., Steele, A., Waterhouse, R.M., Neafsey, D.E., Sharakhov, I.V., Jiang, X., Hall, A.B., Catteruccia, F., Kakani, E., et al. (2015). Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science *347*, 1258524–1258524.

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

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.

Govere, J., Durrheim, D., Coetzee, M., Hunt, R.H., and La Grange, J. (2000). Captures of mosquitoes of the Anopheles gambiae complex (Diptera: Culicidae) in the Lowveld region of Mpumalanga Province, South Africa. Afr. Entomol. *8*, 91–99.

Gray, A.P. (1958). Bird hybrids. A check-list with bibliography.

Gray, A.P. (1972). Mammalian hybrids. A check-list with bibliography. In Technical Communications, (Commonwealth Bureau of Animal Breeding and Genetics), p.

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

Hauser, F. (2006). Identifying neuropeptide and protein hormone receptors in Drosophila melanogaster by exploiting genomic data. Brief. Funct. Genomic. Proteomic. *4*, 321–330.

Hillers, K.J. (2004). Crossover interference. Curr. Biol. *14*, R1036–R1037.

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

Hunt, R.H., Coetzee, M., and Fettene, M. (1998). The Anopheles gambiae complex: a new species from Ethiopia. Trans. R. Soc. Trop. Med. Hyg. *92*, 231–235.

Hurst, L.D., and Pomiankowski, A. (1991). Causes of Sex Ratio Bias May Account for Unisexual Sterility in Hybrids: A New Explanation of Haldane's Rule and Related Phenomena. Genetics *128*, 841–858.

Jaenike, J. (2001). Sex chromosome meiotic drive. Annu. Rev. Ecol. Syst. 25–49.

Kamali, M., Xia, A., Tu, Z., and Sharakhov, I.V. (2012). A new chromosomal phylogeny supports the repeated origin of vectorial capacity in malaria mosquitoes of the Anopheles gambiae complex. Plos Pathog. *8*, e1002960.

Kirkpatrick, M. (2006). Chromosome Inversions, Local Adaptation and Speciation. Genetics *173*, 419–434.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.

Lifschytz, E., and Lindsley, D.L. (1972). The Role of X-Chromosome Inactivation during Spermatogenesis. Proc. Natl. Acad. Sci. *69*, 182–186.

Liu, B.H. (1997). Statistical genomics: linkage, mapping, and QTL analysis (CRC Press).

Llopart, A. (2012). The rapid evolution of X-linked male-biased gene expression and the large-X effect in Drosophila yakuba, D. santomea, and their hybrids. Mol. Biol. Evol. *29*, 3873–3886.

Marchand, R.P. (1984). Field Observations on Swarming and Mating in Anopheles gambiae Mosquitoes in Tanzania. Neth J Zool *34*, 367–387.

Martinez, O., and Curnow, R.N. (1992). Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. *85*, 480–488.

Masly, J.P., and Presgraves, D.C. (2007). High-resolution genome-wide dissection of the two rule of speciation in Drosophila. PLOS Biol. *5*, e243.

McDermott, S.R., and Noor, M.A. (2010). The role of meiotic drive in hybrid male sterility. Philos. Trans. R. Soc. B Biol. Sci. *365*, 1265–1272.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

Meisel, R.P., and Connallon, T. (2013). The faster-X effect: integrating theory and data. Trends Genet. *29*, 537–544.

Muller, H. (1942). Isolating mechanisms, evolution and temperature. Biol Symp 71–125.

Mwangangi, J.M., Mbogo, C., Nzovu, J.G., Githure, J.I., Yan, G., and Beier, J.C. (2003). Blood-meal analysis for anopheline mosquitoes sampled along the Kenyan coast. J. Am. Mosq. Control Assoc. *19*, 371–375.

Nolan, T., Papathanos, P., Windbichler, N., Magnusson, K., Benton, J., Catteruccia, F., and Crisanti, A. (2011). Developing transgenic Anopheles mosquitoes for the sterile insect technique. Genetica *139*, 33–39.

Noor, M.A., Grams, K.L., Bertucci, L.A., and Reiland, J. (2001a). Chromosomal inversions and the reproductive isolation of species. Proc. Natl. Acad. Sci. *98*, 12084–12088.

Noor, M.A., Cunningham, A.L., and Larkin, J.C. (2001b). Consequences of recombination rate variation on quantitative trait locus mapping studies: simulations based on the Drosophila melanogaster genome. Genetics *159*, 581–588.

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 Turelli, M. Dominance and Haldane's rule. Genetics *143*, 613.

Park, J.-S., Lee, S.-H., Na, H.-J., Pyo, J.-H., Kim, Y.-S., and Yoo, M.-A. (2012). Age- and oxidative stress-induced DNA damage in Drosophila intestinal stem cells as marked by Gamma-H2AX. Exp. Gerontol. *47*, 401–405.

Payseur, B.A., Krenz, J.G., and Nachman, M.W. (2004). Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. Evolution *58*, 2064–2078.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., and Hoekstra, H.E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE *7*, e37135.

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

Pock Tsy, J.-M.L., Duchemin, J.-B., Marrama, L., Rabarison, P., Le Goff, G., Rajaonarivelo, V., and Robert, V. (2003). Distribution of the species of the Anopheles gambiae complex and the first evidence of Anopheles merus as a malaria vector in Madagascar. Malar. J. *2*, 1–7.

Pombi, M., Stump, A.D., Della Torre, A., and Besansky, N.J. (2006). Variation in recombination rate across the X chromosome of Anopheles gambiae. Am J Trop Med Hyg *75*, 901–903.

Presgraves, D.C. (2002). Patterns of postzygotic isolation in Lepidoptera. Evolution *56*, 1168–1183.

Presgraves, D.C. (2008). Sex chromosomes and speciation in Drosophila. Trends Genet. *24*, 336–343.

Presgraves, D.C. (2010). The molecular evolutionary basis of species formation. Nat. Rev. Genet. *11*, 175–180.

Presgraves, D.C., and Orr, H.A. (1998). Haldane's rule in taxa lacking a hemizygous X. Science *282*, 952–954.

Ranson, H., N'Guessan, R., Lines, J., Moiroux, N., Nkuni, Z., and Corbel, V. (2011). Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? Trends Parasitol. *27*, 91–98.

Reddy, M.R., Overgaard, H.J., Abaga, S., Reddy, V.P., Caccone, A., Kiszewski, A.E., and Slotman, M.A. (2011). Outdoor host seeking behaviour of Anopheles gambiae mosquitoes following initiation of malaria vector control on Bioko Island, Equatorial Guinea. Malar. J. *10*, 184.

Schilthuizen, M., Giesbers, M., and Beukeboom, L.W. (2011). Haldane's rule in the 21st century. Heredity *107*, 95–102.

Seaton, G., Haley, C.S., Knott, S.A., Kearsey, M., and Visscher, P.M. (2002). QTL Express: mapping quantitative trait loci in simple and complex pedigrees. Bioinformatics *18*, 339–340.

Sharakhova, M.V., and Sharakhov, I.V. (2010). Organization and evolution of heterochromatin in malaria mosquitoes. Russ. J. Genet. *46*, 1250–1253.

Singh, N.D., Larracuente, A.M., and Clark, A.G. (2008). Contrasting the efficacy of selection on the X and autosomes in Drosophila. Mol. Biol. Evol. *25*, 454–467.

Sinka, M.E., Bangs, M.J., Manguin, S., Coetzee, M., Mbogo, C.M., Hemingway, J., Patil, A.P., Temperley, W.H., Gething, P.W., Kabaria, C.W., et al. (2010). The dominant Anopheles vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis.

Slotman, M., Della Torre, A., and Powell, J.R. (2004). The genetics of inviability and male sterility in hybrids between Anopheles gambiae and An. arabiensis. Genetics *167*, 275–287.

Slotman, M.A., Reimer, L.J., Thiemann, T., Dolo, G., Fondjo, E., and Lanzaro, G.C. (2006). Reduced Recombination Rate and Genetic Differentiation Between the M and S Forms of Anopheles gambiae s.s. Genetics *174*, 2081–2093.

Snoek, M., Teuscher, C., and van Vugt, H. (1998). Molecular analysis of the major MHC recombinational hot spot located within the G7c gene of the murine class III region that is involved in disease susceptibility. J. Immunol. *160*, 266–272.

Stump, A.D., Pombi, M., Goeddel, L., Ribeiro, J.M.C., Wilder, J.A., Torre, A.D., and Besansky, N.J. (2007). Genetic exchange in 2La inversion heterokaryotypes of Anopheles gambiae. Insect Mol. Biol. *16*, 703–709.

Tao, Y., Masly, J.P., Ke, Y., and Hartl, D.L. (2007b). a sex-ratio meiotic drive system in Drosophila simulans I: An autosomal suppressor. PLOS Biol. *5*, e292.

Tao, Y., Malsy, J.P., Araripe, L., Ke, Y., and Hartl, D. (2007a). a sex-ratio meiotic drive system in Drosophila simulans II: an X-linked distorter. PLOS Biol. *5*, e293.

Teeter, K.C., Payseur, B.A., Harris, L.W., Bakewell, M.A., Thibodeau, L.M., O'Brien, J.E., Krenz, J.G., Sans-Fuentes, M.A., Nachman, M.W., and Tucker, P.K. (2007). Genome-wide patterns of gene flow across a house mouse hybrid zone. Genome Res. *18*, 67–76.

Temu, E.A., Minjas, J.N., Coetzee, M., Hunt, R.H., and Shift, C.J. (1998). The role of four anopheline species (Diptera: Culicidae) in malaria transmission in coastal Tanzania. Trans. R. Soc. Trop. Med. Hyg. *92*, 152–158.

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

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

Wang, S., and Jacobs-Lorena, M. (2013). Genetic approaches to interfere with malaria transmission by vector mosquitoes. Trends Biotechnol. *31*, 185–193.

Watson, E.T., and Demuth, J.P. (2012). Haldane's Rule in Marsupials: What Happens When Both Sexes Are Functionally Hemizygous? J. Hered. *103*, 453–458.

Weetman, D., Steen, K., Rippon, E.J., Mawejje, H.D., Donnelly, M.J., and Wilding, C.S. (2014). Contemporary gene flow between wild An. gambiae ss and An. arabiensis. Parasit. Vectors *7*, 345.

White, G.B. (1971). Chromosomal evidence for natural interspecific hybridization by mosquitoes of the Anopheles gambiae complex. Nature *231*, 184–185.

White, G.B. (1973). Comparative Studies on Sibling Species of Anopheles-Gambiae Giles Complex (Dipt Culicidae) 3. Distribution, Ecology, Behavior and Vectorial Importance of Species-D in Bwamba-County, Uganda, with an Analysis of Biological, Ecological, Morphological and Cytogenetical Relationships of Ugandan Species-D. Bull. Entomol. Res. *63*, 65–97.

White, B.J., Collins, F.H., and Besansky, N.J. (2011). Evolution of Anopheles gambiae in Relation to Humans and Malaria. Annu. Rev. Ecol. Syst. *42*, 111–132.

White, B.J., Kundert, P.N., Turissini, D.A., Van Ekeris, L., Linser, P.J., and Besansky, N.J. (2013). Dose and developmental responses of Anopheles merus larvae to salinity. J. Exp. Biol. *216*, 3433–3441.

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

Wu, C.-I., Johnson, N.A., and Palopoli, M.F. (1996). Haldane's rule and its legacy: Why are there so many sterile males? Trends Ecol. Evol. *11*, 281–284.

Yan, Z., Zhang, W., He, Y., Gorczyca, D., Xiang, Y., Cheng, L.E., Meltzer, S., Jan, L.Y., and Jan, Y.N. (2012). Drosophila NOMPC is a mechanotransduction channel subunit for gentle-touch sensation. Nature *493*, 221–225.

Zheng, L., Benedict, M., Cornel, A.J., Collins, F.H., and Kafatos, F.C. (1996). An integrated genetic map of the African human malaria vector mosquito, Anopheles gambiae. Genetics *143*, 941–952.

FIGURES

Figure Legends:

Figure 2.1 Backcross progeny provide the phenotypic variation needed for genetic mapping of sterility. A) *Anopheles coluzzii* backcross male progeny (*col*^{BC}) will inherit one recombined autosome and one *An. coluzzii* autosome; a recombined X chromosome; and a non-recombined *An. coluzzii* Y chromosome. B) *Anopheles merus* backcross male progeny (*mer*^{BC}) will inherit one recombined autosome and one *An. merus* autosome; a recombined X chromosome; and a non-recombined *An. merus* Y chromosome. C and D) Male progeny from both backcrosses display a spectrum of phenotypes from fully sterile to fully fertile providing the foundation for QTL mapping.

Figure 2.2 Ancestry reconstruction of *An. coluzzii* backcross progeny elucidates genomic regions associated with sterility. A hidden Markov model was used to reconstruct ancestry for 787 *col*^{BC} males from ddRADseq data. Inferred genotypes for chromosome 2 (left), chromosome 3 (center), and the X (right) are horizontally plotted for each mosquito. Red is homozygous or hemizygous for *An. coluzzii*, blue is heterozygous for one *An. coluzzii* allele and one *An. merus* allele, and green is hemizygous for *An. merus*. Mosquitoes are grouped by phenotype and crossover breakpoints can be visualized with high resolution. On the autosomes a strong correlation between genotype and sterility is evident, while X chromosome genotype appears to be a weaker predictor of phenotype.

Figure 2.3 Ancestry reconstruction of *An. merus* backcross progeny elucidates genomic regions associated with sterility. A hidden Markov model was used to reconstruct ancestry for 900 *mer*^{BC} males from ddRADseq data. Inferred genotypes across chromosome 2 (left), chromosome 3 (center), and the X (right) are horizontally plotted for each mosquito. Green is homozygous or hemizygous for *An. merus*, blue is heterozygous for one *An. merus* allele and one *An. coluzzii* allele, and red is hemizygous for *An. coluzzii*. Mosquitoes are grouped by phenotype and crossover breakpoints are mapped at high resolution. A striking correlation between X chromosome ancestry and sterility is evident. To a lesser degree, autosomal genotype also correlates with sterility.

Figure 2.4 Dominant *An. merus* factors primarily on chromosome 2 interact with the *An. coluzzii* X to cause sterility. A) Ancestry maps for *col*^{BC} progeny that inherited a fully *An. coluzzii* X reveal a strong correlation between heterozygosity of chromosome 2 and sterility. Chromosome 3 genotype is also correlated with sterility, but to a lesser degree. B) Grey circles denote the percentage of chromosome 2 (left) and chromosome 3 (right) that is heterozoygous for each mosquito included in the ancestry map. The average heterozygosity (+/- SEM) for each phenotypic class is marked by a black line. A red regression line shows the overall relationship between phenotype and genotype.

Figure 2.5 Three heterozygous autosomal genomic regions cause sterility when paired with an *An. coluzzii* X. QTL mapping was performed on *col*^{BC} individuals with a fully *An. coluzzii* X. (A) Composite and (B) Bayesian interval mapping methods identified overlapping major QTL for sterility on chromosome 2 (marked A and B) and chromosome 3 (marked C). Presumably, dominant *An. merus* alleles in each QTL are incompatible with an *An. coluzzii* X. Thresholds of significance for logs of odd ratio (LOD) and log posterior density (LPD) are denoted by a red line. Bayesian mapping reports many significant QTL not identified using a maximum-likelihood framework.

Figure 2.6 Dominant *An. coluzzii* factors on both autosomes interact with the *An. merus* X to cause sterility. A) Ancestry maps for *mer*^{BC} progeny that inherited a fully *An. merus* X reveal a strong correlation between heterozygosity of both autosomes and sterility. B) Grey circles denote the percentage of chromosome 2 (left) and chromosome 3 (right) that is heterozoygous for each mosquito included in the ancestry map. The average heterozygosity (+/- SEM) for each phenotypic class is marked by a black line. A red regression line shows the overall relationship between phenotype and genotype.

Figure 2.7 Three heterozygous autosomal genomic regions cause sterility when paired with an *An. merus* X. QTL mapping was performed on *mer*^{BC} individuals with a fully *An. merus* X. (A) Composite and (B) Bayesian interval mapping methods identified overlapping major QTL for sterility on chromosome 2 (marked A and B) and chromosome 3 (marked C). Dominant *An. coluzzii* alleles in each QTL are likely incompatible with both dominant autosomal *An. merus* alleles and the *An. merus* X (see main text). Thresholds of significance for logs of odd ratio (LOD) and log posterior density (LPD) are denoted by a red line. Bayesian mapping reports multiple epistatic QTL suggesting a highly complex genetic basis of sterility.

Figure 2.8 Ancestry reconstruction on an individual mosquito eliminates noisy SNPs. GATK SNP genotype calls for an individual *col*^{BC} male are plotted as tick marks across each chromosome. Red is homozygous for *An. coluzzii*, blue is heterozygous, and green is homozygous for *An. merus*. 'Noisy' SNPs resulting from low coverage are common, especially in putative heterozygous regions. For example, we know that the autosomes in a *col*^{BC} individual should not harbor any *An. merus* homozygous SNPs, however, they are relatively common. Application of a custom hidden-markov model to call ancestry removes noise and precisely delineates crossover breakpoints.

Figure 2.9 Ancestry reconstruction delineates crossover breakpoints at high resolution. The average crossover breakpoint width detected by the HMM for all backcross males is plotted into 50kb bins. Nearly 50% of breakpoints are less than 100kb and greater than 90% are less 500kb. Breakpoints larger than 500kb come from a relatively small number of individuals with very low coverage.

Figure 2.10 Crossing over varies by genomic position. Recombination rates in (A) *An. coluzzii x An merus* F1 hybrids and (B) *An. merus x An. coluzzii* F1 hybrids were estimated by counting the number of crossover breakpoints that occurred in one megabase intervals across the genome. Recombination is positively correlated with distance from the centromere. We observed little to no recombination between alternative arrangements on 2R. We observed slight quantitative differences in recombination between the two reciprocal hybrids.

Figure 2.11 QTL mapping with EM and HK algorithms provide an estimate of chromosome-wide influence on sterility. The EM and HK algorithms were unable to localize QTL on the different chromosomes. However, they do provide relative estimates of the importance of each chromosome to the fertility phenotype. A) In the *col*^{BC} chromosome 2 plays the most important role in sterility, followed by chromosome 3, and then the X chromosome. (B) Contrastingly, in the *mer*^{BC} the X chromosome plays a major role in sterility. Compared to the X, the autosomes make a more minor, albeit highly significant, contribution to *mer*^{BC} sterility.

Figure 2.1

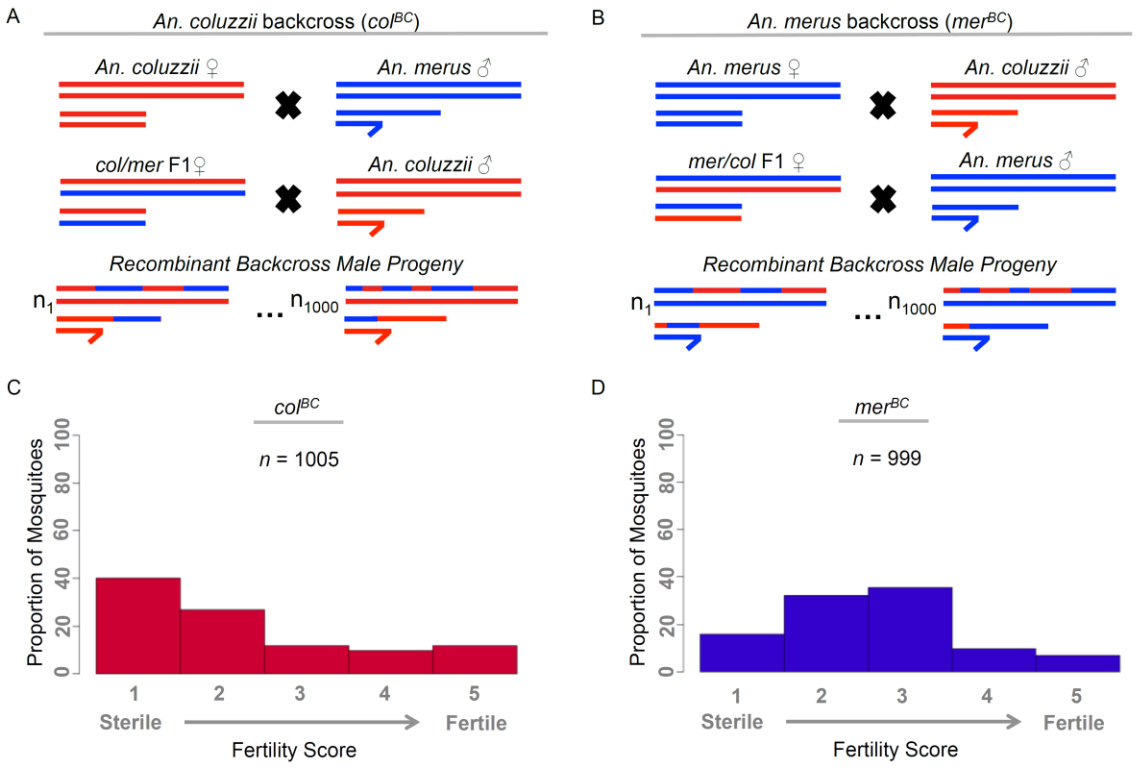

A
An. coluzzii backcross (colBC)
An. coluzzii ♀
An. merus ♂
col/mer F1♀
An. coluzzii ♂
Recombinant Backcross Male Progeny
n1
n1000
B
An. merus backcross (merBC)
An. merus ♀
An. coluzzii ♂
mer/col F1 ♀
An. merus ♂
Recombinant Backcross Male Progeny
n1
n1000
C
colBC
n = 1005
Proportion of Mosquitoes
0
20
40
60
80
100
1
2
3
4
5
Sterile
Fertile
Fertility Score
D
merBC
n = 999
Proportion of Mosquitoes
0
20
40
60
80
100
1
2
3
4
5
Sterile
Fertile
Fertility Score

Figure 2.2

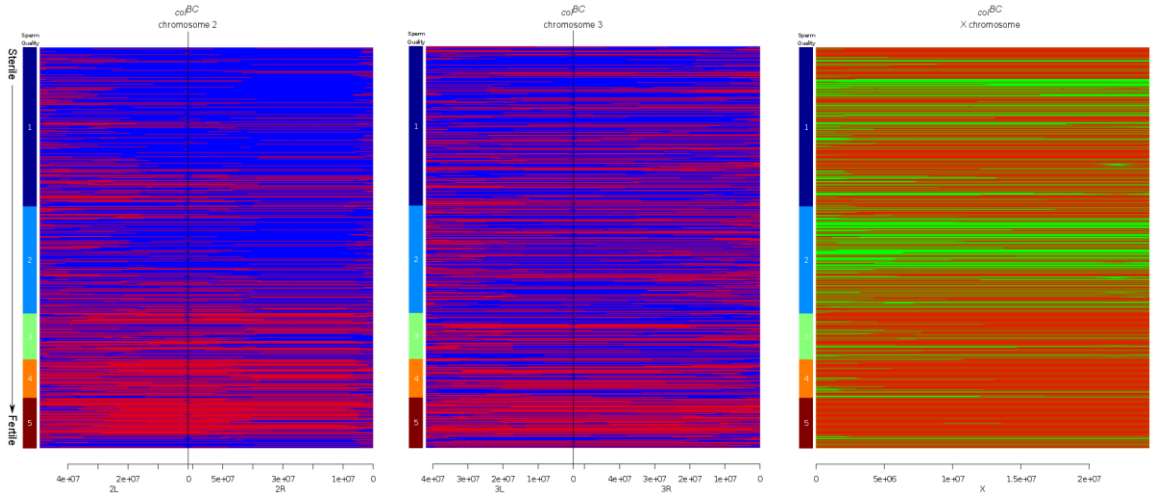

col^BC
chromosome 2
col^BC
chromosome 3
col^BC
X chromosome
Sperm Quality
Sterile
Fertile
1
2
3
4
5
4e+07
2e+07
0
5e+07
3e+07
1e+07
2L
2R
3L
3R
5e+06
1.5e+07
X

Figure 2.3

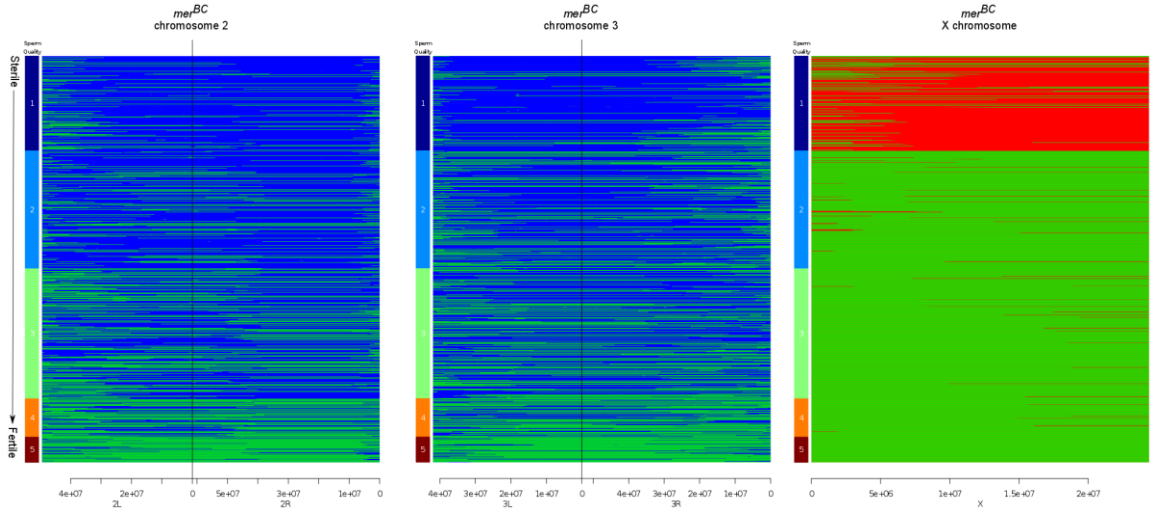

merBC
chromosome 2
merBC
chromosome 3
merBC
X chromosome
Sterile
Fertile
Sperm Quality
1
2
3
4
5
4e+07
2e+07
0
5e+07
3e+07
1e+07
2L
2R
3L
3R
5e+06
1.5e+07
X

Figure 2.4

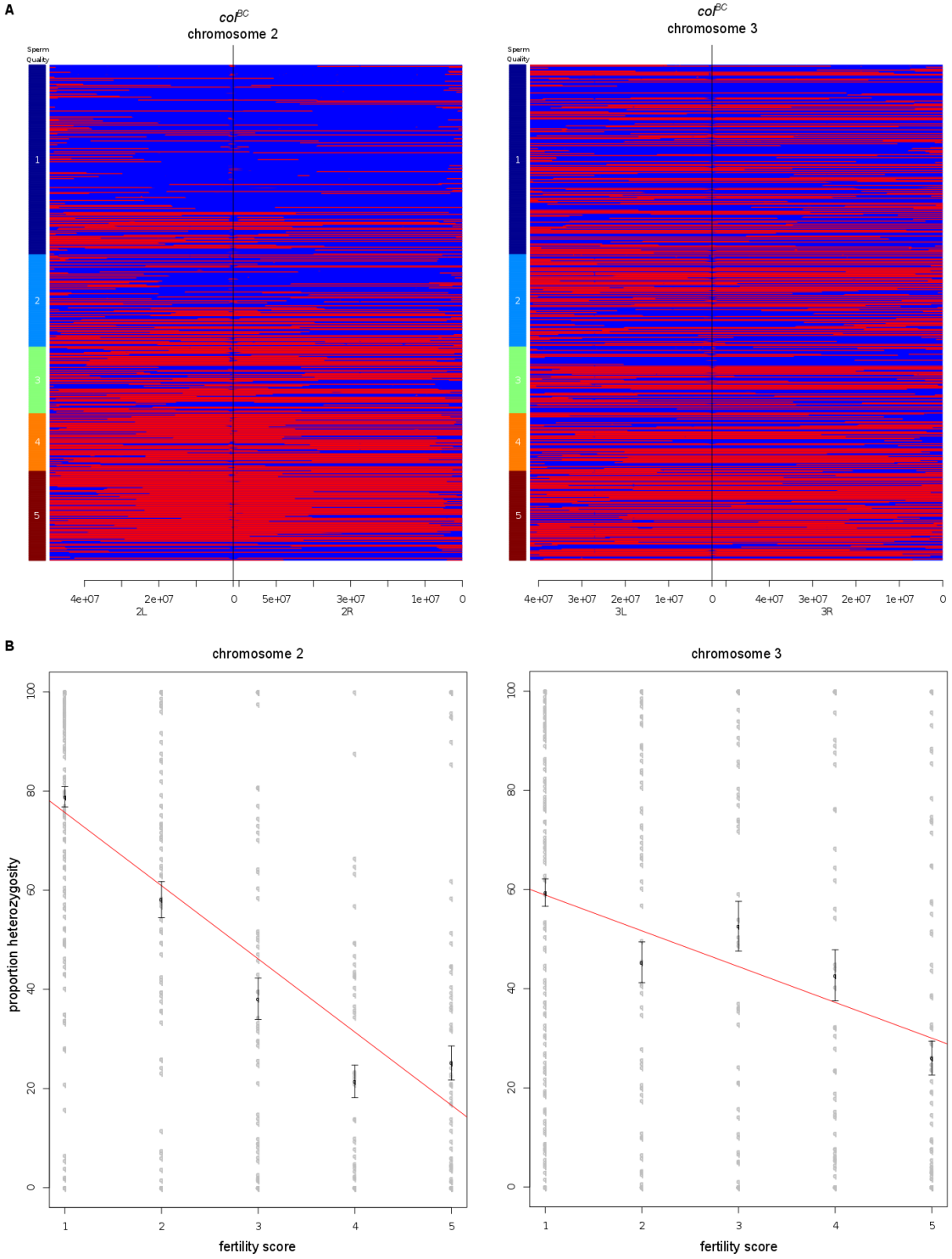

A
col^BC
chromosome 2
Sperm Quality
1
2
3
4
5
4e+07
2e+07
0
5e+07
3e+07
1e+07
0
2L
2R
col^BC
chromosome 3
Sperm Quality
1
2
3
4
5
4e+07
3e+07
2e+07
1e+07
0
4e+07
3e+07
2e+07
1e+07
0
3L
3R
B
chromosome 2
proportion heterozygosity
0
20
40
60
80
100
1
2
3
4
5
fertility score
chromosome 3
0
20
40
60
80
100
1
2
3
4
5
fertility score

Figure 2.5

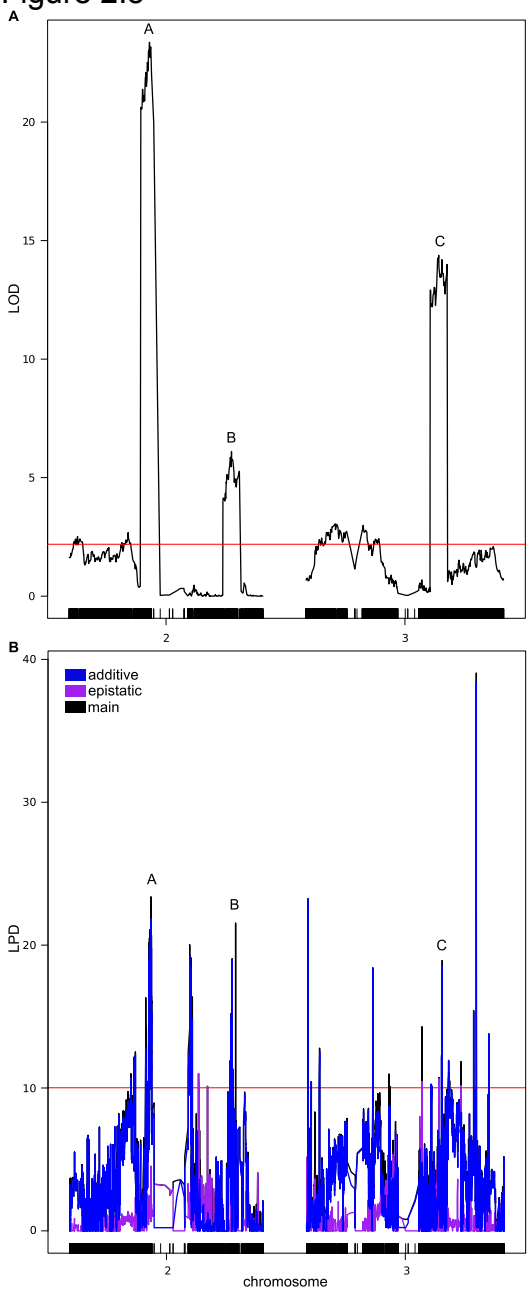

A
A
B
C
LOD
20
15
10
5
0
2
3
B
additive
epistatic
main
40
30
20
10
0
A
B
C
LPD
2
3
chromosome

Figure 2.6

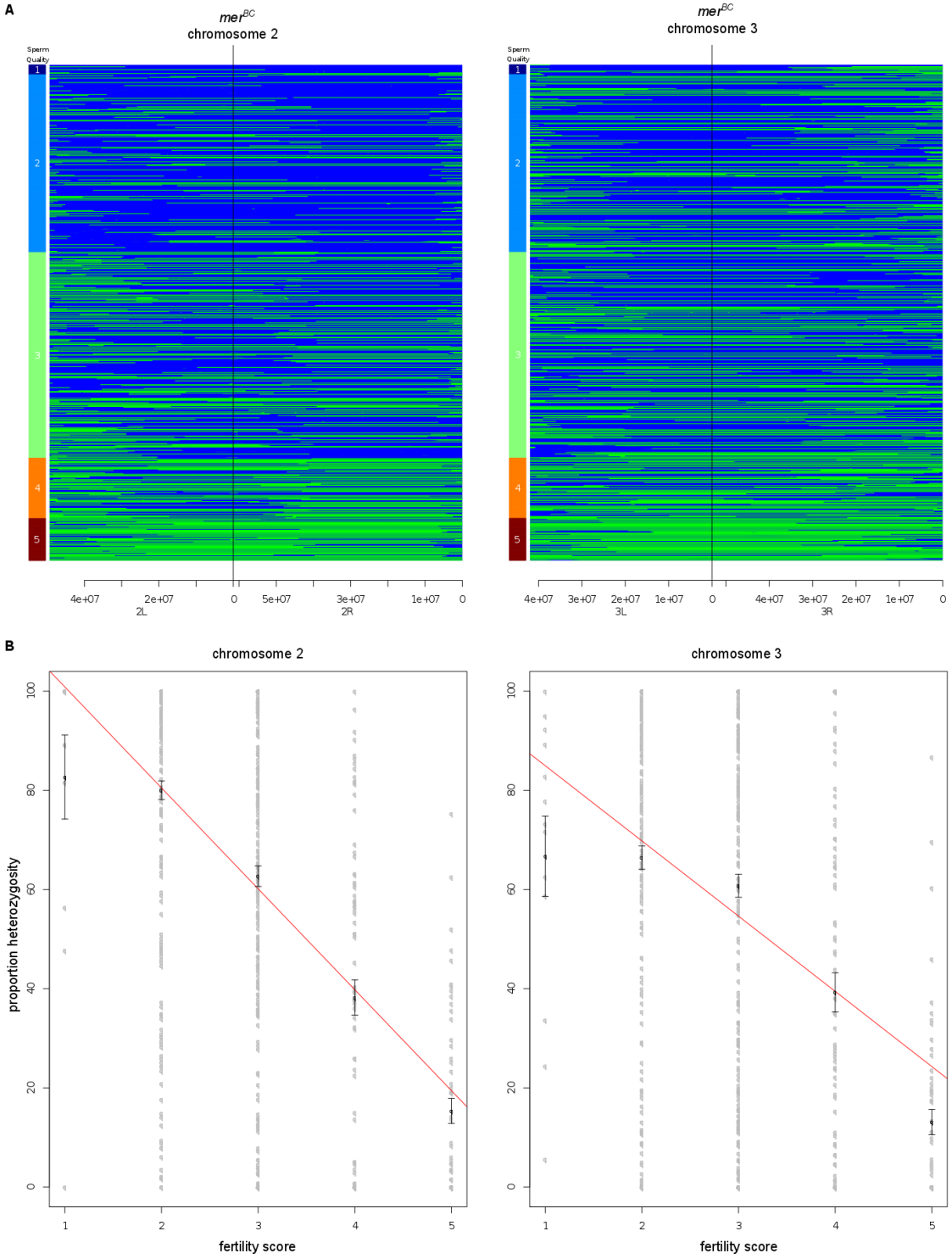

A
merBC
chromosome 2
Sperm Quality
1
2
3
4
5
4e+07
2e+07
0
5e+07
3e+07
1e+07
0
2L
2R
merBC
chromosome 3
Sperm Quality
1
2
3
4
5
4e+07
3e+07
2e+07
1e+07
0
4e+07
3e+07
2e+07
1e+07
0
3L
3R
B
chromosome 2
proportion heterozygosity
0
20
40
60
80
100
1
2
3
4
5
fertility score
chromosome 3
0
20
40
60
80
100
1
2
3
4
5
fertility score

Figure 2.7

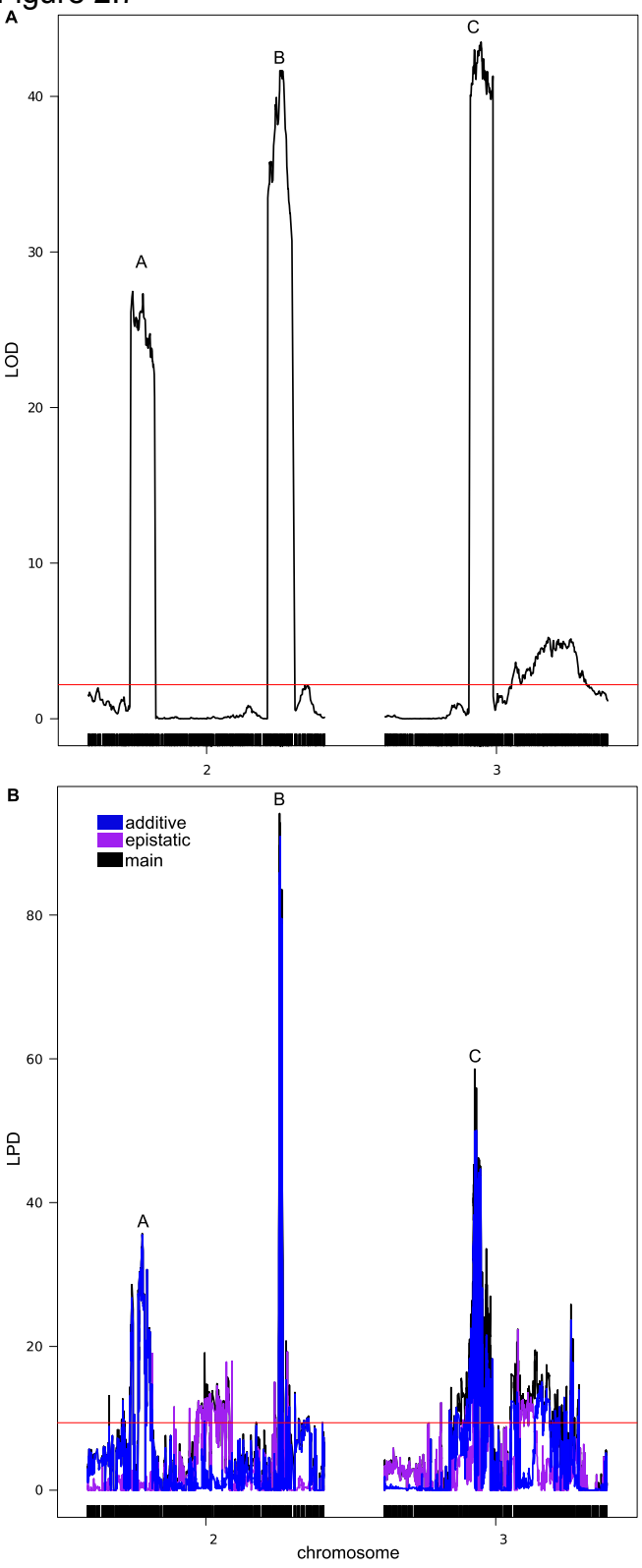

A
B
C
LOD
40
30
20
10
0
2
3
B
additive
epistatic
main
80
60
40
20
0
LPD
chromosome

Figure 2.8

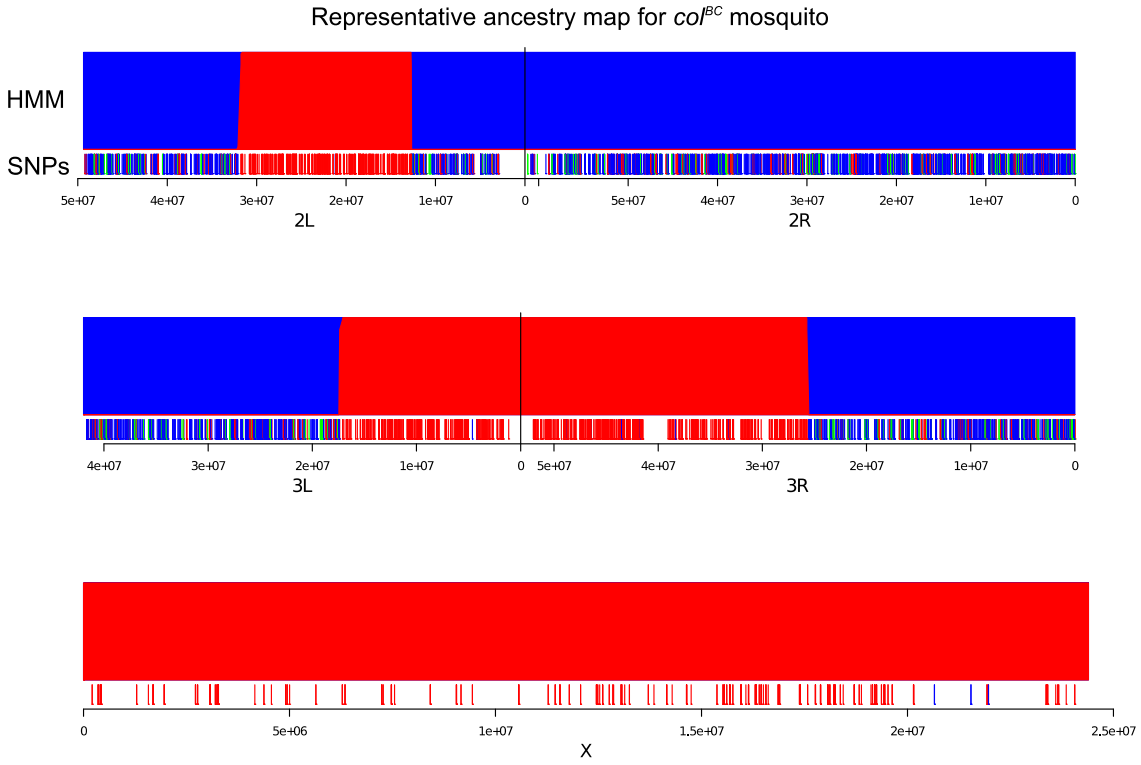

Representative ancestry map for col^BC mosquito
HMM
SNPs
5e+07
4e+07
3e+07
2e+07
1e+07
0
2L
2R
3L
3R
X
5e+06
1.5e+07
2.5e+07

Figure 2.9

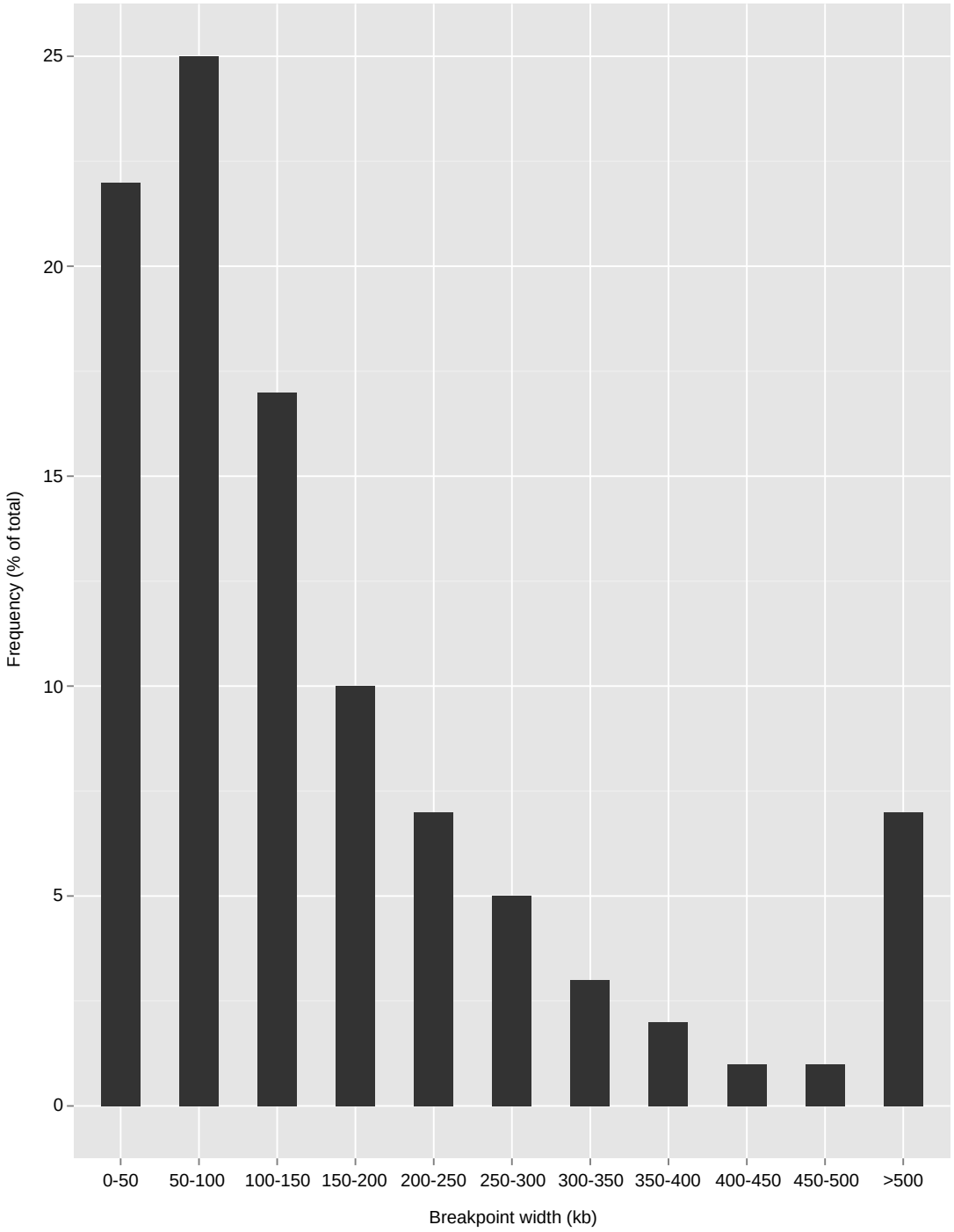

25
20
15
10
5
0
Frequency (% of total)
0-50
50-100
100-150
150-200
200-250
250-300
300-350
350-400
400-450
450-500
>500
Breakpoint width (kb)

Figure 2.10

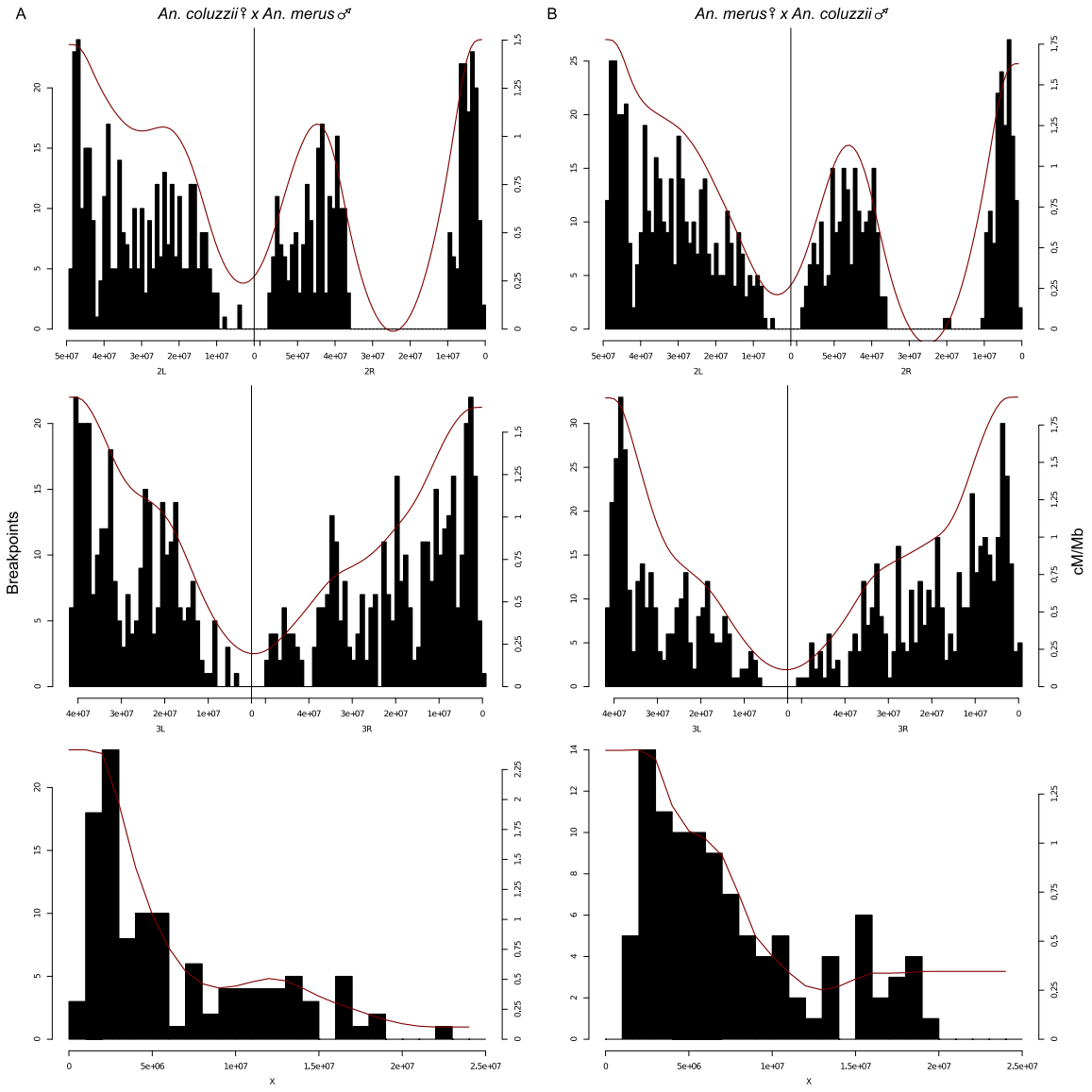
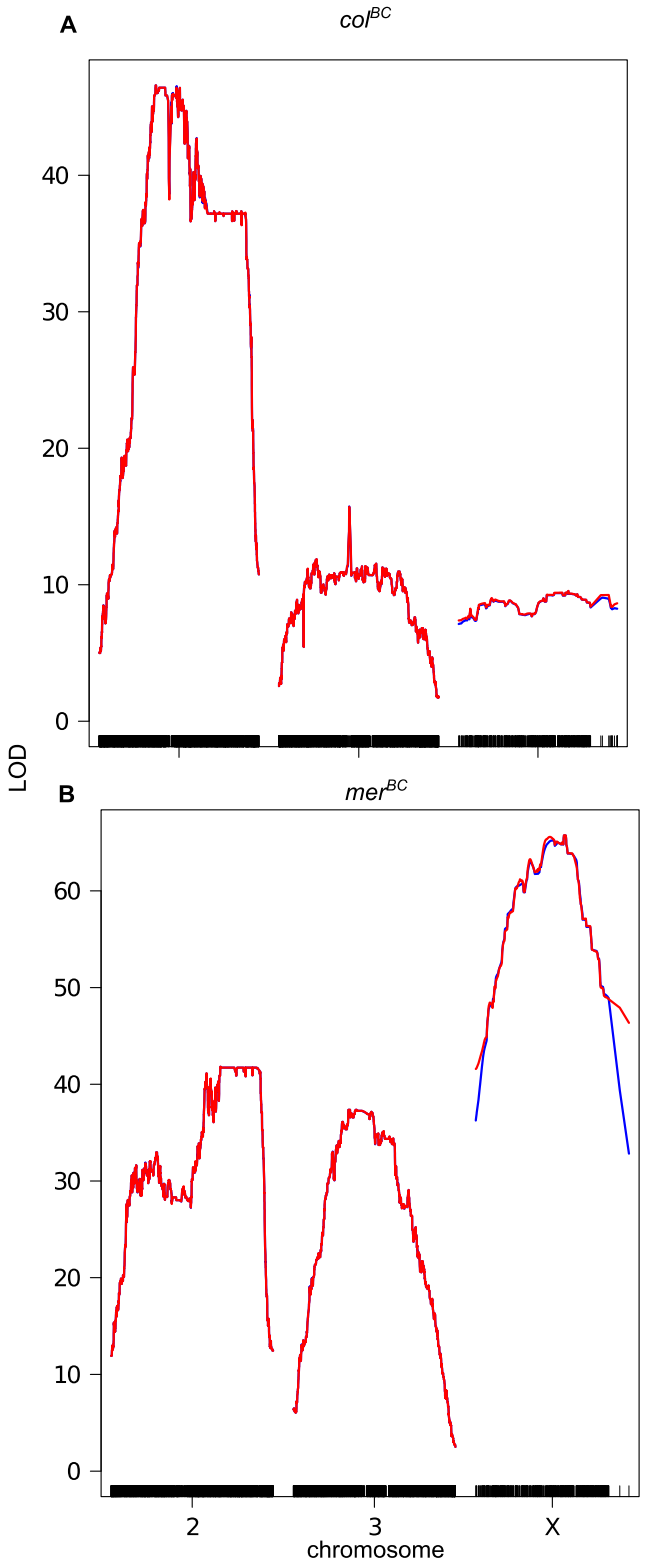

A
An. coluzzii♀ x An. merus♂
B
An. merus♀ x An. coluzzii♂
Breakpoints
cM/Mb
2L
2R
3L
3R
X

Figure 2.11


A
col^BC
40
30
20
10
0
LOD
B
mer^BC
60
50
40
30
20
10
0
2
3
X
chromosome

TABLES

Table 2.1 BIM QTL positions, peaks, and effect for col^{BC}

QTL Name	QTL Position (cM)	Peak Position (bp)	LOD 1.5 Interval (bp)	Size (Kb)	Peak LPD	LOD in model	%Variance in model	F value	P value (chi2)	P value (F)	QTL	Notes
2.1	3.5	2L: 46316137	2L: 46382559-2L: 46307393	75.166	15.18929	1.831	0.8783	8.386	0.004	0.00395	Minor	
2.2	32.5	2L: 20123236	2L: 20289821-2L: 20121852	167.969	19.587561	1.485	0.7112	6.791	0.009	0.00944	Minor	
2.3 **(A)**	45.1	2L: 9769422	2L: 9771109-2L: 9769392	1.717	29.9781	8.486	4.1981	40.08	0	5.50E-010	**Major**	Abuts centromere
2.5 **(B)**	68.2	2R: 24441234	2R: 30539129-2R: 17971261	12567.868	20.90428	6.368	3.1191	29.78	0	7.68E-008	**Major**	*2Rop* inversion
3.1	22.6	3L: 26988030	3L: 29594672-3L: 26966435	2628.237	15.655975	5.123	2.4949	23.82	0	1.43E-006	Minor	
3.3 **(C)**	57.8	3R: 30527198	3R: 30524551-3R: 30528265	3.714	19.19247	8.384	4.1455	39.578	0	6.97E-010	**Major**	

Model LOD:71.83607; **Model % Variance:**48.46766; **Peak:** Highest LPD of non-overlapping QTL; **%Variance in model:** percentage of phenotypic variance in sterility explained by each QTL.

Table 2.2 CIM (Imputation method) QTL positions, peaks, and effect for col^{BC}

Window Length	QTL Position (cM)	Peak Position (bp)	Interval (bp)	Size (Kb)	Peak LOD	QTL	Agree with BIM?
5	32.9	2L: 40210627	2L: 49348982- 2L: 17976184	31372.798	4.59	Minor	**Yes (2.2)**
5	44.9	2L: 9636556	2L: 1:10088235 - 2R: 52147679:61545105	19485.661	23.6	**Major**	**Yes (2.3)**
5	67.9	2R: 36906969	2R: 38795702 - 2R: 7230042	31565.66	4.78	**Major**	**Yes (2.5)**
5	14.8	3L: 33981385	3L: 34376696 - 3L: 33748820	627.876	2.21	Minor	No
5	16.2	3L: 33111225	3L: 33622122- 3L: 33111212	510.91	2.31	Minor	No
5	17.6	3L: 32812009	3L: 32838064 - 3L: 32764055	74.009	2.39	Minor	No
5	22.6	3L: 26239103	3L: 27076303 - 3L: 25848020	1228.283	2.39	Minor	**Yes (3.1)**
5	24.1	3L: 24752441	3L: 24792836 - 3L: 24717605	75.231	2.18	Minor	No
5	50.6	3R: 32821419	3R: 34549326 - 3R: 28934182	5615.144	14.6	**Major**	**Yes (3.3)**
10	32.9	2L: 40210627	2L: 49348982 - 2L:17976184	31372.798	4.59	Minor	**Yes (2.2)**
10	44.7	2L: 10185180	2L: 1:13814525 - 2R: 49501173:61545105	25858.457	23.8	**Major**	**Yes (2.3)**
10	67.9	2R: 36907114	2R: 40365435 - 2R: 6326449	34038.986	4.78	**Major**	**Yes (2.5)**
10	14.7	3L: 34302210	3L: 34376696 - 3L: 33748820	627.876	2.21	Minor	No
10	16.2	3L: 33111225	3L: 33622122 - 3L: 33111212	510.91	2.31	Minor	No
10	17.6	3L: 32812083	3L: 32838064 - 3L: 32764055	74.009	2.39	Minor	No
10	22.6	3L: 26414124	3L: 27076303 - 3L: 25848020	1228.283	2.39	Minor	No
10	24.1	3L: 24792833	3L: 24792836 - 3L: 24717605	75.231	2.18	Minor	**Yes (3.1)**
10	50.6	3R: 32821428	3R: 37347473 - 3R: 24626933	12720.54	14.6	**Major**	**Yes (3.3)**
20	44.7	2L: 10223509	2L: 1:49348982 - 2R: 44262625:61545105	66631.462	23.8	**Major**	**Yes (2.3)**
20	59.3	2R: 43283893	2R: 43293725 - 2R: 43278275	15.45	2.18	Minor	No
20	67.9	2R: 36912016	2R: 40998246 - 2R: 4172937	36825.309	4.78	**Major**	**Yes (2.5)**
20	14.5	3L: 34344154	3L: 34376696 - 3L: 33748820	627.876	2.21	Minor	No
20	16.2	3L: 33111225	3L: 33622122 - 3L: 33111212	510.91	2.31	Minor	No
20	17.6	3L: 32800333	3L: 32838064 - 3L: 32764055	74.009	2.39	Minor	No

20	22.6	3L: 25978092	3L: 27076303 - 3L: 25848020	1228.283	2.39	Minor	**Yes (3.1)**
20	24.1	3L: 24792833	3L: 24792836 - 3L: 24717605	75.231	2.18	Minor	No
20	56.5	3R: 22739209	3L: 3223518 - 3R:19330876:53200684	37093.326	14.8	**Major**	**Yes (3.3)**

Interval: length of the QTL above threshold of 2.18 LOD; **Agree with BIM?:** gives QTL that agree with BIM and the corresponding BIM QTL names.

Table 2.3 CIM (HK method) QTL positions, peaks, and effect for col^{BC}

Window Size	QTL Position (cM)	Peak Position (bp)	Interval (bp)	Size (Kb)	Peak LOD	QTL	Agree with BIM?
5	32.9	2L: 40200624	2L: 49348982 – 2L: 18060426	31288.56	4.59	Minor	**Yes (2.2)**
5	44.9	2L: 9610684	2L: 1:9797189 – 2R: 52147679:61545105	19194.62	23.59	**Major**	**Yes (2.3)**
5	67.9	2R: 36906969	2R: 38795702 – 2R: 7230042	31565.66	4.78	**Major**	**Yes (2.5)**
5	16.2	3L: 33111225	3L: 33111225 – 3L: 33111212	0.013	2.31	Minor	No
5	17.6	3L: 32838052	3L: 32838064 – 3L: 32764055	74.009	2.39	Minor	No
5	22.6	3L: 26106919	3L: 26426525 - 3L: 25848020	578.505	2.39	Minor	**Yes (3.1)**
5	50.6	3R: 32830761	3R: 34549326 – 3R: 28934182	5615.144	14.56	**Major**	**Yes (3.3)**
10	32.9	2L: 40200624	2L: 49348982 – 2L: 18060426	31288.56	4.59	Minor	**Yes (2.2)**
10	44.7	2L: 9830137	2L: 1:13814525 – 2R: 49501173:61545105	25858.46	23.76	**Major**	**Yes (2.3)**
10	67.9	2R: 36906969	2R: 40365435 - 2R: 6326449	34038.99	4.78	**Major**	No
10	16.2	3L: 33111225	3L: 33111225 – 3L: 33111212	0.013	2.31	Minor	No
10	17.6	3L: 32838052	3L: 32838064 – 3L: 32764055	74.009	2.39	Minor	No
10	22.8	3L: 25862031	3L: 26426525 – 3L: 25848020	578.505	2.39	Minor	**Yes (3.1)**
10	50.6	3R: 32830761	3R: 37347473 – 3R: 24626933	12720.54	14.56	**Major**	**Yes (3.3)**
20	44.7	2L: 9830137	2L: 1:49348982 – 2R: 44262625:61545105	66631.46	23.76	**Major**	**Yes (2.3)**
20	67.9	2R: 36906969	2R: 40916401 - 2R: 4172937	36743.46	4.78	**Major**	**Yes (2.5)**
20	16.2	3L: 33111225	3L: 33111225 – 3L: 33111212	0.013	2.31	Minor	No
20	17.6	3L: 32838052	3L: 32838064 – 3L: 32764055	74.009	2.39	Minor	No
20	22.6	3L: 26226812	3L: 26426525 – 3L: 25848020	578.505	2.39	Minor	**Yes (3.1)**
20	56.5	3R: 22717642	3L: 1:3223518 – 3R:19330876:53200684	37093.33	14.8	**Major**	**Yes (3.3)**

Interval: length of the QTL above threshold of 2.3 LOD; **Agree with BIM?:** gives QTL that agree with BIM and the corresponding BIM QTL names.

Table 2.4 BIM QTL and corresponding CIM QTL gene counts for col^{BC}

QTL name	# of genes in BIM interval	# of genes in CIM IMP interval	# of genes in CIM HK interval
2.1	25	N/A	N/A
2.2	43	2141	2134
2.3 **(A)**	1	911	900
2.5 **(B)**	849	2052	2052
3.1	77	35	12
3.3 **(C)**	0 (2 partial)	273	273

Intervals as described in previous tables. Only QTL intervals with the window size of 5 were used.

Table 2.5 BIM QTL positions, peaks, and effect for mer^{BC}

QTL Name	QTL Position (cM)	Peak Position (bp)	LOD 1.5 Interval (bp)	Size (Kb)	Peak LPD	LOD in model	%Variance in model	F value	Pvalue (chi2)	Pvalue (F)	QTL	Notes
2.1 **(A)**	22	2L: 24147023-24151126	2L: 24147018 - 2L: 25992722	1845.704	38.51634	26.074	7.9933	129.73	0	< 2e-16	**Major**	
2.3	48	2L: 6885036	2L: 6880309– 2L: 6891318	11.009	28.990248	9.819	2.8403	23.05	0	2.14E-10	Minor	
2.5 **(B)**	71.6	2R: 20859199 -20859190	2R: 20859115 – 2R: 20865738	6.623	113.95331	38.051	12.1835	98.87	0	< 2e-16	**Major**	*2Rop* inversion
2.6	82.1	2R: 5812824	2R: 5807610 – 2R: 5826915	19.305	19.37842	2.86	0.8072	13.1	0	0.000318	Minor	
3.2 **(C)**	35.5	3L: 2064631	3L: 1:2064678-3R: 49130117: 53200684	6135.244	57.1559	32.042	10.0375	81.45	0	< 2e-16	**Major**	Crosses Centromere
3.4	54	3R: 27452768 -27452698	3R: 27452696 – 3R: 27462876	10.18	20.027494	10.272	2.9762	24.15	0	7.66E-11	Minor	
3.5	60.1	3R: 10242452	3R: 10241654 – 3R: 10244979	3.325	38.93813	2.708	0.764	12.4	0	0.00046	Minor	

Model LOD: 131.18; **Model % Variance:** 60.07227; **Peak:** Highest LPD of non-overlapping QTL; **%Variance in model:** percentage of phenotypic variance in sterility explained by each QTL

Table 2.6 BIM Epistatic QTL positions, peaks, and effect for mer^{BC}

Epistatic QTL	LOD in model	%Variance in model
2.3 : 3.2	9.652	2.7904
2.5 : 3.4	10.026	2.9024

Table 2.7 CIM (Imputation method) QTL positions, peaks, and effect for mer^{BC}

Window Size	QTL Position (cM)	Peak Position (bp)	Interval (bp)	Size (Kb)	Peak LOD	QTL	Agree with BIM?
5	17.8	2L: 23930605	22261621 – 25747005	3485.384	27.22	**Major**	**Yes (2.1)**
5	71.4	2R: 29614768	39538119 – 8340401	31197.72	44.71	**Major**	**Yes (2.5)**
5	80.7	2R: 4371915	4724422 – 4325394	399.028	2.47	minor	No
5	80.9	2R: 4243012	4279081 – 4178753	100.328	2.29	minor	No
5	82.5	2R: 3565205	4154295 – 3565163	589.132	2.59	minor	No
5	35.7	3R: 48900756	3L:1:14878355 – 3R:44002275:53200684	24076.76	42.24	**Major**	**Yes(3.2)**
5	48.3	3R: 27404258	27827895 – 24985271	2842.624	3.22	minor	**Yes (3.4)**
5	51.1	3R: 23398867	24080513 – 22717642	1362.871	2.71	minor	No
5	53.2	3R: 21885737	22526720 – 20991534	1535.186	2.81	minor	No
5	69.7	3R: 7954950	20739358 – 4782497	15956.86	4.84	minor	**Yes(3.5)**
5	74.8	3R: 4619780	4627358 – 4619780	7.578	2.4	minor	No
10	17.8	2L: 23930605	2L: 43855404 – 2L: 27360161	16495.24	27.22	**Major**	**Yes (2.1)**
10	71.6	2R: 18387200	41321049 – 6326449	34994.6	44.71	**Major**	**Yes (2.5)**
10	80.7	2R: 4480019	4724422 – 4325394	399.028	2.47	minor	No
10	80.9	2R: 4197141	4279081 – 4178753	100.328	2.29	minor	No
10	82.5	2R: 3574083	4154295 – 3565163	589.132	2.59	minor	No
10	35.7	3R: 49882968	3L:1:18478320 – 3R:36393069:53200684	35285.93	42.24	**Major**	**Yes(3.2)**
10	48.3	3R: 27440781	27827895 -24985271	2842.624	3.22	minor	**Yes (3.4)**
10	51.2	3R: 23325083	24080513 – 22717642	1362.871	2.71	minor	No
10	53.2	3R: 21821767	22526720 – 20991534	1535.186	2.81	minor	No
10	69.7	3R: 7999653	20739358 – 4782497	15956.86	4.84	minor	**Yes(3.5)**
10	74.8	3R: 4619780	4627358 – 4619780	7.578	2.4	minor	No
20	17.8	2L: 23930605	2L: 45943399 – 2L: 31089522	14853.88	27.22	**Major**	**Yes (2.1)**
20	71.4	2R: 30137923	45777396 – 4178753	41598.64	44.71	**Major**	**Yes (2.5)**
20	82.5	2R: 3565205	4154295 – 3565163	589.132	2.59	minor	No
20	35.7	3R: 48675669	3L:1:24383380 – 3R:31723160:53200684	45860.9	42.24	**Major**	**Yes(3.2)**
20	48.3	3R: 27357996	27827895 – 24985271	2842.624	3.22	minor	**Yes (3.4)**
20	51.1	3R: 23398867	24080513 – 22717642	1362.871	2.71	minor	No
20	53.2	3R: 21885738	22526720 – 20991534	1535.186	2.81	minor	No
20	69.7	3R: 7965252	20739358 – 4782497	15956.86	4.84	minor	**Yes(3.5)**
20	74.8	3R: 4627358	4627358 – 4619780	7.578	2.4	minor	No

Interval: length of the QTL above threshold of 2.26 LOD; **Agree with BIM?:** gives QTL that agree with BIM and the corresponding BIM QTL names.

Table 2.8 CIM (HK method) QTL positions, peaks, and effect for mer^{BC}

Window Size	QTL Position (cM)	Peak Position (bp)	Interval (bp)	Size (Kb)	Peak LOD	QTL	Agree with BIM?
5	17.8	2L: 23930605	22261621 – 25747005	3485.384	27.22	**Major**	**Yes(2.1)**
5	71.4	2R: 20342023	39538119 – 8340401	31197.72	44.71	**Major**	**Yes(2.5)**
5	80.7	2R: 4495703	4509751 – 4325394	184.357	2.47	Minor	No
5	82.5	2R: 3574072	4154295 – 3565163	589.132	2.59	Minor	No
5	35.7	3R: 48570579	3L:1:14878355 – 3R:44002275:53200684	24077.76	42.24	**Major**	**Yes(3.2)**
5	48.3	3R: 27414848	27827895 – 25413587	2414.308	3.22	Minor	**Yes(3.4)**
5	51.1	3R: 23349896	24080513 – 22717642	1362.871	2.71	Minor	No
5	53.2	3R: 21822088	22390806 – 20991534	1399.272	2.81	Minor	No
5	69.7	3R: 7881757	20739358 – 4819092	15920.27	4.84	Minor	**Yes(3.5)**
5	74.8	3R: 4627358	4627358 – 4619780	7.578	2.4	Minor	No
10	17.8	2L: 23930605	43855404 – 27360161	16495.24	27.22	**Major**	**Yes(2.1)**
10	71.4	2R: 20342023	41321049 – 6326449	34994.6	44.71	**Major**	**Yes(2.5)**
10	80.7	2R: 4495703	4509751 – 4325394	184.357	2.47	Minor	No
10	82.5	2R: 3574072	4154295 – 3565163	589.132	2.59	Minor	No
10	35.7	3R: 48570579	3L:1:18478320 – 3R:36393069:53200684	35285.93	42.24	**Major**	**Yes(3.2)**
10	48.3	3R: 27414848	27827895 - 25413587	2414.308	3.22	Minor	**Yes(3.4)**
10	51.1	3R: 23349896	24080513 – 22717642	1362.871	2.71	Minor	No
10	53.2	3R: 21822088	22390806 – 20991534	1399.272	2.81	Minor	No
10	69.7	3R: 7881757	20739358 – 4819092	15920.27	4.84	Minor	**Yes(3.5)**
10	74.8	3R: 4627358	4627358 – 4619780	7.578	2.4	Minor	No
20	17.8	2L: 23930605	45943399 – 31089522	14853.88	27.22	**Major**	**Yes(2.1)**
20	71.6	2R: 19355367	45777396 – 4197141	41580.25	44.71	**Major**	**Yes(2.5)**
20	82.5	2R: 3574072	4154295 – 3565163	589.132	2.59	Minor	No
20	35.7	3R: 49908327	3L:1:24383380 – 3R:31723160:53200684	45860.9	42.24	**Major**	**Yes(3.2)**
20	48.3	3R: 27414850	27827895 – 25413587	2414.308	3.22	Minor	**Yes(3.4)**
20	51.1	3R: 23349896	24080513 – 22717642	1362.871	2.71	Minor	No
20	53.2	3R: 21822088	22390806 – 20991534	1399.272	2.81	Minor	No
20	69.7	3R: 7881757	20739358 – 4819092	15920.27	4.84	Minor	**Yes(3.5)**
20	74.8	3R: 4627358	4627358 – 4619780	7.578	2.4	Minor	No

Interval: length of the QTL above threshold of 2.36 LOD; **Agree with BIM?:** gives QTL that agree with BIM and the corresponding BIM QTL names.

Table 2.9 BIM QTL and corresponding CIM QTL gene counts for mer^{BC}

QTL name	# of genes in BIM interval	# of genes in CIM IMP interval	# of genes in CIM HK interval
2.1 **(A)**	103	230	257
2.3	0 (1 predicted)	N/A	N/A
2.5 **(B)**	0 (2 partial)	2065	2065
2.6	0 (3mRNA hits)	N/A	N/A
3.2 **(C)**	273	1259	1259
3.4	0 (1 predicted)	104	95
3.5	0 (2 partial)	984	975

Intervals as described in previous tables. Only QTL intervals with the window size of 5 were used.

Table 2.10 GO Overrepresentation Test of for col^{BC} BIM QTL genes
A.)

***An. coluzzii* orthologs**					
GO name	**# in REF**	**# in QTL**	**# expected**	**Fold Enrichment**	**P-value**
Molecular Function					
unclassified	5062	324	321.83	1.01	0.00E+000
Biological Process					
intracellular cholesterol transport	11	9	0.7	12.87	7.21E-005
cholesterol transport	20	9	1.27	7.08	9.45E-003
sterol transport	21	9	1.34	6.74	1.39E-002
intracellular sterol transport	12	9	0.76	11.8	1.49E-004
intracellular lipid transport	15	9	0.95	9.44	9.39E-004
flavonoid glucuronidation	22	9	1.4	6.43	1.99E-002
cellular glucuronidation	22	9	1.4	6.43	1.99E-002
glucuronate metabolic process	22	9	1.4	6.43	1.99E-002
uronic acid metabolic process	22	9	1.4	6.43	1.99E-002
flavonoid metabolic process	22	9	1.4	6.43	1.99E-002
flavonoid biosynthetic process	22	9	1.4	6.43	1.99E-002
unclassified	5531	335	351.64	-0.95	0.00E+000
Cellular Component					
unclassified	5667	351	360.29	-0.97	0.00E+000

B.)

An. merus orthologs					
GO name	**# in REF**	**# in QTL**	**# expected**	**Fold Enrichment**	**P-value**
Molecular Function					
unclassified	5062	312	322.24	-0.97	0.00E+000
Biological Process					
intracellular cholesterol transport	11	9	0.7	12.85	7.29E-005
cholesterol transport	20	9	1.27	7.07	9.55E-003
sterol transport	21	9	1.34	6.73	1.40E-002
intracellular sterol transport	12	9	0.76	11.78	1.51E-004
intracellular lipid transport	15	9	0.95	9.43	9.49E-004
insecticide catabolic process	22	9	1.4	6.43	2.01E-002
insecticide metabolic process	22	9	1.4	6.43	2.01E-002
toxin metabolic process	22	9	1.4	6.43	2.01E-002
xenobiotic metabolic process	23	9	1.46	6.15	2.84E-002
cellular response to xenobiotic stimulus	23	9	1.46	6.15	2.84E-002
response to xenobiotic stimulus	24	9	1.53	5.89	3.94E-002
response to insecticide	22	9	1.4	6.43	2.01E-002
toxin catabolic process	22	9	1.4	6.43	2.01E-002
secondary metabolic catabolic process	22	9	1.4	6.43	2.01E-002
xenobiotic catabolic process	23	9	1.46	6.15	2.84E-002
response to DDT	22	9	1.4	6.43	2.01E-002
unclassified	5531	315	352.1	-0.89	0.00E+000
Cellular Component					
unclassified	5667	336	360.75	-0.93	0.00E+000

Table 2.11 GO Overrepresentation Test of for col^{BC} CIM IMP QTL genes

A.)

***An. coluzzii* orthologs**					
GO name	**# in REF**	**# in QTL**	**# expected**	**Fold Enrichment**	**P-value**
Molecular Function					
unclassified	5062	1679	1732.46	-0.97	0.00E+000
Biological Process					
unclassified	5531	1809	1892.97	-0.96	0.00E+000
Cellular Component					
unclassified	5667	1879	1939.52	-0.97	0.00E+000

B.)

***An. merus* orthologs**					
GO name	**# in REF**	**# in QTL**	**# expected**	**Fold Enrichment**	**P-value**
Molecular Function					
unclassified	5062	1615	1691.35	-0.95	0.00E+000
Biological Process					
unclassified	5531	1740	1848.05	-0.94	0.00E+000
Cellular Component					
unclassified	5667	1817	1893.49	-0.96	0.00E+000

Table 2.12 GO Overrepresentation Test of for col^{BC} CIM HK QTL genes

A.)

***An. coluzzii* orthologs**					
GO name	**# in REF**	**# in QTL**	**# expected**	**Fold Enrichment**	**P-value**
Molecular Function					
unclassified	5062	1641	1691.35	-0.97	0.00E+000
Biological Process					
unclassified	5531	1768	1848.05	-0.96	0.00E+000
Cellular Component					
unclassified	5667	1827	1893.49	-0.96	0.00E+000

B.)

***An. merus* orthologs**					
GO name	**# in REF**	**# in QTL**	**# expected**	**Fold Enrichment**	**P-value**
Molecular Function					
unclassified	5062	1577	1644.25	-0.95	0.00E+000
Biological Process					
unclassified	5531	1702	1800.86	-0.95	0.00E+000
Cellular Component					
unclassified	5667	1762	1845.15	-0.95	0.00E+000

Table 2.13 GO Overrepresentation Test of for mer^{BC} BIM QTL genes

A.)

***An. coluzzii* orthologs**					
GO name	**# in REF**	**# in QTL**	**# expected**	**Fold Enrichment**	**P-value**
Molecular Function					
structural constituent of cuticle	156	30	2.28	13.17	2.13E-021
structural molecule activity	375	35	5.48	6.39	2.21E-015
unclassified	5062	52	73.92	-0.7	0.00E+000
Biological Process					
unclassified	5531	95	80.76	1.18	0.00E+000
Cellular Component					
extracellular space	297	20	4.34	4.61	8.92E-006
extracellular region part	319	20	4.66	4.29	2.83E-005
extracellular region	525	22	7.67	2.87	4.70E-003
unclassified	5667	78	82.75	-0.94	0.00E+000

B.)

***An. merus* orthologs**					
GO name	**# in REF**	**# in QTL**	**# expected**	**Fold Enrichment**	**P-value**
Molecular Function					
structural constituent of cuticle	156	19	2.03	9.34	3.16E-010
structural molecule activity	375	25	4.89	5.11	2.46E-008
unclassified	5062	52	66.03	-0.79	0.00E+000
Biological Process					
unclassified	5531	84	72.14	1.16	0.00E+000
Cellular Component					
extracellular space	297	16	3.87	4.13	9.94E-004
extracellular region part	319	16	4.16	3.85	2.44E-003
unclassified	5667	69	73.92	-0.93	0.00E+000

Table 2.14 GO Overrepresentation Test of for mer^{BC} CIM IMP QTL genes
A.)

***An. coluzzii* orthologs**					
GO name	**# in REF**	**# in QTL**	**# expected**	**Fold Enrichment**	**P-value**
Molecular Function					
monooxygenase activity	127	60	30.27	1.98	9.31E-004
heme binding	151	68	35.98	1.89	9.57E-004
tetrapyrrole binding	152	68	36.22	1.88	1.19E-003
iron ion binding	161	70	38.37	1.82	2.20E-003
oxidoreductase activity	162	68	38.61	1.76	9.22E-003
unclassified	5062	1118	1206.33	-0.93	0.00E+000
Biological Process					
cellular protein metabolic process	1136	339	270.72	1.25	1.80E-002
cellular macromolecule metabolic process	1921	558	457.79	1.22	4.82E-004
cellular metabolic process	2725	760	649.39	1.17	9.32E-004
cellular process	4277	1162	1019.25	1.14	2.73E-004
Single-organism process	3201	864	762.83	1.13	1.76E-002
unclassified	5531	1173	1318.09	-0.89	0.00E+000
Cellular Component					
extracellular space	297	107	70.78	1.51	1.37E-002
extracellular region part	319	112	76.02	1.47	2.51E-002
cytoplasm	2147	592	511.65	1.16	3.18E-002
unclassified	5667	1251	1350.5	-0.93	0.00E+000

B.)

An. merus orthologs					
GO name	**# in REF**	**# in QTL**	**# expected**	**Fold Enrichment**	**P-value**
Molecular Function					
monooxygenase activity	127	59	30.31	1.95	1.97E-003
heme binding	151	67	36.03	1.86	1.94E-003
tetrapyrrole binding	152	67	36.27	1.85	2.40E-003
iron ion binding	161	71	38.42	1.85	1.21E-003
oxidoreductase activity	162	69	38.66	1.78	5.25E-003
unclassified	5062	1093	1207.99	-0.9	0.00E+000
Biological Process					
translation	259	97	61.81	1.57	2.16E-02
cellular macromolecule biosynthetic process	704	230	168	1.37	2.06E-003
macromolecule biosynthetic process	709	230	169.19	1.36	3.33E-003
macromolecule metabolic process	2557	703	610.2	1.15	2.26E-002
metabolic process	3885	1062	927.11	1.15	7.92E-005
organic substance biosynthetic process	1078	328	257.25	1.28	6.10E-003
biosynthetic process	1136	346	271.09	1.28	2.97E-003
cellular biosynthetic process	1047	327	249.85	1.31	7.04E-004
cellular metabolic process	2725	791	650.29	1.22	6.78E-007
cellular process	4277	1183	1020.66	1.16	3.13E-007
cellular macromolecule metabolic process	1921	582	458.42	1.27	8.41E-007
gene expression	889	274	212.15	1.29	1.62E-002
cellular protein metabolic process	1136	355	271.09	1.31	2.04E-004
primary metabolic process	3100	846	739.78	1.14	5.76E-003
peptide biosynthetic process	262	97	62.52	1.55	3.32E-002
amide biosynthetic process	283	103	67.53	1.53	3.67E-002
Single-organism metabolic process	1308	384	312.14	1.23	2.10E-002
Single-organism process	3201	875	763.88	1.15	2.73E-003
unclassified	5531	1141	1319.91	-0.86	0.00E+000
Cellular Component					
intracellular organelle part	1433	430	341.97	1.26	2.55E-004
intracellular part	3615	982	862.68	1.14	4.45E-004
intracellular	3885	1053	927.11	1.14	1.89E-004
cell part	4535	1199	1082.22	1.11	2.22E-003
cell	4559	1206	1087.95	1.11	1.81E-003
intracellular organelle	2840	773	677.73	1.14	9.95E-003
organelle	2861	778	682.74	1.14	1.04E-002
organelle part	1446	433	345.07	1.25	2.87E-004
cytoplasmic part	1438	423	343.16	1.23	2.26E-003
cytoplasm	2147	623	512.36	1.22	4.07E-005
macromolecular complex	1489	428	355.33	1.2	1.62E-002
intracellular membrane-bound organelle	2447	671	583.95	1.15	1.93E-002

membrane-bound organelle	2507	685	598.27	1.14	2.34E-002
unclassified	5667	1217	1352.36	-0.9	0.00E+000

Table 2.15 GO Overrepresentation Test of for mer^{BC} CIM HK QTL genes A.)

***An. coluzzii* orthologs**					
GO name	**# in REF**	**# in QTL**	**# expected**	**Fold Enrichment**	**P-value**
Molecular Function					
monooxygenase activity	127	60	30.04	2	7.34E-004
heme binding	151	68	35.71	1.9	7.41E-004
tetrapyrrole binding	152	68	35.95	1.89	9.24E-004
iron ion binding	161	70	38.08	1.84	1.70E-003
oxidoreductase activity	162	68	38.31	1.77	7.27E-003
unclassified	5062	1110	1197.19	-0.93	0.00E+000
Biological Process					
cellular protein metabolic process	1136	337	268.67	1.25	1.66E-002
cellular macromolecule metabolic process	1921	556	454.33	1.22	3.04E-004
cellular metabolic process	2725	756	644.48	1.17	6.98E-004
cellular process	4277	1154	1011.53	1.14	2.58E-005
Single-organism process	3201	854	757.05	1.13	3.51E-002
unclassified	5531	1166	1308.11	-0.89	0.00E+000
Cellular Component					
extracellular space	297	105	70.24	1.49	2.47E-002
extracellular region part	319	110	75.45	1.46	4.32E-002
unclassified	5667	1248	1340.28	-0.93	0.00E+000

B.)

An. merus orthologs					
GO name	**# in REF**	**# in QTL**	**# expected**	**Fold Enrichment**	**P-value**
Molecular Function					
monooxygenase activity	127	59	30.02	1.97	1.47E-003
heme binding	151	67	35.69	1.88	1.41E-003
tetrapyrrole binding	152	67	35.92	1.87	1.75E-003
iron ion binding	161	71	38.05	1.87	8.69E-004
oxidoreductase activity	162	69	38.29	1.8	3.84E-003
unclassified	5062	1084	1196.36	-0.91	0.00E+000
Biological Process					
translation	259	95	61.21	1.55	3.94E-002
cellular macromolecule biosynthetic process	704	228	166.38	1.37	2.13E-003
macromolecule biosynthetic process	709	228	167.57	1.36	3.43E-003
macromolecule metabolic process	2557	697	604.32	1.15	2.13E-002
metabolic process	3885	1051	918.19	1.14	1.07E-004
organic substance biosynthetic process	1078	326	254.78	1.28	4.85E-003
biosynthetic process	1136	344	268.48	1.28	2.26E-003
cellular biosynthetic process	1047	325	247.45	1.31	5.56E-004
cellular metabolic process	2725	784	644.03	1.22	6.93E-007
cellular process	4277	1171	1010.83	1.16	4.40E-007
cellular macromolecule metabolic process	1921	577	454.01	1.27	8.44E-007
gene expression	889	270	210.11	1.29	2.63E-002
cellular protein metabolic process	1136	352	268.48	1.31	2.04E-004
primary metabolic process	3100	838	732.66	1.14	6.19E-003
Single-organism metabolic process	1308	378	309.13	1.22	4.02E-002
Single-organism process	3201	864	756.53	1.14	5.02E-003
unclassified	5531	1134	1307.2	-0.87	0.00E+000
Cellular Component					
intracellular organelle part	1433	425	338.68	1.25	3.64E-004
intracellular part	3615	969	854.37	1.13	9.94E-004
intracellular	3885	1039	918.19	1.13	4.58E-004
cell part	4535	1183	1071.81	1.1	5.30E-003
cell	4559	1189	1077.48	1.1	5.10E-003
intracellular organelle	2840	762	671.21	1.14	2.06E-002
organelle	2861	767	676.17	1.13	2.13E-002
organelle part	1446	428	341.75	1.25	4.04E-004
cytoplasmic part	1438	414	339.86	1.22	8.32E-003
cytoplasm	2147	612	507.42	1.21	1.59E-004
macromolecular complex	1489	423	351.91	1.2	2.14E-002
intracellular membrane-bound organelle	2447	661	578.33	1.14	4.00E-002

membrane-bound organelle	2507	675	592.51	1.14	4.68E-002
unclassified	5667	1213	1339.35	-0.91	0.00E+000

Chapter III

Impaired silencing of the X chromosome during spermatogenesis in sterile hybrid malaria mosquitoes

Abstract

Over 40 years ago, Lifschytz and Lindsley hypothesized that disruption of meiotic sex chromosome inactivation (MSCI) in the male germline is a common cause of hybrid sterility. While a compelling argument, empirical evidence supporting the hypothesis remains sparse. Comparative transcriptomic profiling during spermatogenesis reveals that X-specific overexpression in the testes of mosquitoes is associated with not only male hybrid sterility, but also the large X-effect. Dominance to additive ratios corroborate these results. Our results have extended the hypothesis of Lifschytz and Lindsley, emphasizing the importance of improper gene regulation in hybrid dysfunction.

Introduction

A long hypothesized, but oft neglected, mechanism that may partially explain both the rapid evolution of male sterility and the large-X effect is disruption of meiotic sex chromosome inactivation (MSCI) in the germline of the male heterogametes (Lifschytz and Lindsley, 1972). In diverse eukaryotes including mammals (Turner, 2007), grasshoppers (Cabrero et al., 2007) and nematodes (Kelly et al., 2002), the X chromosome is inactivated and transcriptionally silenced during gametogenesis in males. Interestingly, strong

repression of X-linked genes also occurs in the germline of male *Drosophila,* but through a mechanism that appears distinct from MSCI. Evidence suggests that in flies transcriptional repression begins in pre-meiotic cells and persists through meiosis (Meiklejohn et al., 2011). Regardless of the underlying cellular pathway, assuming that transcriptional silencing occurs via a chromosome-wide epigenetic mechanism, a simple verbal model explains how misregulation of the X could lead to hybrid male sterility in a diversity of eukaryotes. In F1 hybrids, factors that regulate expression of the X in one species may not properly repress the 'foreign' X from the alternate species resulting in widespread overexpression of X-linked genes and failure of spermatogenesis in hybrids (Lifschytz and Lindsley, 1972). Recent studies of hybrid sterility between sub-species of the house mouse *Mus musculus* implicate breakdown of MSCI in the initial evolution of hybrid male sterility (Campbell et al., 2013; Good et al., 2010). Also consistent with the misregulation hypothesis, sterile males from certain *Drosophila* crosses show modest overexpression of X-linked, but not autosomal, genes in the testes relative to their parents (Gomes and Civetta, 2014; Moehring et al., 2007). While evidence remains sparse, misregulation and subsequent overexpression of the X during gametogenesis could be an underappreciated cause of both hybrid sterility and the associated large X-effect (Good et al., 2010; Meiklejohn and Tao, 2010). The intriguing hypothesis warrants investigation in additional taxa, which is now possible with the decreasing cost of next generation sequencing and increasing availability of computational analysis tools.

We used targeted transcriptomics to dissect the genetic basis of hybrid male sterility between two members of the *An. gambaie* complex: *An. coluzzii* (previously *An. gambiae* M form) and *An. merus*. Both species are synanthropic and highly efficient vectors of human malaria where present (Coetzee et al., 2013; Cuamba and Mendis, 2009; Govere et al., 2000; Mwangangi et al., 2003; Pock Tsy et al., 2003; Temu et al., 1998). Currently, the two species have allopatric distributions; the range of *An. coluzzii* extends from Northern Senegal in the west to East-Central Africa and south to Angola (Coetzee et al., 2013), while the brackish water breeding *An. merus* is primarily restricted to the coasts of east Africa and Mozambique (Sinka et al., 2010). In addition to potential insights into the formation and maintenance of the species boundaries, elucidation of genes critical to proper *Anopheles* sperm development could enable population suppression through genetic-based sterile insect technique (Nolan et al., 2011; Wang and Jacobs-Lorena, 2013). Discovery of such novel malaria control tools will be critical to curbing disease transmission as physiological and behavioral resistance to insecticidal interventions continues to spread through mosquito populations (Ranson et al., 2011; Reddy et al., 2011).

METHODS

Mosquito Rearing and crosses

All mosquitoes were reared and crossed in accordance with standard protocols (White et al., 2013). In brief, mosquitoes were maintained in insectaries

under controlled conditions of 27°C, 65% relative humidity, and a 12h:12h light:dark cycle with 1 h dawn and dusk transitions. *Anopheles coluzzii* larvae were reared in freshwater (dH2O) at a density of 200 larvae/L of water, while *An. merus* larvae were reared in 25% saltwater (7.5 g NaCl/L) at the same density. F1 hybrid individuals were reared in dH20 with no adverse effects on survival. Each larval tray was fed ~100 mg per day of a 4:1 mixture of finely ground fish pellets to baker's yeast. Reciprocal F1 crosses were performed with 200-400 virgins from each parental colony.

RNA isolation and cDNA library preparation

A total of 23 different species/tissues/replicates were included in the experiment as follows: a) *An. coluzzii* carcass without testes - 4 biological replicates, b) *An. merus* carcass without testes – 4 biological replicates, c) *An. coluzzii* testes – 4 biological replicates, d) *An. merus* testes – 4 biological replicates, e) CxM^{F1} hybrid testes – 3 biological replicates, and f) MXC^{F1} hybrid testes – 4 biological replicates. Total RNA was extracted from pools of 200 mosquito testes or 20 carcasses using Trizol (Life Technologies; Carlsbad, CA). RNA (13 ng per sample) was used to generate two adaptor-ligated double-stranded cDNA libraries for RNA-seq using the NEB Next Ultra RNA library prep kit for Illumina following the manufacturer's protocol. Samples were diluted to 2 nM and pooled to generate two multiplexed cDNA libraries consisting of 11 or 12 adaptor-tagged sample pools each. Samples were multiplexed to avoid having

more than two replicates for any species/tissue in a single library to counteract possible lane bias when sequencing.

Transcriptomic analysis

Paired-end 100bp sequencing was performed on each cDNA library using the Illumina HiSeq2500 platform in the UCR Genomics Core. Raw reads were imported into CLC Genomics Workbench (v6.5.1, CLC Bio) and trimmed for quality, adapter indexes, and poly (A) tails using the default settings (Ambiguous limit = 2, quality limit = 0.05). Redundant reads were removed using the Duplicate Removal plugin in CLC. The raw sequence reads can be retrieved from the NCBI short sequence read archive under the accession number SRP047496.

Two cDNA libraries consisting of 23-pooled samples were sequenced, which generated 617 million raw 100 bp paired end reads. After demultiplexing, trimming poor quality reads, adapters and poly(A) sequences, and removing duplicate reads, ~72% of reads were retained (between 7 and 32 million non-redundant reads per sample). The *An. gambiae* PEST [AgamP4.2; 13,624 genes] and *An. merus* [AmerM1.2; 14,415 genes] reference gene sets were retrieved from VectorBase (Megy et al., 2012). Preprocessed reads were aligned to the reference gene sets using the map to reference function in CLC Genomics Workbench and the following parameters: Similarity Fraction = 0.95; Length Fraction = 0.95; default settings herein. Robustness of the read counting scheme was validated using BWA v1.2.3 (Li and Durbin, 2009) in the MCIC-Galaxy

pipeline (Goecks et al., 2010).

Using the CLC Bio RNA-Seq Analysis tool, read counts for broken and paired reads were normalized by calculating the number of unique (i.e. unambiguous) reads per kilobase of exon model per million mapped reads (RPKM) (Mortazavi et al., 2008). Data from one *An. coluzzii* testes and one *An. coluzzii* whole body replicate were omitted prior to subsequent analysis due to low read mapping counts. Quality control parameters indicated the RPKM file for each *Anopheles* species contained samples that were normally distributed and homogenous. Principal component analyses (PCA) of the read counts derived from the 21 cDNA libraries revealed distinct clustering of samples by *Anopheles* species and tissue (Figure 3.1). Clustering of the parental species and reciprocal hybrid testes samples was less discrete.

Genes differentially expressed between tissues were identified using the R module DESeq (Anders and Huber, 2010), which is based on a negative binomial distribution model. Additional DESeq analyses were carried out to examine differential expression between the parental species and the reciprocal hybrid crosses in each species genetic background. Significance was defined at a FDR < 0.05 (Benjamini and Hochberg, 1995) for all analyses. For each contrast, the dataset was filtered to contain only expressed genes, which were required to have a minimum RPKM value of four for all replicates in at least one species/hybrid. To determine testes-specific genes, we required each sample to have a carcass RPKM value of one or less. A multivariate ANOVA in R was

performed to determine the effect of tissue, species, and their interaction on chromosome-wide gene expression. To compare differences in expression between chromosomes, but within a species/tissue, we performed pairwise t-tests with Holm-Bonferroni correction between the parental species and the reciprocal hybrid crosses in each species genetic background. Significance was defined at a FDR < 0.05 for all analyses. For each contrast, the dataset was filtered to contain only expressed genes, which were required to have a minimum RPKM value of four for all replicates in at least one species/hybrid. To determine testes-specific genes, we required each sample to have a carcass RPKM value of one or less. A multivariate ANOVA in R was performed to determine the effect of tissue, species, and their interaction on chromosome-wide gene expression. To compare differences in expression between chromosomes, but within a species/tissue, we performed pairwise t-tests with Holm-Bonferroni correction.

Hybrid-Parental Expression Pattern Classification

Differentially expressed gene RPKM values of the F1 and parental testes were used to determine the dominance to additive effect of expression using the d/a ratio formula modified from Supar et al. (Stupar et al., 2008). Two ratios (d/a) were determine by looking at either High versus Low expression (T1 ratio) or Maternal versus Paternal expression (T2 ratio). To determine the T1 ratio, the d value = F1 – ((High expressing parent + Low expressing parent)/2), and the a value = High expressing parent - ((High expressing parent + Low expressing parent)/2). Additive expression would result in a d/a ratio of 0, parent-like

expression would be a d/a ratio of 1 or -1, where 1 is High expressing parent-like and -1 is Low expressing parent-like. Genes where the d/a ratio is greater than 1 or less than -1 are non-additively expressed, which greater than 1 being expressed more than the High expressing parent, and less than -1 being expressed less than the Low expressing parent. The T2 ratio was determined by using the d value = F1 – ((Parental expression + Maternal expression)/2), and a value = Parental expression – ((Parental expression + Maternal expression)/2). Additive expression would result in a d/a ratio of 0, parent-like expression would be a d/a ratio of 1 or -1, where 1 is Paternal-like and -1 is Maternal-like. Non-additive expression, where the d/a ratio is greater than 1 or less than -1, are expressed outside of the parental range. Using the arbitrary ranges set up in Renaut et al. 2009, the expression of a gene was considered additive if the d/a ratio fell between -0.5 and +0.5, dominant if the ratio was -1.5 to -0.5 (Maternally dominant) or +0.5 to +1.5 (Paternally dominant), and non-additive if the ratio was greater than +1.5 or less than -1.5. The ratios were applied to each chromosome separately. Outliers outside the 1.5 Interquartile range were identified using Tukey's method and removed (Dhana, 2016). A t-test was used to find if Type I and Type II ratio means for each chromosome were significantly different than -1, 0, and 1. The Gene Ontology terms of the genes that fell within each range for each hybrid were analyzed using PantherDB's statistical overrepresentation test (Thomas et al., 2003). Genes were not separated by chromosome for Gene Ontology analysis.

Results and Discussion

Transcriptional Repression in Parental Species

From quantitative genomic analysis of backcross progeny we can infer that the X chromosome plays a critical role in the sterility of CxM^{F1} males, but a more minor role in MxC^{F1} sterility (Unpublished data Chapter II). As detailed in the introduction, improper inactivation/repression of the X chromosome in the male germline may be an underappreciated cause of sterility in hybrid males (Lifschytz and Lindsley, 1972). We hypothesized that the disproportionate role of the *An. coluzzii* X chromosome in sterility is due to impaired silencing of this chromosome during CxM^{F1} spermatogenesis. Under this model, dominant *An. merus* factors fail to properly repress the *An. coluzzii* X in CxM^{F1} hybrids resulting in widespread overexpression of X-linked genes and impaired sperm development. To test the hypothesis, we performed RNA-SEQ on both testes and carcass (body without testes) of adult males from each species and testes of reciprocal F1 hybrid males. If DMIs cause misregulation and subsequent over-expression of the *An. coluzzii* X chromosome, we make two key predictions concerning gene expression. First, X-linked, but not autosomal genes, should be highly down regulated in the testes of both *An. coluzzii* and *An. merus*, proving that the X chromosome is inactivated during normal sperm development. Second, X-linked, but not autosomal, genes should be widely overexpressed in the testes of CxM^{F1} males relative to both parent species. In MxC^{F1} males, where

the X plays a more minor role in sterility, sex-linked genes should not be overexpressed in the male germline.

Illumina sequencing of 23 cDNA libraries followed by stringent quality filtering resulted in 444 million 100 bp, paired-end reads for downstream analysis. Preprocessed reads were mapped to the newest reference gene sets for each species and RPKM normalized transcript expression levels were calculated for each sample independently (Trapnell et al., 2010). Approximately 46% (*n* = 6,198) and 43% (*n* = 6,142) of genes were expressed in *An. coluzzii* and *An. merus* testes, respectively. For the subset of genes expressed in *An. coluzzii* testes, 445 were testes-specific (i.e. expressed only in testes) and 2,567 were testes-enriched (i.e. upregulated in testes relative to carcass), while *An. merus* had 527 testes-specific and 1,641 testes-enriched genes. Fewer genes were carcass-specific and carcass-enriched in both *An. coluzzii* (*n* = 325 and 1,159, respectively) and *An. merus* (*n* = 431 and 847, respectively). A chi-squared test with Yates correction revealed a significant underrepresentation of testes-expressed, -enriched, and -specific genes on the X chromosome of both species ($P < 1 \times 10^{-4}$ for all comparisons; Table 3.1).

Next, we calculated mean transcript abundance of testes-expressed genes for each chromosome. As shown in Figure 3.2, average RPKM are similar for genes located on the two autosomes in both *An. coluzzii* and *An. merus* (*P > 0.2* for all comparisons). However, X-linked genes show much lower expression, with average RPKM values 50-75% lower than autosomal genes in both species

($P < 1 \times 10^{-6}$ for all comparisons). No similar down-regulation of X-linked genes is observed in the carcass ($P > 0.4$ in all comparisons). Previous tissue-specific expression profiling of *An. coluzzii* using microarrays uncovered a near identical pattern, with X-linked genes being strongly repressed in the testes, but not somatic tissues (Baker et al., 2011). Additionally, Magnusson *et al* (Magnusson et al., 2012) recently utilized transgenic constructs to prove that when active testes autosomal promoters are transferred to the X they do not induce high male germline expression. Taken together with these two previous studies, our results demonstrate that – as in a diversity of eukaryotes – the X chromosome is inactivated/repressed in *An. gambiae* complex members during spermatogenesis. However, it is unclear if transcriptional silencing of the sex chromosomes in *Anopheles* occurs via mammalian-like MSCI or an alternative pathway as appears to be the case in *Drosophila* (Meiklejohn et al., 2011).

Differential Expression in Hybrids

Having established that the X chromosome is transcriptionally repressed during spermatogenesis in both parental species, we looked for evidence that the X-silencing process is impaired in hybrids. Examination of testes transcripts levels in reciprocal F1 hybrids revealed an exceptional pattern (Figure 3.2). As in the parental species, MxC^{F1} males show strong downregulation of X-linked, but not autosomal genes in the germline ($P < 1 \times 10^{-14}$). In striking contrast, mean expression levels of X-linked genes in the testes of CxM^{F1} hybrids are not

significantly different from the expression level of autosomal genes (*P > 0.6*). This pattern very strongly suggests that the *An. coluzzii* X chromosome is not properly repressed when in a hybrid autosomal background. Next, we identified genes that were differentially expressed between the testes of parents and CxM^{F1} hybrids. On the X chromosome, 294 genes are significantly differentially expressed between the testes of *An. coluzzii* and CxM^{F1} hybrids; of these genes, a staggering 290 (98.6%) are overexpressed in the hybrid. A similar result is observed when comparing the testes of the CxM^{F1} hybrids to *An. merus*; of the 226 X-linked genes differentially expressed in the testes, 220 (97.4%) are overexpressed in the hybrid. In support of X-specific misregulation, we do not observe an excess of overexpressed genes on either autosome (Table 3.2). In fact, when compared to either parent, CxM^{F1} hybrids show a slight excess of under-expressed autosomal genes. Further, examination of volcano plots (Figure 3.3) shows that nearly all X-linked genes (significant and non-significant) show higher expression in CxM^{F1} hybrids relative to both parents, whereas autosomal gene expression is much more balanced between the hybrid and parents.

As expected, the genomic distribution of differentially expressed testes genes between MxC^{F1} hybrids and parental species does not follow the same pattern. In fact, when differentially expressed, most X-linked genes are under-expressed in MxC^{F1} hybrids relative to both parents (159/170 or 93.5% when compared to *An. coluzzii*; 98/114 or 86.0% when compared to *An. merus;* Table

3.2). Volcano plots also show a skew towards under-expression of most X-linked genes in the hybrid, albeit not as dramatic as the overexpression seen in MxC^{F1} hybrids (Figure 3.4). The excess of under-expressed X genes is surprising as no significant difference is observed in mean X-linked transcript abundance between the testes of MxC^{F1} males and parents (Figure 3.2). In contrast to the X, differentially expressed autosomal genes show a bias towards over-expression in MxC^{F1} males. In sum, our RNA-SEQ results strongly support the possibility that improper repression of the *An. coluzzii* X chromosome by dominant *An. merus* autosomal factors underlies the large influence of the X on CxM^{F1} male sterility.

Non-Additive Expression in Hybrids

We compared levels of additive, parental-like, and non-additive expression of all the genes expressed in the testes of each hybrid using two dominance/additive (d/a) ratio tests. As explained in the methods, both ratio types give the same F1 genes that are additively expressed. The Type I d/a ratio is used to determine if the F1 transcript level is closer to the higher-expressing parent or the lower-expressing parent, and if the non-additive expressing genes are more highly expressed than the high parent or lower than the low parent. Gene that were expressed above our cut-off for the high-expressing parent are considered over-expressed, and those genes expressing lower than the low-expressing parent are considered under-expressed. The Type II d/a ratio gives us whether the F1 expression is like the Maternal or Paternal expression, or outside

the parental range. We see in Table 3.4 that the *CxM*^{F1} follows the pattern of expression that we found where, on the X chromosome, 93.9% of expressed genes are non-additively over-expressed. We do see a divergence from our volcano plots with the autosome expression, were we find that the highest percentage of gene were non-additively over-expressed, this is because we are taking into account both parental males, where our differential expression analysis looks at the comparisons between F1 and each individual parental. In Figure 3.5 we see the mean expression and find that chromosome 3's average most closely resembles the high expressing parent (mean= 0.985, P = 0.8383, mu = 1) while chromosome 2 and the X chromosome have means that fall outside of the parental ranges no matter which ratio type was used (Figure 3.5 and 3.6). In *MxC*^{F1} males we see the same pattern as found by our differential expression analysis, a larger percentage of X-linked genes are under-expressed (19.7% under-expressed versus 8.3% over-expressed) and a higher proportion of genes on the autosomes that are over-expressed (Chromosome 2 = 46.5%, Chromosome 3 = 40.2%). Interestingly, *MxC*^{F1} males also show a higher percentage of additively expressed genes on X than non-additive (37.9% additive versus 28% non-additive) and a higher percentage than either of the autosomes (Table 3.3). The Type II ratios show us that there is relatively even amounts of genes in *CxM*^{F1} males that are expressed either maternal-like or paternal-like, all chromosomes taken together 13.8% are expressed like paternal genes, and 13.4% are expressed like maternal genes (Table 3.4). The *MxC*^{F1} males show

slightly more variation, with 16.9% of genes expressing paternal-like and 14.6% expressing maternal-like. However, this is misleading because the X chromosome actually follows maternal expression more than paternal expression (20.4% maternal versus 14.9% paternal). While none of the chromosomal ratio means for the Type I ratio are similar enough to either parental or additive expression (Figure 3.5), the Type II ratio does show that the X chromosome is being expressed in a mostly additive fashion (mean= -0.103, P = 0.2229, mu=0)(Figure 3.6). This agrees with our finding that the X chromosome has a less important role in MxC^{F1} hybrid sterility.

Gene Ontology analysis of Hybrid additive, non-additive, and parental-like expression

Given that the genes most likely to influence fertility in the hybrids are those that are non-additively expressed, we explored them further using an enrichment test (Renaut et al., 2009; Stupar et al., 2007). We also analyzed those genes that fell within the following categories for each hybrid: additively expressed, maternal-like expression, paternal-like expression and their results are in Table 3.5, 3.6, and 3.7, respectively. For CxM^{F1} males we found that the most enriched genes for each GO category that were non-additively over-expressed were structural constituent of ribosome (1.98 Fold Enrichment (FE), Molecular Function (MF)), vesicle-mediated transport (1.87 FE, Biological Process (BP)), and endoplasmic reticulum (2.14 FE, Cellular Component

(CC))(Table 3.8 A). There were also many under-represented GO terms involved in receptor activity, sensory perception, and unclassified genes (Table 3.8 A). Those genes that were non-additively under-expressed in CxM^{F1} males were enriched for nucleoside-triphosphatase activity (2.24 FE, MF), protein modification by small protein removal (4.46 FE, BP), and outer membrane (6.69 FE, CC) (Table 3.9 A). The MxC^{F1} genes that were over-expressed were found to be enriched for NAD binding (3.23 FE, MF), vesicle-mediated transport (1.71 FE, BP), and intercellular organelle lumen (1.67 FE, CC) (Table 3.8 B). Like the CxM^{F1} over-expressed under-represented genes, the MxC^{F1} males were also under-represented for sensory perception. The under-expressed MxC^{F1} genes over-represented in our GO analysis were cellular protein metabolic process (2.03 FE, BP) and peptidase complex (6.48 FE, CC) (Table 3.9 B).

Insights into the Rapid Evolution of Male Hybrid Sterility

Based on our results, we argue that such misregulation of the X may partially explain why hybrid male sterility appears more rapidly than other types of post-zygotic isolation following lineage splitting. In males with heteromorphic sex chromosomes, both sex chromosomes are inactivated/repressed during male gametogenesis, although the cellular and molecular processes leading to transcriptional silencing may vary between taxa. Despite its ubiquity and clear importance to organism fitness, there is good reason to believe that the genes regulating sex chromosome expression in the germline may be rapidly evolving

relative to rest of the genome. It has been hypothesized that sex chromosome repression is a result of evolutionary pressures to suppress selfish genetic elements including both segregation distorters and transposable elements (Ellis et al., 2005; Haig and Grafen, 1991; Hamilton, 1967; Meiklejohn and Tao, 2010; Tao et al., 2007b). If the theory is correct, autosomal factors that recognize and repress conspecific X chromosomes may undergo rapid divergence after species split as they evolve to counteract genes on the X that attempt to circumvent the limits of normal Mendelian inheritance. When placed into a hybrid genetic background, the rapidly evolving autosomal factors that regulate expression of the X in the male germline may not properly repress the heterospecific X due to the independent 'arms-races' since the two species split (Lifschytz and Lindsley, 1972; McDermott and Noor, 2010; Meiklejohn et al., 2011). In support of this hypothesis, hybrid mice with disrupted MSCI produce mostly daughters suggesting that silent distorters on the X are released in the absence on inactivation (Cocquet et al., 2009). Unfortunately, we cannot test for drive in our F1 hybrid males since even old males (14+ days) exhibit no signs of fertility.

Indeed, disruption of X-repression is a very enticing explanation for hybrid male sterility and data from the CxM^{F1} male hybrids is consistent with the hypothesis. In contrast, interpreting the causes of hybrid male sterility in MxC^{F1} is not as straightforward. In this cross, it appears that sterility is caused by complex DMIs involving negative interactions between genes on all three chromosomes (Unpublished data, Chapter II). Overall, the X chromosome appears to be less

influential on the expression of hybrid sterility than either autosome. In the absence of major effect loci on the sex chromosomes, it is unclear why MxC^{F1} male sterility evolved prior to male inviability or any type of female dysfunction. Without identification of specific DMI genes we cannot test the hypothesis that sexual selection on autosomal male fertility factors drove the rapid appearance of sterility in MxC^{F1} males (Malone and Michalak, 2008; Presgraves and Orr, 1998; Wu and Davis, 1993). Interestingly, we do observe an excess of under-expressed X-linked genes in MxC^{F1} hybrids. Although the misregulation is not as gross as that observed in CxM^{F1} hybrids it does warrant further investigation. In sum, our data demonstrates that male sterility may arise through different mechanisms and highlights improper sex chromosome gene regulation as a potentially underappreciated and taxonomically widespread (from mice to mosquitoes) mechanism underlying Haldane's Rule.

A Role for Z Misregulation in Female Sterility?

In *Aves* (birds) and *Lepidoptera* females are heterogametic (ZW), while males are homogametic (ZZ). In crosses that display unisexual sterility, both taxa largely adhere to Haldane's Rule with female sterility appearing first (Aves 24/27; Lepidoptera 17/18). If either increased sexual selection on male fertility factors or spermatogenic sensitivity were a major driver of male sterility, one would predict that ZW taxa should violate Haldane's rule with respect to sterility. As this is clearly not the case, we hypothesize that ZW female sterility could also be

caused by sex chromosome misregulation during gametogenesis. While the most recent cytogenetic study found no evidence of mammalian-like MSCI during chicken oogenesis (Guioli et al., 2012; Schoenmakers et al., 2009), various transcriptomic analyses do find that expression of Z-linked genes is highly repressed in female meiotic cells (Ellegren, 2011; Storchova and Divina, 2006), mirroring the pattern of X chromosome expression in the testes of XY males. It is possible that, like *Drosophila* males, female chickens utilize a non-canonical epigenetic pathway to as least transiently silence the Z chromosome during meiosis (Schoenmakers et al., 2010). Additional data from other taxa with heterogametic females, especially *Lepidoptera*, would allow for more definitive conclusions and potentially provide better insight into the evolutionary drivers of meiotic sex chromosome silencing. Although further evidence is needed, X (Z) misregulation potentially explains why both male and female heterogametic taxa so closely follow Haldane's with respect to sterility.

Overexpression and the Large X-Effect

We have found evidence for a large X-effect in CxM^{F1} hybrids (Unpublished data, Chapter II) and targeted transcriptomics suggests it is a consequence of X chromosome overexpression during spermatogenesis. Improper repression of the X should lead to a higher density – but not necessarily increased effect size – of X-linked sterility factors, since most overexpressed chromosomal segments have the potential to cause defects in sperm development via misregulation of key fertility genes or the release of selfish

genetic elements. Thus, we argue that X-overexpression in the hybrid germline is an intriguing mechanistic explanation for the large X-effect and warrants further taxonomic and molecular exploration.

Summary and Next Steps

Using tissue-specific transcriptomics we provide evidence that disruption of X silencing underlies male sterility in one direction of a reciprocal cross between *An. coluzzii* and *An. merus*. Future efforts to profile gene expression in the testes of other sterile F1 hybrids within the An. gambiae complex would be helpful to test if this expression pattern is common within *Anopheles* species. If X overexpression were a common cause of male sterility, genes regulating X chromosome expression would be promising targets for malaria vector control through sterile insect technique (Lees et al., 2014; Nolan et al., 2011; Wang and Jacobs-Lorena, 2013).

Acknowledgements

I'd like to acknowledge Bryan Cassone for RNA-seq analysis, differential expression analysis, and PCA plots.

References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Baker, D.A., Nolan, T., Fischer, B., Pinder, A., Crisanti, A., and Russell, S. (2011). A comprehensive gene expression atlas of sex-and tissue-specificity in the malaria vector, Anopheles gambiae. BMC Genomics *12*, 296.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. 289–300.

Cabrero, J., Teruel, M., Carmona, F., Jimenez, R., and Camacho, J. (2007). Histone H3 lysine 9 acetylation pattern suggests that X and B chromosomes are silenced during entire male meiosis in a grasshopper. Cytogenet. Genome Res. *119*, 135–142.

Campbell, P., Good, J.M., and Nachman, M.W. (2013). Meiotic sex chromosome inactivation is disrupted in sterile hybrid male house mice. Genetics *193*, 819–828.

Cocquet, J., Ellis, P.J.I., Yamauchi, Y., Mahadevaiah, S.K., Affara, N.A., Ward, M.A., and Burgoyne, P.S. (2009). The Multicopy Gene Sly Represses the Sex Chromosomes in the Male Mouse Germline after Meiosis. PLoS Biol. *7*, e1000244.

Coetzee, M., Hunt, R.H., Wilkerson, R., Della Torre, A., Coulibaly, M.B., and Besansky, N.J. (2013). Anopheles coluzzii and Anopheles amharicus, new members of the Anopheles gambiae complex. Zootaxa *3619*, 246–274.

Cuamba, N., and Mendis, C. (2009). The role of Anopheles merus in malaria transmission in an area of southern Mozambique. J. Vector Borne Dis. *46*, 157–159.

Dhana, K. (2016). Identify, describe, plot, and remove the outliers from the dataset.

Ellegren, H. (2011). Emergence of male-biased genes on the chicken Z-chromosome: sex-chromosome contrasts between male and female heterogametic systems. Genome Res. *21*, 2082–2086.

Ellis, P.J.I., Clemente, E.J., Ball, P., Toure, A., Ferguson, L., Turner, J.M.A., Loveland, K.L., Affara, N.A., and Burgoyne, P.S. (2005). Deletions on mouse Yq

lead to upregulation of multiple X- and Y-linked transcripts in spermatids. Hum. Mol. Genet. *14*, 2705–2715.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. *11*, R86.

Gomes, S., and Civetta, A. (2014). Misregulation of spermatogenesis genes in Drosophila hybrids is lineage-specific and driven by the combined effects of sterility and fast male regulatory divergence. J. Evol. Biol.

Good, J.M., Giger, T., Dean, M.D., and Nachman, M.W. (2010). Widespread over-expression of the X chromosome in sterile F1 hybrid mice. PLoS Genet. *6*, e1001148.

Govere, J., Durrheim, D., Coetzee, M., Hunt, R.H., and La Grange, J. (2000). Captures of mosquitoes of the Anopheles gambiae complex (Diptera: Culicidae) in the Lowveld region of Mpumalanga Province, South Africa. Afr. Entomol. *8*, 91–99.

Guioli, S., Lovell-Badge, R., and Turner, J.M.A. (2012). Error-Prone ZW Pairing and No Evidence for Meiotic Sex Chromosome Inactivation in the Chicken Germ Line. PLoS Genet. *8*, e1002560.

Haig, D., and Grafen, A. (1991). Genetic scrambling as a defence against meiotic drive. J. Theor. Biol. *153*, 531–558.

Hamilton, W.D. (1967). Extraordinary Sex Ratios. Science *156*, 477–488.

Kelly, W.G., Schaner, C.E., Dernburg, A.F., Lee, M.-H., Kim, S.K., Villeneuve, A.M., and Reinke, V. (2002). X-chromosome silencing in the germline of C.elegans. Development *129*, 479–492.

Lees, R.S., Knols, B., Bellini, R., Benedict, M.Q., Bheecarry, A., Bossin, H.C., Chadee, D.D., Charlwood, J., Dabiré, R.K., Djogbenou, L., et al. (2014). Review: Improving our knowledge of male mosquito biology in relation to genetic control programmes. Acta Trop. *132*, S2–S11.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.

Lifschytz, E., and Lindsley, D.L. (1972). The Role of X-Chromosome Inactivation during Spermatogenesis. Proc. Natl. Acad. Sci. *69*, 182–186.

Magnusson, K., Lycett, G.J., Mendes, A.M., Lynd, A., Papathanos, P.-A., Crisanti, A., and Windbichler, N. (2012). Demasculinization of the Anopheles gambiae X chromosome. BMC Evol. Biol. *12*, 1–13.

Malone, J.H., and Michalak, P. (2008). Physiological Sex Predicts Hybrid Sterility Regardless of Genotype. Science *319*, 59.

McDermott, S.R., and Noor, M.A. (2010). The role of meiotic drive in hybrid male sterility. Philos. Trans. R. Soc. B Biol. Sci. *365*, 1265–1272.

Megy, K., Emrich, S.J., Lawson, D., Campbell, D., Dialynas, E., Hughes, D.S.T., Koscielny, G., Louis, C., MacCallum, R.M., Redmond, S.N., et al. (2012). VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. Nucleic Acids Res. *40*, D729–D734.

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

Meiklejohn, C.D., Landeen, E.L., Cook, J.M., Kingan, S.B., and Presgraves, D.C. (2011). Sex Chromosome-Specific Regulation in the Drosophila Male Germline But Little Evidence for Chromosomal Dosage Compensation or Meiotic Inactivation. PLoS Biol. *9*, e1001126.

Moehring, A.J., Teeter, K.C., and Noor, M.A. (2007). Genome-wide patterns of expression in Drosophila pure species and hybrid males II Examination of multiple-species hybridizations, platforms, and life cycle stages. Mol. Biol. Evol. *24*, 137–145.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

Mwangangi, J.M., Mbogo, C., Nzovu, J.G., Githure, J.I., Yan, G., and Beier, J.C. (2003). Blood-meal analysis for anopheline mosquitoes sampled along the Kenyan coast. J. Am. Mosq. Control Assoc. *19*, 371–375.

Nolan, T., Papathanos, P., Windbichler, N., Magnusson, K., Benton, J., Catteruccia, F., and Crisanti, A. (2011). Developing transgenic Anopheles mosquitoes for the sterile insect technique. Genetica *139*, 33–39.

Pock Tsy, J.-M.L., Duchemin, J.-B., Marrama, L., Rabarison, P., Le Goff, G., Rajaonarivelo, V., and Robert, V. (2003). Distribution of the species of the Anopheles gambiae complex and the first evidence of Anopheles merus as a malaria vector in Madagascar. Malar. J. *2*, 1–7.

Presgraves, D.C., and Orr, H.A. (1998). Haldane's rule in taxa lacking a hemizygous X. Science *282*, 952–954.

Ranson, H., N'Guessan, R., Lines, J., Moiroux, N., Nkuni, Z., and Corbel, V. (2011). Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? Trends Parasitol. *27*, 91–98.

Reddy, M.R., Overgaard, H.J., Abaga, S., Reddy, V.P., Caccone, A., Kiszewski, A.E., and Slotman, M.A. (2011). Outdoor host seeking behaviour of Anopheles gambiae mosquitoes following initiation of malaria vector control on Bioko Island, Equatorial Guinea. Malar. J. *10*, 184.

Renaut, S., Nolte, A.W., and Bernatchez, L. (2009). Gene Expression Divergence and Hybrid Misexpression between Lake Whitefish Species Pairs (Coregonus spp. Salmonidae). Mol. Biol. Evol. *26*, 925–936.

Schoenmakers, S., Wassenaar, E., Hoogerbrugge, J.W., Laven, J.S.E., Grootegoed, J.A., and Baarends, W.M. (2009). Female Meiotic Sex Chromosome Inactivation in Chicken. PLoS Genet. *5*, e1000466.

Schoenmakers, S., Wassenaar, E., Laven, J.S.E., Grootegoed, J.A., and Baarends, W.M. (2010). Meiotic silencing and fragmentation of the male germline restricted chromosome in zebra finch. Chromosoma *119*, 311–324.

Sinka, M.E., Bangs, M.J., Manguin, S., Coetzee, M., Mbogo, C.M., Hemingway, J., Patil, A.P., Temperley, W.H., Gething, P.W., Kabaria, C.W., et al. (2010). The dominant Anopheles vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis.

Storchova, R., and Divina, P. (2006). Nonrandom representation of sex-biased genes on chicken Z chromosome. J. Mol. Evol. *63*, 676–681.

Stupar, R.M., Hermanson, P.J., and Springer, N.M. (2007). Nonadditive Expression and Parent-of-Origin Effects Identified by Microarray and Allele-Specific Expression Profiling of Maize Endosperm. PLANT Physiol. *145*, 411–425.

Stupar, R.M., Gardiner, J.M., Oldre, A.G., Haun, W.J., Chandler, V.L., and Springer, N.M. (2008). Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. BMC Plant Biol. *8*, 33.

Tao, Y., Masly, J.P., Ke, Y., and Hartl, D.L. (2007). a sex-ratio meiotic drive system in Drosophila simulans I: An autosomal suppressor. PLOS Biol. *5*, e292.

Temu, E.A., Minjas, J.N., Coetzee, M., Hunt, R.H., and Shift, C.J. (1998). The role of four anopheline species (Diptera: Culicidae) in malaria transmission in coastal Tanzania. Trans. R. Soc. Trop. Med. Hyg. *92*, 152–158.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. *13*, 2129–2141.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Turner, J.M.A. (2007). Meiotic sex chromosome inactivation. Development *134*, 1823–1831.

Wang, S., and Jacobs-Lorena, M. (2013). Genetic approaches to interfere with malaria transmission by vector mosquitoes. Trends Biotechnol. *31*, 185–193.

White, B.J., Kundert, P.N., Turissini, D.A., Van Ekeris, L., Linser, P.J., and Besansky, N.J. (2013). Dose and developmental responses of Anopheles merus larvae to salinity. J. Exp. Biol. *216*, 3433–3441.

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

FIGURES

Figure Legends:

Figure 3.1 RNA-SEQ samples from the same tissue cluster together. Principal component analysis shows tight clustering of parental samples by tissue, while hybrid samples show less discrete clustering.

Figure 3.2 The X chromosome is not properly silenced in the testes of *CxM*^{F1} hybrids. Mean transcript abundance for expressed genes on each chromosome in 6 different species/tissue combinations are given. In parental carcass tissue (*An. coluzzii* carcass: orange, *An. merus* carcass: yellow), gene expression is equivalent regardless of chromosome. Contrastingly, in the parental testes (*An. coluzzii* testes: green, *An. merus* testes: teal), expression of the X chromosome is repressed relative to the autosomes. Testes expression in *MxC*^{F1} mosquitoes (purple) follows a similar pattern with decreased transcript abundance of X-linked genes. In striking contrast, the testes of *CxM*^{F1} mosquitoes (blue) show no downregulation of X-linked genes relative to autosomal genes. Species, tissue, and their interaction have a significant effect on chromosome-wide gene expression ($P < 0.001$). Red dots represent mean expression.

Figure 3.3 Nearly every X-linked gene is overexpressed in the testes of *CxM^{F1}* hybrids. Genes expressed in the testes are represented by a circle. Green circles are genes significantly differentially expressed ($P < 0.05$ after multiple test correction) between *CxM^{F1}* and parental mosquitoes. Genes on the right half of each graph are more highly expressed in the hybrid relative to the parent. The vast majority of genes located on the X chromosome show higher expression in *CxM^{F1}* hybrids relative to both *An. coluzzii* (top) and *An. merus* (bottom). In contrast, autosomal genes show a slight bias towards underexpression in the hybrid.

Figure 3.4 Most X-linked genes are under-expressed in the testes of *MxC^{F1}* hybrids. Genes expressed in the testes are represented by a circle. Green circles are genes significantly differentially expressed ($P < 0.05$ after multiple test correction) between *MxC^{F1}* and parental mosquitoes. Genes on the left half of each graph are under expressed in the hybrid relative to the parent. A majority of genes located on the X chromosome show lower expression in *MxC^{F1}* hybrids relative to both *An. coluzzii* (top) and *An. merus* (bottom). In contrast, autosomal genes show a slight bias towards overexpression in the hybrid.

Figure 3.5 Type I d/a ratio density plots. The density of Type I d/a ratios for each chromosome for each hybrid. The gene vertical line is at -1 and represents genes expressing like the low-expressing parent and the orange line is at 1 and represents high-expressing parent-like gene expression. The center black line is at zero, representing additive expression.

Figure 3.6 Type II d/a ratio density plots. The density of Type II d/a ratios for each chromosome for each hybrid. The red line is for *An. coluzzii*-like expression and the blue line is for *An. merus*-like expression. The center black line is at zero, representing additive expression.

Figure 3.1

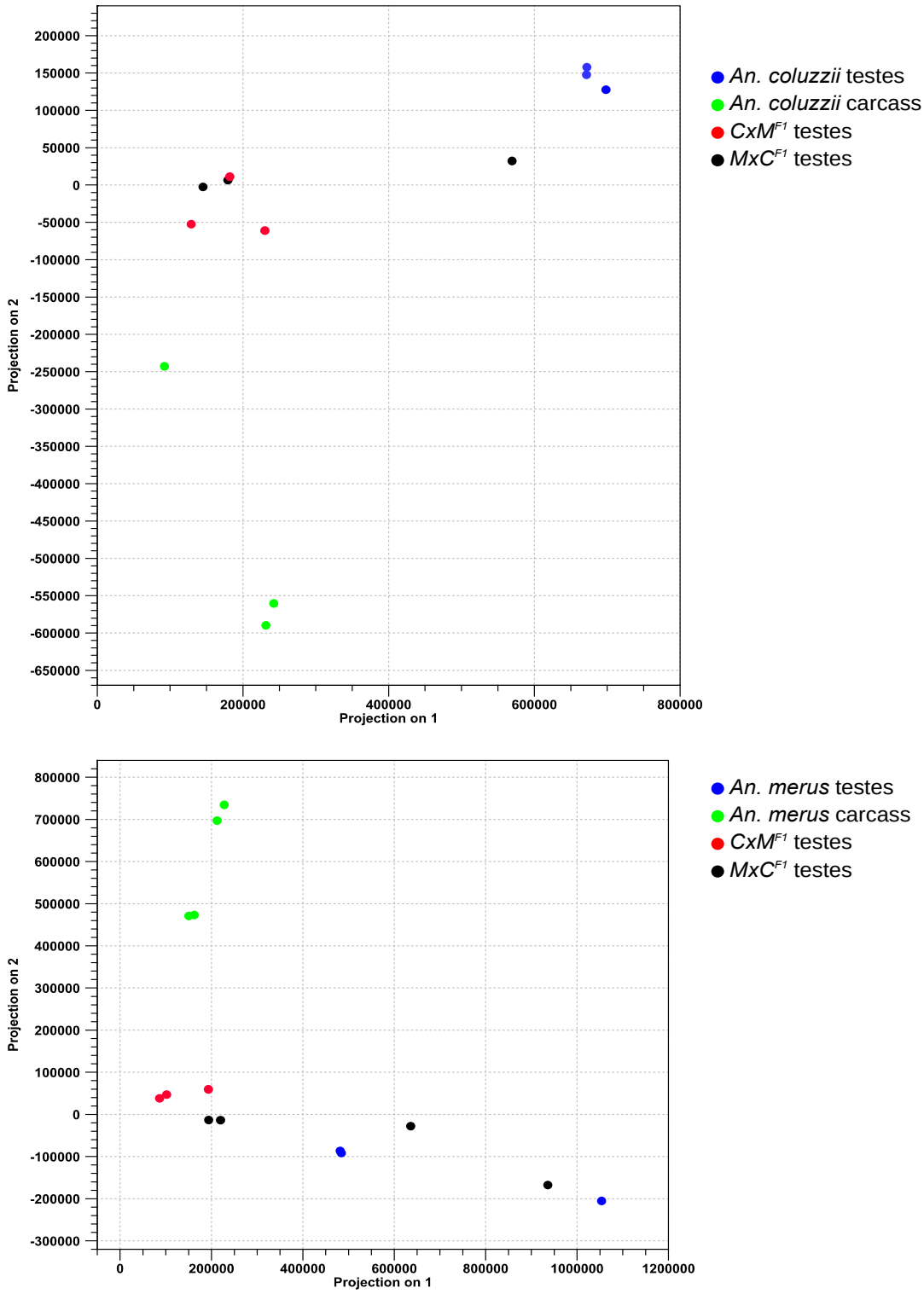

200000
150000
100000
50000
0
-50000
-100000
-150000
-200000
-250000
-300000
-350000
-400000
-450000
-500000
-550000
-600000
-650000
Projection on 2
0
200000
400000
600000
800000
Projection on 1
An. coluzzii testes
An. coluzzii carcass
CxM^F1 testes
MxC^F1 testes
800000
700000
600000
500000
400000
300000
200000
100000
0
-100000
-200000
-300000
Projection on 2
0
200000
400000
600000
800000
1000000
1200000
Projection on 1
An. merus testes
An. merus carcass
CxM^F1 testes
MxC^F1 testes

Figure 3.2

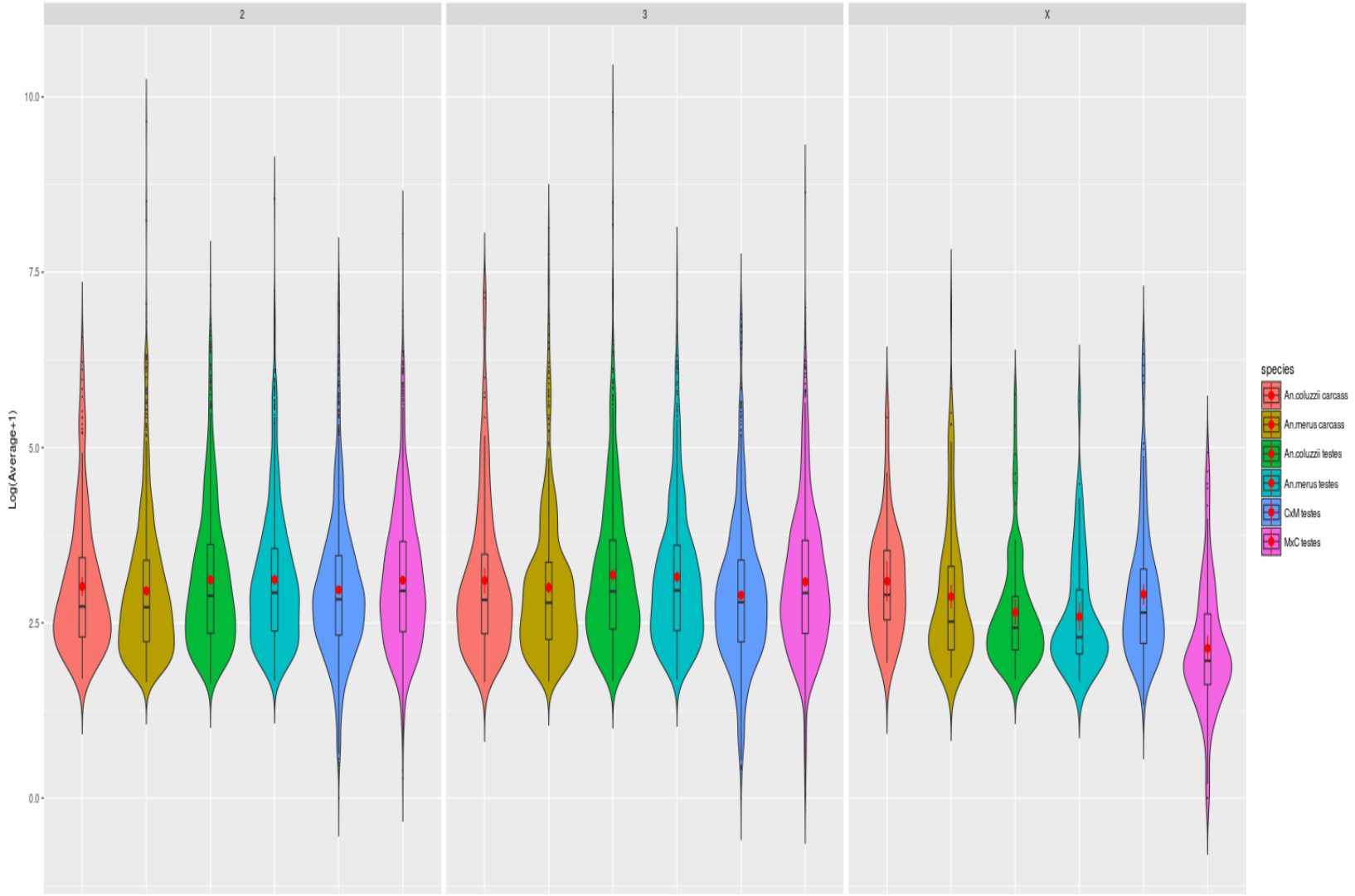

2
3
X
10.0
7.5
5.0
2.5
0.0
Log(Average+1)
species
An.coluzzii carcass
An.merus carcass
An.coluzzii testes
An.merus testes
CxM testes
MxC testes

Figure 3.3

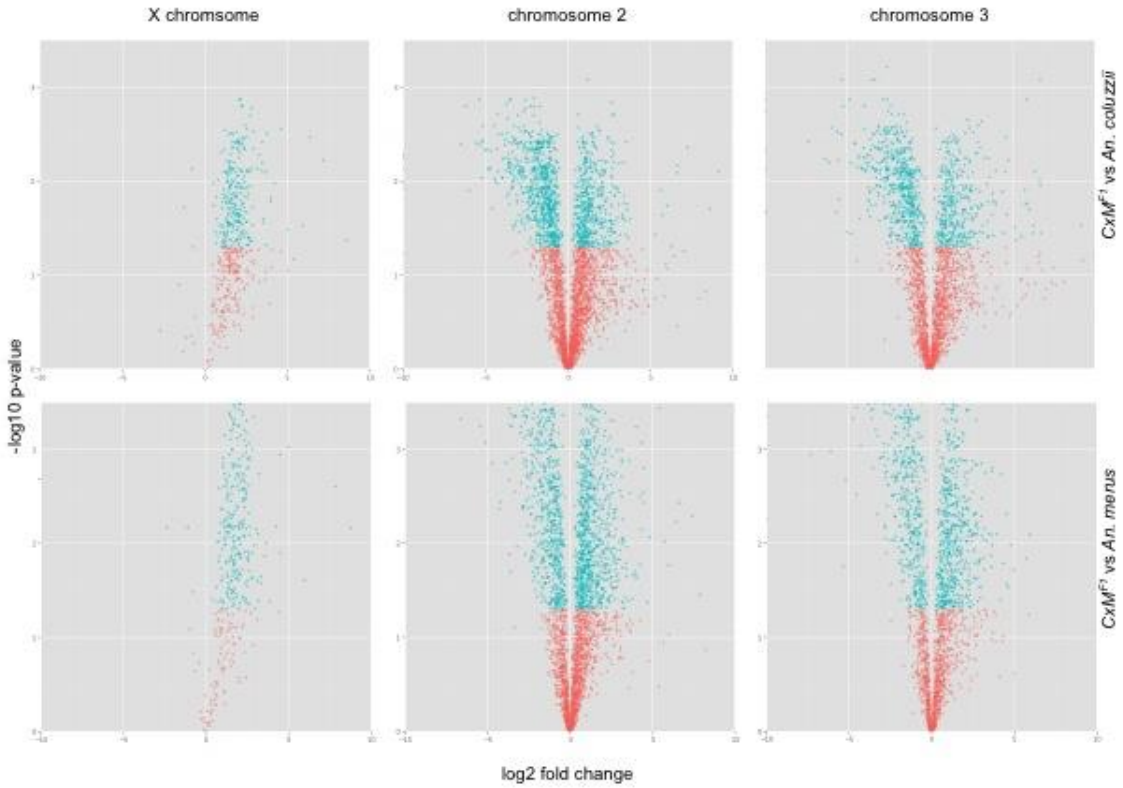

X chromsome
chromosome 2
chromosome 3
CxM^F1 vs An. coluzzii
CxM^F1 vs An. merus
-log10 p-value
log2 fold change

Figure 3.4

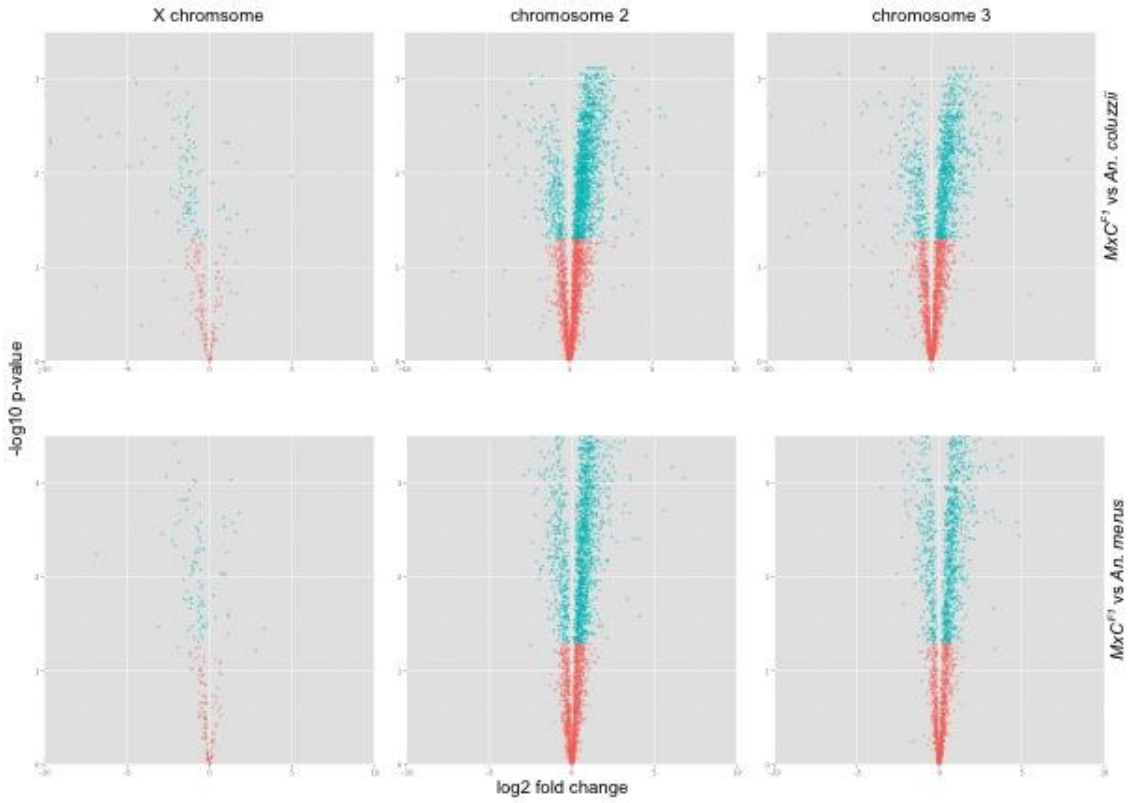

X chromsome
chromosome 2
chromosome 3
-log10 p-value
log2 fold change
MxC^F1 vs An. coluzzii
MxC^F1 vs An. menus

Figure 3.5

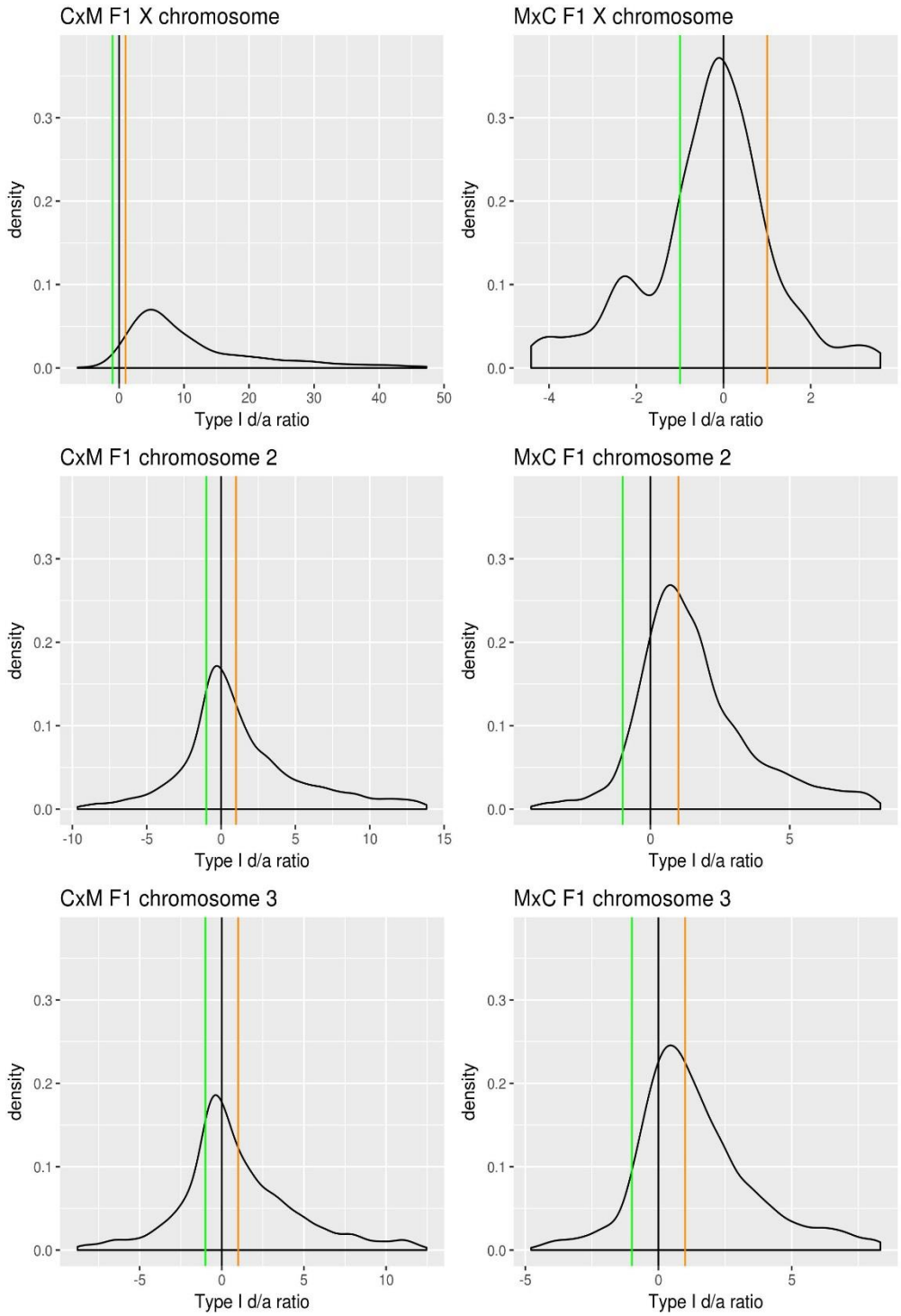
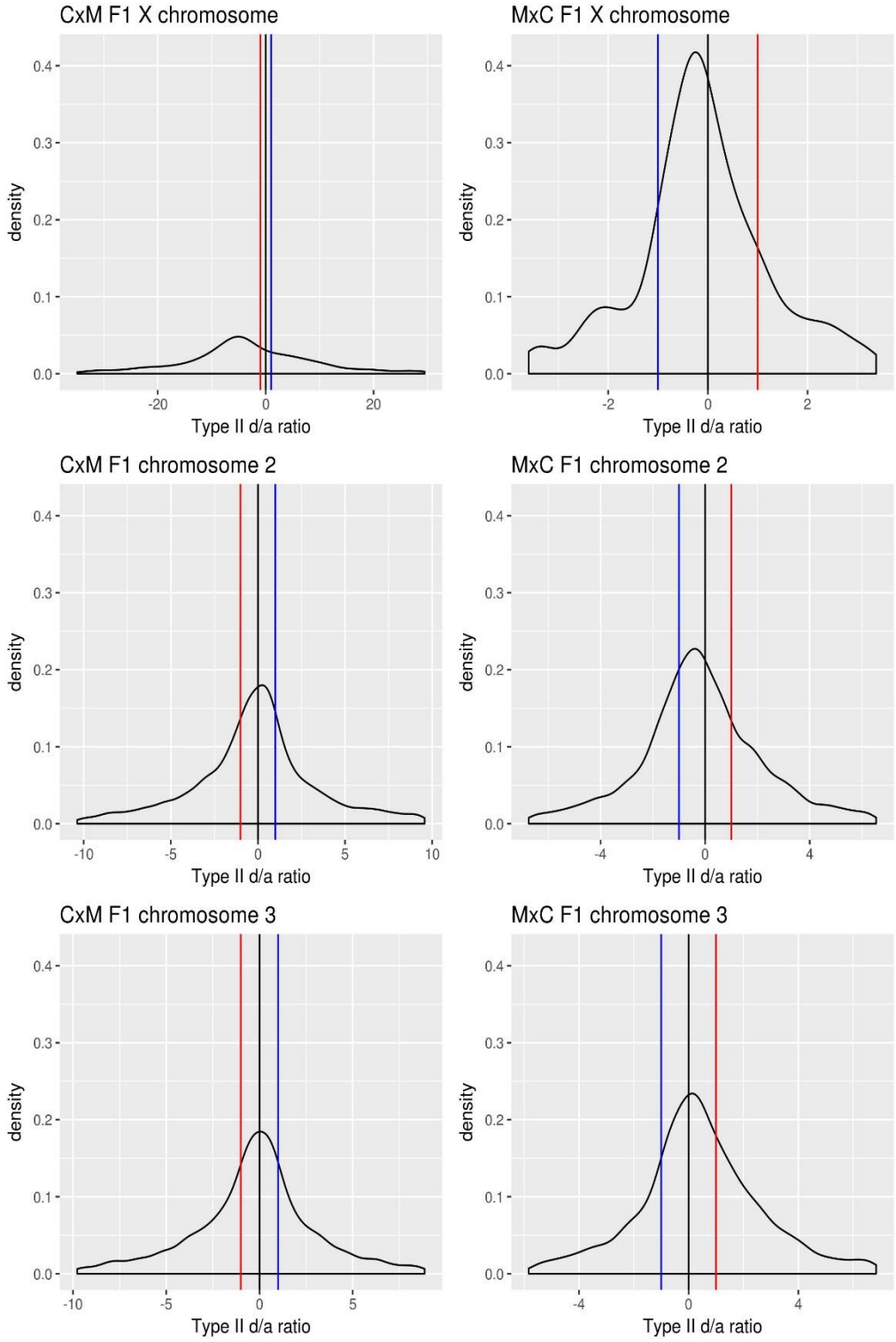

CxM F1 X chromosome
MxC F1 X chromosome
CxM F1 chromosome 2
MxC F1 chromosome 2
CxM F1 chromosome 3
MxC F1 chromosome 3
density
Type I d/a ratio

Figure 3.6


CxM F1 X chromosome
MxC F1 X chromosome
CxM F1 chromosome 2
MxC F1 chromosome 2
CxM F1 chromosome 3
MxC F1 chromosome 3
density
Type II d/a ratio

TABLES

Table 3.1 Chromosomal distribution of testes transcripts.

	Total	Expr (Obs)	Expr (Exp)	Spec (Obs)	Spec (Exp)	Up (Obs)	Up (Exp)
col chrom 2	6925	3413	3226	222	**232**	1692	1572
col chrom 3	4979	2403	2320	197	167	1222	1130
col X chrom	1097	242	511	17	37	38	249
mer chrom 2	5189	2961	2754	187	189	1012	933
mer chrom 3	3556	1969	1887	155	1129	710	639
mer X chrom	905	191	480	9	33	13	163

Total, number of genes on each chromosome; **Expr (obs)**, genes expressed in testes per chromosome; **Exp (exp)**, genes expected to be expressed in testes per chromosome; **Spec**, testes specific genes per chromosome; **Up**, number of genes significantly upregulated in testes relative to carcass per chromosome.

Table 3.2 Chromosomal distribution of genes differentially expressed between the testes of hybrids and parents.

Comparison	Chromosome 2		Chromosome 3		X chromosome	
	Over	Under	Over	Under	Over	Under
CxM^{F1} vs *coluzzii*	478	1054	364	784	290	4
CxM^{F1} vs *merus*	535	974	376	685	220	6
MxC^{F1} vs *coluzzii*	1112	674	669	597	11	159
MxC^{F1} vs *merus*	564	988	572	442	16	98

Over, number of genes significantly overexpressed in the hybrid; under, number of genes significantly underexpressed in the hybrid

Table 3.3 Type I d/a ratio

Chromosome	Additive	High-like	Low-like	Non-additive >1.5	Non-additive < -1.5	Chromosome total
CxM F1						
2	585 (16.9%)	432 (12.4%)	532 (15.3%)	1291 (37.2%)	630 (18.2%)	3470
3	448 (18.3%)	285 (11.7%)	405 (16.6%)	872 (35.6%)	436 (17.8%)	2446
X	12 (2.6%)	13 (2.8%)	4 (0.9%)	432 (93.3%)	2 (0.4%)	463
total	1045 (16.4%)	730 (11.4%)	941 (14.8%)	2595 (40.7%)	1068 (16.7%)	6379
MxC F1						
2	683 (19.4%)	766 (21.7%)	267 (7.6%)	1639 (46.5%)	172 (4.9%)	3527
3	576 (23.1%)	559 (22.4%)	226 (9%)	1004 (40.2%)	132 (5.3%)	2497
X	100 (37.9%)	41 (15.5%)	49 (18.6%)	22 (8.3%)	52 (19.7%)	264
total	1359 (21.6%)	1366 (21.7%)	542 (8.6%)	2665 (42.4%)	356 (5.7%)	6288

Non-additive > 1.5 is over-expressed; Non-additive < -1.5 is under-expressed.

Table 3.4 Type II d/a ratio

Chromosome	Additive	Paternal-like	Maternal-like	Non-additive	Chromosome total
CxM F1					
2	585 (17.6%)	489 (14.7%)	475 (14.3%)	1771 (53.3%)	3320
3	448 (18.8%)	348 (14.6%)	342 (14.4%)	1241 (52.2%)	2379
X	12 (2.7%)	11 (2.5%)	6 (1.4%)	411 (93.4%)	440
total	1045 (17%)	848 (13.8%)	823 (13.4%)	3423 (55.8%)	6139
MxC F1					
2	683 (20.2%)	555 (16.4%)	478 (14.1%)	1663 (49.2%)	3379
3	576 (23.7%)	429 (17.7%)	356 (14.7%)	1067 (43.9%)	2428
X	100 (39.2%)	38 (14.9%)	52 (20.4%)	65 (25.5%)	255
total	1359 (22.4%)	1022 (16.9%)	886 (14.6%)	2795 (46.1%)	6062

Non-additive are genes expressing outside the parental range.

Table 3.5 GO over-representation test results for additively expressed hybrid genes.

A. *CxM*^{F1}

GO Term	GO #	# in REF	#	expected	Fold Enrichment	Pvalue
Molecular Function						
RNA binding	(GO:0003723)	347	49	25.88	1.89	2.00E-02
Unclassified	(UNCLASSIFIED)	5202	433	387.91	1.12	0.00E+00
signal transducer activity	(GO:0004871)	401	11	29.9	-0.37	4.56E-02
serine-type peptidase activity	(GO:0008236)	330	6	24.61	-0.24	5.59E-03
serine hydrolase activity	(GO:0017171)	330	6	24.61	-0.24	5.59E-03
serine-type endopeptidase activity	(GO:0004252)	306	5	22.82	-0.22	5.52E-03
Biological Process						
cellular macromolecule metabolic process	(GO:0044260)	1854	184	138.25	1.33	3.27E-02
Unclassified	(UNCLASSIFIED)	5766	452	429.97	1.05	0.00E+00
multicellular organismal process	(GO:0032501)	460	12	34.3	-0.35	9.12E-03
system process	(GO:0003008)	226	3	16.85	< -0.2	4.74E-02
sensory perception of smell	(GO:0007608)	180	1	13.42	< -0.2	2.26E-02
sensory perception of chemical stimulus	(GO:0007606)	195	1	14.54	< -0.2	7.82E-03
sensory perception	(GO:0007600)	211	1	15.73	< -0.2	2.50E-03
Cellular Component						
macromolecular complex	(GO:0032991)	1415	145	105.52	1.37	2.49E-02
nucleus	(GO:0005634)	1429	145	106.56	1.36	3.93E-02
intracellular membrane-bounded organelle	(GO:0043231)	2246	224	167.48	1.34	8.44E-04
membrane-bounded organelle	(GO:0043227)	2308	229	172.11	1.33	8.97E-04
intracellular part	(GO:0044424)	3357	331	250.33	1.32	1.55E-06
intracellular organelle	(GO:0043229)	2635	257	196.49	1.31	6.13E-04
organelle	(GO:0043226)	2655	258	197.98	1.3	7.76E-04
intracellular	(GO:0005622)	3616	341	269.64	1.26	1.04E-04
cell part	(GO:0044464)	4141	363	308.79	1.18	4.45E-02
cell	(GO:0005623)	4166	365	310.66	1.17	4.37E-02
Unclassified	(UNCLASSIFIED)	6047	443	450.92	-0.98	0.00E+00
intrinsic component of membrane	(GO:0031224)	2572	145	191.79	-0.76	2.31E-02
integral component of membrane	(GO:0016021)	2563	144	191.12	-0.75	2.00E-02
extracellular region	(GO:0005576)	356	6	26.55	-0.23	6.39E-04

B. *MxC*^{F1}

GO Term	GO #	# in REF	#	expected	Fold Enrichment	Pvalue
Molecular Function						
Unclassified	(UNCLASSIFIED)	5202	496	514.65	-0.96	0.00E+00
olfactory receptor activity	(GO:0004984)	124	1	12.27	< -0.2	4.88E-02
structural constituent of cuticle	(GO:0042302)	154	1	15.24	< -0.2	2.97E-03
Biological Process						
cellular macromolecule localization	(GO:0070727)	265	49	26.22	1.87	4.16E-02
protein localization	(GO:0008104)	295	54	29.19	1.85	2.22E-02
cellular localization	(GO:0051641)	368	65	36.41	1.79	9.92E-03
organonitrogen compound biosynthetic process	(GO:1901566)	425	70	42.05	1.66	4.11E-02
organelle organization	(GO:0006996)	607	93	60.05	1.55	3.46E-02
cellular component organization	(GO:0016043)	830	122	82.11	1.49	1.30E-02
cellular component organization or biogenesis	(GO:0071840)	950	137	93.99	1.46	9.20E-03
single-organism cellular process	(GO:0044763)	1597	209	158	1.32	1.93E-02
cellular macromolecule metabolic process	(GO:0044260)	1854	237	183.42	1.29	2.22E-02
cellular metabolic process	(GO:0044237)	2553	326	252.58	1.29	3.19E-04
cellular process	(GO:0009987)	4000	487	395.73	1.23	2.50E-05
Unclassified	(UNCLASSIFIED)	5766	528	570.45	-0.93	0.00E+00
multicellular organismal process	(GO:0032501)	460	18	45.51	-0.4	2.52E-03
sensory perception of smell	(GO:0007608)	180	2	17.81	< -0.2	3.39E-03
sensory perception of chemical stimulus	(GO:0007606)	195	2	19.29	< -0.2	8.76E-04
sensory perception	(GO:0007600)	211	2	20.87	< -0.2	2.04E-04
neurological system process	(GO:0050877)	219	2	21.67	< -0.2	9.80E-05
system process	(GO:0003008)	226	2	22.36	< -0.2	5.15E-05
detection of chemical stimulus	(GO:0009593)	132	1	13.06	< -0.2	3.24E-02
detection of chemical stimulus involved in sensory perception	(GO:0050907)	132	1	13.06	< -0.2	3.24E-02
detection of stimulus involved in sensory perception	(GO:0050906)	133	1	13.16	< -0.2	2.96E-02
detection of stimulus	(GO:0051606)	143	1	14.15	< -0.2	1.16E-02
Cellular Component						
ribosome	(GO:0005840)	165	34	16.32	2.08	3.40E-02
cytoplasmic part	(GO:0044444)	1281	196	126.73	1.55	3.27E-07
cytoplasm	(GO:0005737)	1932	292	191.14	1.53	1.86E-11
intracellular non-membrane-bounded organelle	(GO:0043232)	656	97	64.9	1.49	3.31E-02
non-membrane-bounded organelle	(GO:0043228)	656	97	64.9	1.49	3.31E-02
macromolecular complex	(GO:0032991)	1415	205	139.99	1.46	9.14E-06
organelle part	(GO:0044422)	1363	192	134.85	1.42	1.93E-04
intracellular organelle part	(GO:0044446)	1350	189	133.56	1.42	3.62E-04
intracellular part	(GO:0044424)	3357	442	332.12	1.33	1.59E-09
intracellular organelle	(GO:0043229)	2635	344	260.69	1.32	4.57E-06
organelle	(GO:0043226)	2655	346	262.67	1.32	4.89E-06
intracellular	(GO:0005622)	3616	469	357.74	1.31	2.14E-09
membrane-bounded organelle	(GO:0043227)	2308	288	228.34	1.26	5.24E-03
intracellular membrane-bounded organelle	(GO:0043231)	2246	280	222.2	1.26	7.81E-03
cell	(GO:0005623)	4166	506	412.16	1.23	5.24E-06
cell part	(GO:0044464)	4141	501	409.68	1.22	1.18E-05
Unclassified	(UNCLASSIFIED)	6047	538	598.25	-0.9	0.00E+00

Table 3.6 GO over-representation test results for maternally expressed hybrid genes.

A. *CxM*^{F1}

GO Term	GO #	# in REF	#	expected	Fold Enrichment	Pvalue
Molecular Function						
ribonucleotide binding	(GO:0032553)	804	82	47.29	1.73	9.81E-04
purine ribonucleotide binding	(GO:0032555)	795	81	46.76	1.73	1.19E-03
purine nucleotide binding	(GO:0017076)	796	81	46.82	1.73	1.25E-03
purine ribonucleoside triphosphate binding	(GO:0035639)	793	80	46.64	1.72	2.01E-03
nucleotide binding	(GO:0000166)	1092	105	64.23	1.63	4.24E-04
nucleoside phosphate binding	(GO:1901265)	1092	105	64.23	1.63	4.24E-04
carbohydrate derivative binding	(GO:0097367)	904	85	53.17	1.6	1.33E-02
small molecule binding	(GO:0036094)	1155	108	67.94	1.59	1.04E-03
anion binding	(GO:0043168)	990	91	58.23	1.56	1.52E-02
heterocyclic compound binding	(GO:1901363)	2368	190	139.28	1.36	2.07E-03
organic cyclic compound binding	(GO:0097159)	2376	190	139.75	1.36	2.58E-03
Unclassified	(UNCLASSIFIED)	5202	298	305.97	-0.97	0.00E+00
serine-type endopeptidase activity	(GO:0004252)	306	3	18	< -0.2	1.23E-02
Biological Process						
organelle organization	(GO:0006996)	607	66	35.7	1.85	2.11E-03
cellular component organization	(GO:0016043)	830	86	48.82	1.76	3.85E-04
cellular component organization or biogenesis	(GO:0071840)	950	95	55.88	1.7	4.20E-04
cellular protein metabolic process	(GO:0044267)	1085	104	63.82	1.63	7.79E-04
nucleic acid metabolic process	(GO:0090304)	881	82	51.82	1.58	3.75E-02
cellular macromolecule metabolic process	(GO:0044260)	1854	168	109.05	1.54	5.86E-06
cellular metabolic process	(GO:0044237)	2553	211	150.16	1.41	6.07E-05
cellular process	(GO:0009987)	4000	305	235.27	1.3	3.91E-05
Unclassified	(UNCLASSIFIED)	5766	321	339.15	-0.95	0.00E+00
system process	(GO:0003008)	226	1	13.29	< -0.2	2.49E-02
Cellular Component						
chromosome	(GO:0005694)	161	28	9.47	2.96	2.84E-04
chromosomal part	(GO:0044427)	143	24	8.41	2.85	3.16E-03
intracellular organelle lumen	(GO:0070013)	421	49	24.76	1.98	3.22E-03
membrane-enclosed lumen	(GO:0031974)	421	49	24.76	1.98	3.22E-03
organelle lumen	(GO:0043233)	421	49	24.76	1.98	3.22E-03
nuclear part	(GO:0044428)	520	60	30.59	1.96	3.92E-04
catalytic complex	(GO:1902494)	424	46	24.94	1.84	3.23E-02
protein complex	(GO:0043234)	862	88	50.7	1.74	1.97E-04
macromolecular complex	(GO:0032991)	1415	144	83.23	1.73	2.14E-08
nucleus	(GO:0005634)	1429	141	84.05	1.68	2.86E-07
organelle part	(GO:0044422)	1363	131	80.17	1.63	6.52E-06
intracellular organelle part	(GO:0044446)	1350	129	79.41	1.62	1.21E-05
intracellular membrane-bounded organelle	(GO:0043231)	2246	214	132.11	1.62	4.00E-11
membrane-bounded organelle	(GO:0043227)	2308	219	135.75	1.61	2.67E-11
cytoplasm	(GO:0005737)	1932	173	113.64	1.52	2.88E-06
organelle	(GO:0043226)	2655	235	156.16	1.5	3.23E-09
cytoplasmic part	(GO:0044444)	1281	113	75.35	1.5	4.40E-03
intracellular organelle	(GO:0043229)	2635	232	154.99	1.5	8.18E-09
intracellular part	(GO:0044424)	3357	294	197.45	1.49	2.55E-12
intracellular	(GO:0005622)	3616	313	212.69	1.47	7.10E-13
cell	(GO:0005623)	4166	326	245.04	1.33	1.19E-07
cell part	(GO:0044464)	4141	323	243.57	1.33	2.37E-07
cellular_component	(GO:0005575)	6143	416	361.32	1.15	1.11E-02
Unclassified	(UNCLASSIFIED)	6047	301	355.68	-0.85	0.00E+00
cell periphery	(GO:0071944)	616	15	36.23	-0.41	1.80E-02
plasma membrane	(GO:0005886)	592	13	34.82	-0.37	6.50E-03

B. *MxC*^{F1}

GO Term	GO #	# in REF	#	expected	Fold Enrichment	Pvalue
Molecular Function						
purine nucleoside binding	(GO:0001883)	139	24	8.91	2.69	1.49E-02
guanyl ribonucleotide binding	(GO:0032561)	139	24	8.91	2.69	1.49E-02
purine ribonucleoside binding	(GO:0032550)	139	24	8.91	2.69	1.49E-02
GTP binding	(GO:0005525)	139	24	8.91	2.69	1.49E-02
guanyl nucleotide binding	(GO:0019001)	139	24	8.91	2.69	1.49E-02
ribonucleoside binding	(GO:0032549)	142	24	9.1	2.64	2.08E-02
nucleoside binding	(GO:0001882)	143	24	9.16	2.62	2.32E-02
heterocyclic compound binding	(GO:1901363)	2368	197	151.72	1.3	3.56E-02
organic cyclic compound binding	(GO:0097159)	2376	197	152.23	1.29	4.34E-02
Unclassified	(UNCLASSIFIED)	5202	323	333.29	-0.97	0.00E+00
molecular transducer activity	(GO:0060089)	382	6	24.47	-0.25	6.01E-03
receptor activity	(GO:0004872)	382	6	24.47	-0.25	6.01E-03
transmembrane receptor activity	(GO:0099600)	330	5	21.14	-0.24	2.07E-02
signaling receptor activity	(GO:0038023)	335	5	21.46	-0.23	1.60E-02
Biological Process						
regulation of macromolecule metabolic process	(GO:0060255)	868	88	55.61	1.58	2.03E-02
regulation of primary metabolic process	(GO:0080090)	821	83	52.6	1.58	3.93E-02
regulation of metabolic process	(GO:0019222)	896	90	57.41	1.57	2.27E-02
cellular macromolecule metabolic process	(GO:0044260)	1854	168	118.78	1.41	2.07E-03
cellular metabolic process	(GO:0044237)	2553	217	163.57	1.33	3.86E-03
Unclassified	(UNCLASSIFIED)	5766	319	369.42	-0.86	0.00E+00
sensory perception	(GO:0007600)	211	1	13.52	< -0.2	2.03E-02
Cellular Component						
catalytic complex	(GO:1902494)	424	50	27.17	1.84	1.72E-02
macromolecular complex	(GO:0032991)	1415	150	90.66	1.65	2.37E-07
intracellular non-membrane-bounded organelle	(GO:0043232)	656	68	42.03	1.62	3.91E-02
non-membrane-bounded organelle	(GO:0043228)	656	68	42.03	1.62	3.91E-02
intracellular organelle part	(GO:0044446)	1350	126	86.49	1.46	5.53E-03
cytoplasmic part	(GO:0044444)	1281	119	82.07	1.45	1.23E-02
organelle part	(GO:0044422)	1363	126	87.33	1.44	8.59E-03
cytoplasm	(GO:0005737)	1932	174	123.78	1.41	7.11E-04
intracellular part	(GO:0044424)	3357	298	215.08	1.39	3.59E-08
intracellular	(GO:0005622)	3616	320	231.67	1.38	5.20E-09
intracellular organelle	(GO:0043229)	2635	227	168.82	1.34	2.82E-04
organelle	(GO:0043226)	2655	228	170.1	1.34	3.34E-04
intracellular membrane-bounded organelle	(GO:0043231)	2246	187	143.9	1.3	3.09E-02
membrane-bounded organelle	(GO:0043227)	2308	191	147.87	1.29	3.53E-02
cell	(GO:0005623)	4166	337	266.91	1.26	5.39E-05
cell part	(GO:0044464)	4141	333	265.31	1.26	1.31E-04
Unclassified	(UNCLASSIFIED)	6047	342	387.42	-0.88	0.00E+00

Table 3.7 GO over-representation test results for paternally expressed hybrid genes.

A. *CxM*^{F1}

GO Term	GO #	# in REF	#	expected	Fold Enrichment	Pvalue
Molecular Function						
Unclassified	(UNCLASSIFIED)	5202	286	309.82	-0.92	0.00E+00
Biological Process						
Unclassified	(UNCLASSIFIED)	5766	309	342.93	-0.9	0.00E+00
multicellular organismal process	(GO:0032501)	460	9	27.36	-0.33	4.08E-02
Cellular Component						
nucleoplasm	(GO:0005654)	168	25	9.99	2.5	1.76E-02
nuclear lumen	(GO:0031981)	335	47	19.92	2.36	4.52E-05
intracellular organelle lumen	(GO:0070013)	421	55	25.04	2.2	3.67E-05
membrane-enclosed lumen	(GO:0031974)	421	55	25.04	2.2	3.67E-05
organelle lumen	(GO:0043233)	421	55	25.04	2.2	3.67E-05
nuclear part	(GO:0044428)	520	55	30.93	1.78	1.72E-02
protein complex	(GO:0043234)	862	79	51.27	1.54	4.80E-02
macromolecular complex	(GO:0032991)	1415	128	84.16	1.52	4.97E-04
intracellular organelle part	(GO:0044446)	1350	115	80.29	1.43	2.63E-02
organelle part	(GO:0044422)	1363	116	81.06	1.43	2.52E-02
cytoplasm	(GO:0005737)	1932	154	114.91	1.34	3.47E-02
intracellular	(GO:0005622)	3616	287	215.06	1.33	3.36E-06
intracellular part	(GO:0044424)	3357	263	199.66	1.32	8.08E-05
cell	(GO:0005623)	4166	307	247.77	1.24	1.31E-03
cell part	(GO:0044464)	4141	302	246.29	1.23	4.40E-03
Unclassified	(UNCLASSIFIED)	6047	327	359.65	-0.91	0.00E+00

B. *MxC*^{F1}

GO Term	GO #	# in REF	#	expected	Fold Enrichment	Pvalue
Molecular Function						
structural constituent of ribosome	(GO:0003735)	143	26	10.82	2.4	4.59E-02
Unclassified	(UNCLASSIFIED)	5202	405	393.46	1.03	0.00E+00
signaling receptor activity	(GO:0038023)	335	7	25.34	-0.28	1.22E-02
transmembrane signaling receptor activity	(GO:0004888)	312	6	23.6	-0.25	1.24E-02
Biological Process						
peptide biosynthetic process	(GO:0043043)	247	44	18.68	2.36	3.47E-04
amide biosynthetic process	(GO:0043604)	266	47	20.12	2.34	1.73E-04
translation	(GO:0006412)	244	43	18.46	2.33	6.24E-04
peptide metabolic process	(GO:0006518)	291	45	22.01	2.04	9.78E-03
organonitrogen compound biosynthetic process	(GO:1901566)	425	64	32.15	1.99	3.12E-04
cellular amide metabolic process	(GO:0043603)	323	48	24.43	1.96	1.36E-02
cellular nitrogen compound biosynthetic process	(GO:0044271)	666	82	50.37	1.63	1.69E-02
macromolecule biosynthetic process	(GO:0009059)	675	81	51.05	1.59	4.54E-02
cellular component organization	(GO:0016043)	830	96	62.78	1.53	3.48E-02
gene expression	(GO:0010467)	843	97	63.76	1.52	3.84E-02
cellular protein metabolic process	(GO:0044267)	1085	124	82.06	1.51	3.82E-03
biosynthetic process	(GO:0009058)	1039	116	78.59	1.48	2.28E-02
cellular biosynthetic process	(GO:0044249)	989	110	74.8	1.47	4.45E-02
organic substance biosynthetic process	(GO:1901576)	998	111	75.48	1.47	4.09E-02
cellular nitrogen compound metabolic process	(GO:0034641)	1367	147	103.39	1.42	1.08E-02
cellular macromolecule metabolic process	(GO:0044260)	1854	192	140.23	1.37	3.75E-03
cellular metabolic process	(GO:0044237)	2553	247	193.1	1.28	1.54E-02
cellular process	(GO:0009987)	4000	361	302.54	1.19	3.49E-02
Unclassified	(UNCLASSIFIED)	5766	417	436.12	-0.96	0.00E+00
multicellular organismal process	(GO:0032501)	460	11	34.79	-0.32	2.11E-03
sensory perception of smell	(GO:0007608)	180	1	13.61	< -0.2	1.88E-02
sensory perception of chemical stimulus	(GO:0007606)	195	1	14.75	< -0.2	6.42E-03
Cellular Component						
ribosomal subunit	(GO:0044391)	124	23	9.38	2.45	4.85E-02
ribosome	(GO:0005840)	165	28	12.48	2.24	4.14E-02
membrane protein complex	(GO:0098796)	211	33	15.96	2.07	4.72E-02
intracellular organelle lumen	(GO:0070013)	421	55	31.84	1.73	3.98E-02
membrane-enclosed lumen	(GO:0031974)	421	55	31.84	1.73	3.98E-02
organelle lumen	(GO:0043233)	421	55	31.84	1.73	3.98E-02
nuclear part	(GO:0044428)	520	66	39.33	1.68	1.89E-02
protein complex	(GO:0043234)	862	109	65.2	1.67	6.43E-05
macromolecular complex	(GO:0032991)	1415	176	107.02	1.64	1.40E-08
cytoplasmic part	(GO:0044444)	1281	152	96.89	1.57	9.27E-06
organelle part	(GO:0044422)	1363	159	103.09	1.54	1.23E-05
intracellular organelle part	(GO:0044446)	1350	157	102.11	1.54	1.87E-05
cytoplasm	(GO:0005737)	1932	212	146.13	1.45	4.58E-06
intracellular part	(GO:0044424)	3357	341	253.91	1.34	1.24E-07
intracellular membrane-bounded organelle	(GO:0043231)	2246	227	169.88	1.34	7.76E-04
membrane-bounded organelle	(GO:0043227)	2308	232	174.57	1.33	8.47E-04
intracellular organelle	(GO:0043229)	2635	264	199.3	1.32	1.48E-04
organelle	(GO:0043226)	2655	266	200.81	1.32	1.30E-04
intracellular	(GO:0005622)	3616	354	273.5	1.29	3.88E-06
cell	(GO:0005623)	4166	377	315.1	1.2	5.68E-03
cell part	(GO:0044464)	4141	373	313.21	1.19	1.04E-02
Unclassified	(UNCLASSIFIED)	6047	401	457.37	-0.88	0.00E+00

Table 3.8 GO over-representation test results for non-additively over-expressed hybrid genes.

A. *CxM*^{F1}

GO Term	GO #	# in REF	#	expected	Fold Enrichment	Pvalue
Molecular Function						
structural constituent of ribosome	(GO:0003735)	143	54	27.31	1.98	3.00E-03
transferase activity	(GO:0016740)	1008	248	192.5	1.29	2.75E-02
Unclassified	(UNCLASSIFIED)	5202	935	993.46	-0.94	0.00E+00
peptidase activity, acting on L-amino acid peptides	(GO:0070011)	560	69	106.95	-0.65	3.43E-02
signal transducer activity	(GO:0004871)	401	38	76.58	-0.5	4.76E-04
molecular transducer activity	(GO:0060089)	382	24	72.95	-0.33	1.23E-08
receptor activity	(GO:0004872)	382	24	72.95	-0.33	1.23E-08
transmembrane receptor activity	(GO:0099600)	330	16	63.02	-0.25	8.51E-10
signaling receptor activity	(GO:0038023)	335	16	63.98	-0.25	4.06E-10
transmembrane signaling receptor activity	(GO:0004888)	312	13	59.58	-0.22	1.65E-10
chitin binding	(GO:0008061)	86	2	16.42	< -0.2	8.84E-03
odorant binding	(GO:0005549)	148	3	28.26	< -0.2	1.60E-06
structural constituent of cuticle	(GO:0042302)	154	2	29.41	< -0.2	5.47E-08
Biological Process						
vesicle-mediated transport	(GO:0016192)	202	72	38.58	1.87	8.90E-04
cellular lipid metabolic process	(GO:0044255)	193	64	36.86	1.74	3.10E-02
organonitrogen compound biosynthetic process	(GO:1901566)	425	121	81.16	1.49	1.75E-02
biosynthetic process	(GO:0009058)	1039	269	198.42	1.36	3.88E-04
organic substance biosynthetic process	(GO:1901576)	998	254	190.59	1.33	3.01E-03
cellular biosynthetic process	(GO:0044249)	989	251	188.88	1.33	4.23E-03
cellular protein metabolic process	(GO:0044267)	1085	275	207.21	1.33	1.52E-03
single-organism metabolic process	(GO:0044710)	1251	313	238.91	1.31	7.78E-04
cellular metabolic process	(GO:0044237)	2553	606	487.56	1.24	2.79E-06
single-organism cellular process	(GO:0044763)	1597	377	304.99	1.24	1.14E-02
cellular macromolecule metabolic process	(GO:0044260)	1854	425	354.07	1.2	4.14E-02
cellular process	(GO:0009987)	4000	880	763.9	1.15	2.60E-04
Unclassified	(UNCLASSIFIED)	5766	1009	1101.17	-0.92	0.00E+00
multicellular organismal process	(GO:0032501)	460	42	87.85	-0.48	2.93E-05
system process	(GO:0003008)	226	11	43.16	-0.25	5.32E-06
neurological system process	(GO:0050877)	219	9	41.82	-0.22	8.50E-07
sensory perception	(GO:0007600)	211	8	40.3	< -0.2	6.18E-07
chitin metabolic process	(GO:0006030)	109	4	20.82	< -0.2	9.43E-03
glucosamine-containing compound metabolic process	(GO:1901071)	111	4	21.2	< -0.2	6.88E-03
amino sugar metabolic process	(GO:0006040)	114	4	21.77	< -0.2	4.27E-03
detection of stimulus	(GO:0051606)	143	4	27.31	< -0.2	3.80E-05
sensory perception of chemical stimulus	(GO:0007606)	195	4	37.24	< -0.2	5.43E-09
sensory perception of smell	(GO:0007608)	180	3	34.38	< -0.2	8.09E-09
behavior	(GO:0007610)	67	1	12.8	< -0.2	4.27E-02
Cellular Component						
endoplasmic reticulum	(GO:0005783)	201	82	38.39	2.14	1.93E-07
nuclear outer membrane-endoplasmic reticulum membrane network	(GO:0042175)	131	50	25.02	2	2.74E-03
endoplasmic reticulum part	(GO:0044432)	135	51	25.78	1.98	2.93E-03
endoplasmic reticulum membrane	(GO:0005789)	130	49	24.83	1.97	4.64E-03
ribosome	(GO:0005840)	165	59	31.51	1.87	2.97E-03
endomembrane system	(GO:0012505)	447	156	85.37	1.83	6.73E-10
ribonucleoprotein complex	(GO:1990904)	367	110	70.09	1.57	1.97E-03
intracellular ribonucleoprotein complex	(GO:0030529)	367	110	70.09	1.57	1.97E-03
cytoplasmic part	(GO:0044444)	1281	372	244.64	1.52	1.94E-13
cytoplasm	(GO:0005737)	1932	519	368.97	1.41	1.24E-13
macromolecular complex	(GO:0032991)	1415	353	270.23	1.31	6.66E-05
intracellular organelle part	(GO:0044446)	1350	333	257.82	1.29	4.40E-04
organelle part	(GO:0044422)	1363	333	260.3	1.28	1.01E-03
intracellular part	(GO:0044424)	3357	786	641.11	1.23	1.16E-08
intracellular	(GO:0005622)	3616	843	690.57	1.22	2.89E-09
intracellular organelle	(GO:0043229)	2635	602	503.22	1.2	2.79E-04
organelle	(GO:0043226)	2655	604	507.04	1.19	4.58E-04
membrane-bounded organelle	(GO:0043227)	2308	513	440.77	1.16	4.30E-02
cell	(GO:0005623)	4166	905	795.61	1.14	5.40E-04
cell part	(GO:0044464)	4141	899	790.83	1.14	6.78E-04
cellular_component	(GO:0005575)	6143	1281	1173.17	1.09	1.87E-03
Unclassified	(UNCLASSIFIED)	6047	1047	1154.83	-0.91	0.00E+00

B. *MxC^{F1}*

GO Term	GO #	# in REF	#	expected	Fold Enrichment	Pvalue
Molecular Function						
NAD binding	(GO:0051287)	28	17	5.26	3.23	2.96E-02
nucleotide binding	(GO:0000166)	1092	290	205.14	1.41	1.96E-06
nucleoside phosphate binding	(GO:1901265)	1092	290	205.14	1.41	1.96E-06
purine nucleotide binding	(GO:0017076)	796	208	149.54	1.39	1.31E-03
purine ribonucleotide binding	(GO:0032555)	795	207	149.35	1.39	1.74E-03
ribonucleotide binding	(GO:0032553)	804	209	151.04	1.38	1.72E-03
purine ribonucleoside triphosphate binding	(GO:0035639)	793	206	148.97	1.38	2.14E-03
small molecule binding	(GO:0036094)	1155	297	216.98	1.37	2.54E-05
anion binding	(GO:0043168)	990	250	185.98	1.34	1.38E-03
nucleic acid binding	(GO:0003676)	1363	332	256.05	1.3	5.92E-04
heterocyclic compound binding	(GO:1901363)	2368	547	444.85	1.23	6.84E-05
organic cyclic compound binding	(GO:0097159)	2376	548	446.35	1.23	8.10E-05
Unclassified	(UNCLASSIFIED)	5202	927	977.24	-0.95	0.00E+00
signal transducer activity	(GO:0004871)	401	37	75.33	-0.49	4.43E-04
heme binding	(GO:0020037)	151	9	28.37	-0.32	1.71E-02
tetrapyrrole binding	(GO:0046906)	152	9	28.55	-0.32	1.50E-02
molecular transducer activity	(GO:0060089)	382	22	71.76	-0.31	2.77E-09
receptor activity	(GO:0004872)	382	22	71.76	-0.31	2.77E-09
transmembrane receptor activity	(GO:0099600)	330	18	61.99	-0.29	2.54E-08
transmembrane signaling receptor activity	(GO:0004888)	312	16	58.61	-0.27	2.44E-08
signaling receptor activity	(GO:0038023)	335	17	62.93	-0.27	3.49E-09
odorant binding	(GO:0005549)	148	5	27.8	< -0.2	1.04E-04
olfactory receptor activity	(GO:0004984)	124	2	23.29	< -0.2	1.69E-05
structural constituent of cuticle	(GO:0042302)	154	1	28.93	< -0.2	5.67E-09
Biological Process						
vesicle-mediated transport	(GO:0016192)	202	65	37.95	1.71	3.99E-02
RNA processing	(GO:0006396)	337	102	63.31	1.61	3.88E-03
nucleic acid metabolic process	(GO:0090304)	881	235	165.5	1.42	7.68E-05
RNA metabolic process	(GO:0016070)	706	186	132.63	1.4	4.13E-03
gene expression	(GO:0010467)	843	222	158.37	1.4	4.59E-04
cellular macromolecule metabolic process	(GO:0044260)	1854	471	348.29	1.35	5.17E-09
macromolecule modification	(GO:0043412)	837	212	157.24	1.35	1.05E-02
nucleobase-containing compound metabolic process	(GO:0006139)	1086	275	204.01	1.35	4.38E-04
cellular nitrogen compound metabolic process	(GO:0034641)	1367	345	256.8	1.34	1.52E-05
heterocycle metabolic process	(GO:0046483)	1131	284	212.47	1.34	5.50E-04

cellular protein metabolic process	(GO:0044267)	1085	270	203.83	1.32	2.18E-03
organic cyclic compound metabolic process	(GO:1901360)	1165	288	218.86	1.32	1.60E-03
cellular aromatic compound metabolic process	(GO:0006725)	1145	283	215.1	1.32	2.03E-03
cellular component organization or biogenesis	(GO:0071840)	950	232	178.47	1.3	4.00E-02
cellular metabolic process	(GO:0044237)	2553	622	479.6	1.3	9.54E-10
nitrogen compound metabolic process	(GO:0006807)	1574	375	295.69	1.27	1.16E-03
macromolecule metabolic process	(GO:0043170)	2488	548	467.39	1.17	2.63E-02
primary metabolic process	(GO:0044238)	2983	651	560.38	1.16	9.35E-03
organic substance metabolic process	(GO:0071704)	3183	684	597.95	1.14	3.42E-02
metabolic process	(GO:0008152)	3793	809	712.55	1.14	1.05E-02
cellular process	(GO:0009987)	4000	852	751.44	1.13	5.96E-03
Unclassified	(UNCLASSIFIED)	5766	1043	1083.19	-0.96	0.00E+00
single-organism developmental process	(GO:0044767)	278	26	52.22	-0.5	4.52E-02
anatomical structure development	(GO:0048856)	246	21	46.21	-0.45	2.69E-02
single-multicellular organism process	(GO:0044707)	236	20	44.33	-0.45	3.50E-02
response to chemical	(GO:0042221)	300	25	56.36	-0.44	2.07E-03
multicellular organismal process	(GO:0032501)	460	30	86.42	-0.35	1.18E-09
neurological system process	(GO:0050877)	219	11	41.14	-0.27	2.46E-05
system process	(GO:0003008)	226	11	42.46	-0.26	9.09E-06
sensory perception	(GO:0007600)	211	9	39.64	-0.23	4.86E-06
sensory perception of smell	(GO:0007608)	180	6	33.81	< -0.2	5.01E-06
sensory perception of chemical stimulus	(GO:0007606)	195	6	36.63	< -0.2	4.59E-07
detection of stimulus	(GO:0051606)	143	4	26.86	< -0.2	5.59E-05
detection of chemical stimulus involved in sensory perception of smell	(GO:0050911)	124	2	23.29	< -0.2	2.35E-05
detection of chemical stimulus	(GO:0009593)	132	2	24.8	< -0.2	5.82E-06
detection of chemical stimulus involved in sensory perception	(GO:0050907)	132	2	24.8	< -0.2	5.82E-06
detection of stimulus involved in sensory perception	(GO:0050906)	133	2	24.99	< -0.2	4.88E-06
Cellular Component						
intracellular organelle lumen	(GO:0070013)	421	132	79.09	1.67	8.09E-06
membrane-enclosed lumen	(GO:0031974)	421	132	79.09	1.67	8.09E-06
organelle lumen	(GO:0043233)	421	132	79.09	1.67	8.09E-06
ribonucleoprotein complex	(GO:1990904)	367	107	68.94	1.55	4.18E-03
intracellular ribonucleoprotein complex	(GO:0030529)	367	107	68.94	1.55	4.18E-03
nuclear lumen	(GO:0031981)	335	97	62.93	1.54	1.38E-02
nuclear part	(GO:0044428)	520	142	97.69	1.45	4.31E-03
endomembrane system	(GO:0012505)	447	122	83.97	1.45	1.84E-02
cytoplasmic part	(GO:0044444)	1281	329	240.65	1.37	2.42E-06

intracellular non-membrane-bounded organelle	(GO:0043232)	656	168	123.24	1.36	2.04E-02
non-membrane-bounded organelle	(GO:0043228)	656	168	123.24	1.36	2.04E-02
membrane-bounded organelle	(GO:0043227)	2308	573	433.58	1.32	1.91E-10
intracellular organelle part	(GO:0044446)	1350	335	253.61	1.32	5.09E-05
organelle part	(GO:0044422)	1363	335	256.05	1.31	1.24E-04
intracellular membrane-bounded organelle	(GO:0043231)	2246	552	421.93	1.31	3.35E-09
cytoplasm	(GO:0005737)	1932	473	362.94	1.3	3.22E-07
macromolecular complex	(GO:0032991)	1415	346	265.82	1.3	1.23E-04
nucleus	(GO:0005634)	1429	349	268.45	1.3	1.21E-04
organelle	(GO:0043226)	2655	645	498.77	1.29	1.84E-10
intracellular organelle	(GO:0043229)	2635	640	495.01	1.29	2.50E-10
intracellular part	(GO:0044424)	3357	811	630.64	1.29	5.15E-14
intracellular	(GO:0005622)	3616	872	679.3	1.28	1.89E-15
cell	(GO:0005623)	4166	919	782.62	1.17	7.33E-07
cell part	(GO:0044464)	4141	913	777.92	1.17	9.85E-07
Unclassified	(UNCLASSIFIED)	6047	1095	1135.98	-0.96	0.00E+00
membrane	(GO:0016020)	2905	460	545.73	-0.84	4.59E-03
membrane part	(GO:0044425)	2732	427	513.23	-0.83	2.58E-03
intrinsic component of membrane	(GO:0031224)	2572	393	483.17	-0.81	5.71E-04
integral component of membrane	(GO:0016021)	2563	390	481.48	-0.81	3.89E-04
cell periphery	(GO:0071944)	616	66	115.72	-0.57	8.79E-05
plasma membrane	(GO:0005886)	592	63	111.21	-0.57	1.19E-04

Table 3.9 GO over-representation test results for non-additively under-expressed hybrid genes.

A. *CxM*^{F1}

GO Term	GO #	# in REF	#	expected	Fold Enrichment	Pvalue
Molecular Function						
nucleoside-triphosphatase activity	(GO:0017111)	323	54	24.14	2.24	5.81E-05
hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	(GO:0016818)	339	55	25.33	2.17	1.13E-04
pyrophosphatase activity	(GO:0016462)	337	54	25.19	2.14	2.16E-04
hydrolase activity, acting on acid anhydrides	(GO:0016817)	344	55	25.71	2.14	1.79E-04
ATPase activity	(GO:0016887)	233	37	17.41	2.12	1.97E-02
ATP binding	(GO:0005524)	656	90	49.03	1.84	3.10E-05
adenyl ribonucleotide binding	(GO:0032559)	658	90	49.17	1.83	3.54E-05
adenyl nucleotide binding	(GO:0030554)	659	90	49.25	1.83	3.79E-05
purine ribonucleoside triphosphate binding	(GO:0035639)	793	106	59.26	1.79	6.47E-06
purine ribonucleotide binding	(GO:0032555)	795	106	59.41	1.78	7.36E-06
purine nucleotide binding	(GO:0017076)	796	106	59.49	1.78	7.85E-06
ribonucleotide binding	(GO:0032553)	804	106	60.09	1.76	1.30E-05
nucleotide binding	(GO:0000166)	1092	134	81.61	1.64	1.01E-05
nucleoside phosphate binding	(GO:1901265)	1092	134	81.61	1.64	1.01E-05
anion binding	(GO:0043168)	990	119	73.99	1.61	2.16E-04
small molecule binding	(GO:0036094)	1155	136	86.32	1.58	8.49E-05
carbohydrate derivative binding	(GO:0097367)	904	106	67.56	1.57	3.13E-03
nucleic acid binding	(GO:0003676)	1363	145	101.86	1.42	8.33E-03
heterocyclic compound binding	(GO:1901363)	2368	238	176.97	1.34	4.21E-04
organic cyclic compound binding	(GO:0097159)	2376	238	177.57	1.34	5.47E-04
Unclassified	(UNCLASSIFIED)	5202	388	388.76	-1	0.00E+00
serine-type endopeptidase activity	(GO:0004252)	306	6	22.87	-0.26	2.17E-02
signal transducer activity	(GO:0004871)	401	4	29.97	< -0.2	2.10E-06
molecular transducer activity	(GO:0060089)	382	3	28.55	< -0.2	9.85E-07
receptor activity	(GO:0004872)	382	3	28.55	< -0.2	9.85E-07
transmembrane receptor activity	(GO:0099600)	330	2	24.66	< -0.2	3.97E-06
transmembrane signaling receptor activity	(GO:0004888)	312	1	23.32	< -0.2	1.13E-06
signaling receptor activity	(GO:0038023)	335	1	25.04	< -0.2	2.08E-07
Biological Process						
protein modification by small protein removal	(GO:0070646)	36	12	2.69	4.46	2.82E-02
DNA-dependent DNA replication	(GO:0006261)	44	13	3.29	3.95	4.49E-02
DNA conformation change	(GO:0071103)	55	15	4.11	3.65	2.93E-02
DNA replication	(GO:0006260)	70	19	5.23	3.63	2.78E-03

microtubule cytoskeleton organization	(GO:0000226)	63	16	4.71	3.4	3.72E-02
DNA metabolic process	(GO:0006259)	190	48	14.2	3.38	9.58E-10
cell cycle process	(GO:0022402)	115	29	8.59	3.37	3.32E-05
DNA repair	(GO:0006281)	118	29	8.82	3.29	5.68E-05
microtubule-based process	(GO:0007017)	128	30	9.57	3.14	9.21E-05
cellular response to DNA damage stimulus	(GO:0006974)	132	30	9.86	3.04	1.75E-04
modification-dependent macromolecule catabolic process	(GO:0043632)	147	33	10.99	3	5.63E-05
ubiquitin-dependent protein catabolic process	(GO:0006511)	144	32	10.76	2.97	1.12E-04
modification-dependent protein catabolic process	(GO:0019941)	144	32	10.76	2.97	1.12E-04
cell cycle	(GO:0007049)	135	30	10.09	2.97	2.79E-04
proteolysis involved in cellular protein catabolic process	(GO:0051603)	167	36	12.48	2.88	3.79E-05
mitotic cell cycle	(GO:0000278)	93	20	6.95	2.88	4.24E-02
cellular protein catabolic process	(GO:0044257)	168	36	12.56	2.87	4.38E-05
protein catabolic process	(GO:0030163)	174	36	13	2.77	1.02E-04
chromosome organization	(GO:0051276)	199	40	14.87	2.69	4.23E-05
ncRNA processing	(GO:0034470)	182	35	13.6	2.57	8.09E-04
cellular macromolecule catabolic process	(GO:0044265)	236	45	17.64	2.55	2.74E-05
protein modification by small protein conjugation or removal	(GO:0070647)	190	34	14.2	2.39	5.35E-03
cellular macromolecular complex assembly	(GO:0034622)	200	34	14.95	2.27	1.52E-02
ncRNA metabolic process	(GO:0034660)	235	39	17.56	2.22	6.27E-03
cellular response to stress	(GO:0033554)	199	33	14.87	2.22	3.28E-02
macromolecule catabolic process	(GO:0009057)	281	46	21	2.19	1.32E-03
organelle organization	(GO:0006996)	607	98	45.36	2.16	1.99E-09
RNA processing	(GO:0006396)	337	54	25.19	2.14	3.01E-04
nucleic acid metabolic process	(GO:0090304)	881	140	65.84	2.13	5.57E-14
nucleobase-containing compound metabolic process	(GO:0006139)	1086	161	81.16	1.98	8.23E-14
heterocycle metabolic process	(GO:0046483)	1131	166	84.52	1.96	6.00E-14
cellular aromatic compound metabolic process	(GO:0006725)	1145	166	85.57	1.94	1.82E-13
organic cyclic compound metabolic process	(GO:1901360)	1165	168	87.06	1.93	1.85E-13
RNA metabolic process	(GO:0016070)	706	99	52.76	1.88	2.86E-06
cellular component biogenesis	(GO:0044085)	450	63	33.63	1.87	2.77E-03
cellular component organization	(GO:0016043)	830	115	62.03	1.85	2.50E-07
cellular component organization or biogenesis	(GO:0071840)	950	131	71	1.85	1.54E-08
cellular macromolecule metabolic process	(GO:0044260)	1854	251	138.56	1.81	1.34E-18
cellular nitrogen compound metabolic process	(GO:0034641)	1367	184	102.16	1.8	3.43E-12
macromolecule modification	(GO:0043412)	837	107	62.55	1.71	7.18E-05
cellular protein modification process	(GO:0006464)	750	91	56.05	1.62	6.15E-03
protein modification process	(GO:0036211)	750	91	56.05	1.62	6.15E-03
cellular metabolic process	(GO:0044237)	2553	305	190.79	1.6	1.35E-15
cellular protein metabolic process	(GO:0044267)	1085	129	81.09	1.59	1.59E-04

nitrogen compound metabolic process	(GO:0006807)	1574	186	117.63	1.58	2.02E-07
gene expression	(GO:0010467)	843	97	63	1.54	2.42E-02
macromolecule metabolic process	(GO:0043170)	2488	263	185.94	1.41	9.30E-07
cellular process	(GO:0009987)	4000	413	298.93	1.38	3.14E-12
primary metabolic process	(GO:0044238)	2983	303	222.93	1.36	1.74E-06
organic substance metabolic process	(GO:0071704)	3183	316	237.88	1.33	7.49E-06
metabolic process	(GO:0008152)	3793	341	283.46	1.2	3.36E-02
Unclassified	(UNCLASSIFIED)	5766	390	430.91	-0.91	0.00E+00
signal transduction	(GO:0007165)	851	35	63.6	-0.55	4.61E-02
response to chemical	(GO:0042221)	300	5	22.42	-0.22	1.06E-02
multicellular organismal process	(GO:0032501)	460	7	34.38	-0.2	1.24E-05
sensory perception	(GO:0007600)	211	1	15.77	< -0.2	2.42E-03
neurological system process	(GO:0050877)	219	1	16.37	< -0.2	1.37E-03
system process	(GO:0003008)	226	1	16.89	< -0.2	8.27E-04
Cellular Component						
outer membrane	(GO:0019867)	16	8	1.2	6.69	1.56E-02
peptidase complex	(GO:1905368)	46	15	3.44	4.36	1.43E-03
microtubule	(GO:0005874)	47	13	3.51	3.7	3.33E-02
mitochondrial membrane part	(GO:0044455)	91	22	6.8	3.23	1.13E-03
microtubule cytoskeleton	(GO:0015630)	129	31	9.64	3.22	1.25E-05
mitochondrial protein complex	(GO:0098798)	71	17	5.31	3.2	1.69E-02
cytoskeletal part	(GO:0044430)	173	35	12.93	2.71	9.90E-05
catalytic complex	(GO:1902494)	424	80	31.69	2.52	5.39E-11
cytoskeleton	(GO:0005856)	213	39	15.92	2.45	2.43E-04
mitochondrial envelope	(GO:0005740)	179	32	13.38	2.39	3.96E-03
chromosome	(GO:0005694)	161	28	12.03	2.33	2.25E-02
mitochondrial membrane	(GO:0031966)	169	29	12.63	2.3	2.10E-02
nuclear lumen	(GO:0031981)	335	56	25.04	2.24	1.79E-05
intracellular organelle lumen	(GO:0070013)	421	67	31.46	2.13	5.30E-06
membrane-enclosed lumen	(GO:0031974)	421	67	31.46	2.13	5.30E-06
organelle lumen	(GO:0043233)	421	67	31.46	2.13	5.30E-06
transferase complex	(GO:1990234)	220	35	16.44	2.13	1.68E-02
mitochondrial part	(GO:0044429)	247	39	18.46	2.11	7.28E-03
protein complex	(GO:0043234)	862	136	64.42	2.11	1.14E-13
nuclear part	(GO:0044428)	520	82	38.86	2.11	1.71E-07
envelope	(GO:0031975)	222	35	16.59	2.11	2.00E-02
organelle envelope	(GO:0031967)	222	35	16.59	2.11	2.00E-02
intracellular non-membrane-bounded organelle	(GO:0043232)	656	96	49.03	1.96	2.37E-07
non-membrane-bounded organelle	(GO:0043228)	656	96	49.03	1.96	2.37E-07
intracellular organelle part	(GO:0044446)	1350	188	100.89	1.86	2.19E-14
organelle part	(GO:0044422)	1363	189	101.86	1.86	2.70E-14
mitochondrion	(GO:0005739)	394	54	29.44	1.83	9.90E-03
macromolecular complex	(GO:0032991)	1415	189	105.75	1.79	1.03E-12
nucleus	(GO:0005634)	1429	184	106.79	1.72	7.47E-11

intracellular organelle	(GO:0043229)	2635	315	196.92	1.6	8.46E-17
organelle	(GO:0043226)	2655	317	198.42	1.6	7.21E-17
intracellular membrane-bounded organelle	(GO:0043231)	2246	266	167.85	1.58	8.68E-13
membrane-bounded organelle	(GO:0043227)	2308	271	172.48	1.57	1.18E-12
intracellular part	(GO:0044424)	3357	377	250.88	1.5	7.88E-17
intracellular	(GO:0005622)	3616	398	270.24	1.47	1.08E-16
cytoplasm	(GO:0005737)	1932	194	144.38	1.34	4.01E-03
cell part	(GO:0044464)	4141	412	309.47	1.33	6.06E-10
cell	(GO:0005623)	4166	414	311.34	1.33	6.00E-10
Unclassified	(UNCLASSIFIED)	6047	411	451.91	-0.91	0.00E+00
membrane	(GO:0016020)	2905	162	217.1	-0.75	2.64E-03
membrane part	(GO:0044425)	2732	145	204.17	-0.71	2.89E-04
integral component of membrane	(GO:0016021)	2563	121	191.54	-0.63	4.43E-07
intrinsic component of membrane	(GO:0031224)	2572	121	192.21	-0.63	3.22E-07
cell periphery	(GO:0071944)	616	23	46.04	-0.5	4.36E-02
plasma membrane	(GO:0005886)	592	21	44.24	-0.47	2.59E-02
extracellular region	(GO:0005576)	356	5	26.61	< -0.2	1.29E-04

B. *MxC^{F1}*

GO Term	GO #	# in REF	#	expected	Fold Enrichment	Pvalue
Molecular Function						
Unclassified	(UNCLASSIFIED)	5202	154	139.55	1.1	0.00E+00
Biological Process						
cellular protein metabolic process	(GO:0044267)	1085	59	29.11	2.03	1.85E-04
cellular macromolecule metabolic process	(GO:0044260)	1854	83	49.73	1.67	1.45E-03
cellular metabolic process	(GO:0044237)	2553	100	68.48	1.46	3.15E-02
Unclassified	(UNCLASSIFIED)	5766	145	154.67	-0.94	0.00E+00
Cellular Component						
peptidase complex	(GO:1905368)	46	8	1.23	6.48	1.87E-02
respiratory chain	(GO:0070469)	53	8	1.42	5.63	4.94E-02
mitochondrial membrane part	(GO:0044455)	91	11	2.44	4.51	2.00E-02
mitochondrion	(GO:0005739)	394	30	10.57	2.84	1.88E-04
catalytic complex	(GO:1902494)	424	30	11.37	2.64	8.28E-04
cytoplasmic part	(GO:0044444)	1281	61	34.36	1.78	3.29E-03
macromolecular complex	(GO:0032991)	1415	63	37.96	1.66	1.78E-02
intracellular organelle part	(GO:0044446)	1350	60	36.21	1.66	2.97E-02
organelle part	(GO:0044422)	1363	60	36.56	1.64	3.87E-02
cytoplasm	(GO:0005737)	1932	79	51.83	1.52	2.90E-02
intracellular organelle	(GO:0043229)	2635	107	70.68	1.51	9.97E-04
organelle	(GO:0043226)	2655	107	71.22	1.5	1.43E-03
intracellular membrane-bounded organelle	(GO:0043231)	2246	89	60.25	1.48	2.70E-02
membrane-bounded organelle	(GO:0043227)	2308	90	61.91	1.45	4.36E-02
intracellular part	(GO:0044424)	3357	127	90.05	1.41	2.79E-03
intracellular	(GO:0005622)	3616	135	97	1.39	2.27E-03
Unclassified	(UNCLASSIFIED)	6047	141	162.21	-0.87	0.00E+00

Conclusion

Summary

Once only studied through the manipulation of *Drosophila,* speciation and HI have recently seen an influx of research using non-model organisms due to increasing accessibility of solved genome sequences and a diminishing cost of sequencing. With sequencing technologies we can get closer to the mechanisms behind HI and the rules of speciation and find how conserved these rules and mechanisms are. Even the traditional crossing experiments can be seen in a new light as our knowledge has expanded since they were first conceived. This dissertation uses both traditional and current research techniques to discover if mosquitoes of the *An. gambiae* complex follow the known patterns of HI.

Chapter I showed us that, while our current colonies do not show similar signs of sex-ratio distortion in the F1 generation compared Davidson's original crosses, we did see male hybrid sterility in all crosses except for the hybrids between the two most recently speciated members of the complex (*An. gambiae* and *An. coluzzii*). The patterns of sex-ratio distortion in our current crosses look like "Darwin's Corollary", or isolation asymmetry, a common pattern considered one of the "rules of speciation", and a rejection of Haldane's rule in multiple crosses that showed female inviability, though only one with major female inviability (Turelli and Moyle, 2006). We can hypothesize that one of the following contribute to both the asymmetric nature of HI and the female sterility: maternal effects (Sawamura, 1996; Turelli and Orr, 2000), chromosome imprinting (Kelsey

and Feil, 2013; Wolf et al., 2014), or simple chromosomal incompatibilities (Bateson, 1909; Dobzhansky, 1937; Muller, 1942). For the MAF x KGB cross that showed almost complete female inviability, we know that it is an incompatibility that acts before the larval stage and so is related to either fertilization or embryonic growth.

In chapter II we used QTL analysis along with high-throughput genomics and found that the X chromosome has a large effect on sterility of *An. coluzzii*♀ by *An. merus*♂ hybrids, but the X plays an equal or lesser role in sterility for the reciprocal hybrid cross. For the cross that adheres to the large-X effect we hypothesize that dominant autosomal factors from *An.merus* influence the expression of the *An. coluzzii* X-linked genes during spermatogenesis. For the reciprocal cross we hypothesize that epistatic interactions between the autosomes plays a larger role than interactions with the X chromosome.

Chapter III looked at the transcriptomics of reciprocal *An. coluzzii* by *An. merus* hybrids. We found that a disruption of sex chromosome inactivation during spermatogenesis is likely the cause of sterility in *An. coluzzii*♀ by *An. merus*♂ hybrid males, and this disruption is possibly a common cause of Haldane's rule in terms of sterility (Campbell et al., 2013; Good et al., 2010; Lifschytz and Lindsley, 1972). While the lack of sex chromosome inactivation causes massive misexpression of the X chromosome, we also see misexpression of the autosomes of both hybrids which may also be linked to sterility.

Taken together this dissertation has explored many different aspects of HI. The first chapter found that while hybrid male sterility is common in all except the most recently diverged mosquito species, major asymmetric inviability is also found in *An. arabiensis* by *An. merus* reciprocal crosses and one side of this same cross also shows a rejection of Haldane's rule as the female hybrid has the most severe phenotype. Chapter II found that hybrid sterility between *An. coluzzii*♀ and *An. merus*♂ are largely influenced by X-linked genes while the reciprocal cross's sterility is more likely due to interactions between the autosomes. The final chapter shows that in *An. coluzzii*♀ and *An. merus*♂ hybrid male testes almost all X chromosome genes are over-expressed when compared to either parentals' testes gene expression, agreeing with the findings of Chapter II. In the reciprocal cross (*An. merus*♀ by *An. coluzzii*♂) we also see an agreement with Chapter II in that misexpression was more common on the autosomes than the X-chromosome. These findings have added new knowledge into how post-zygotic reproductive isolation is evolving in Anopheline mosquitoes.

Future Directions

Each chapter of this dissertation has raised more research questions that can be further explored. The most interesting findings of Chapter I are the rejection of Haldane's rule, asymmetric isolation, and the seeming evolution of

less extreme phenotype most of the colony hybrid sex-ratios. Each finding requires more experimentation to uncover the mechanisms at work. Maternal effects, either through cytoplasmic incompatibilities or piwiRNA mechanisms could be responsible for the rejection of Haldane's rule, while differences in X chromosome – autosome interactions seems likely to be causing the asymmetric isolation. The colony evolution can be explored through hybrid sex-ratio comparisons between current colony populations and populations of wild mosquitoes. Chapters II and III, taken together, can narrow down candidate sterility genes. The list of genes found in Chapter II's QTLs can be trimmed by making sure that they are expressed in the hybrid testes. Candidate genes are also likely to be misexpressed in the hybrid. This shortened list of candidate genes can then be searched for signatures of faster evolution as would be expected under the faster-male and faster-X theories (Coyne and Orr, 1997). The candidates of this small list can then be tested through genetic manipulation of the mosquito embryo with either RNAi to disrupt the genes' expression or direct genetic editing (Iordanou et al., 2011). RNA-seq analysis of other *An. gambiae* complex hybrids can also show if disruption of sex chromosome inactivation is a common mechanism of hybrid sterility in mosquitoes.

References

Bateson, W. (1909). Heredity and variation in modern lights. In Darwin and Modern Science, (Cambridge: Cambridge University Press), pp. 85–101.

Campbell, P., Good, J.M., and Nachman, M.W. (2013). Meiotic sex chromosome inactivation is disrupted in sterile hybrid male house mice. Genetics *193*, 819–828.

Coyne, J.A., and Orr, H.A. (1997). “Patterns of Speciation in Drosophila” Revisited. Evolution *51*, 295.

Dobzhansky, T. (1937). Genetics and the Origin of Species (New York: Columbia University Press).

Good, J.M., Giger, T., Dean, M.D., and Nachman, M.W. (2010). Widespread over-expression of the X chromosome in sterile F1 hybrid mice. PLoS Genet. *6*, e1001148.

Iordanou, E., Chandran, R.R., Blackstone, N., and Jiang, L. (2011). RNAi Interference by dsRNA Injection into Drosophila Embryos. J. Vis. Exp.

Kelsey, G., and Feil, R. (2013). New insights into establishment and maintenance of DNA methylation imprints in mammals. Philos. Trans. R. Soc. B Biol. Sci. *368*, 20110336.

Lifschytz, E., and Lindsley, D.L. (1972). The Role of X-Chromosome Inactivation during Spermatogenesis. Proc. Natl. Acad. Sci. *69*, 182–186.

Muller, H. (1942). Isolating mechanisms, evolution and temperature. Biol Symp 71–125.

Sawamura, K. (1996). Maternal effect as a cause of exceptions for Haldane’s rule. Genetics *143*, 609.

Turelli, M., and Moyle, L.C. (2006). Asymmetric Postmating Isolation: Darwin’s Corollary to Haldane’s Rule. Genetics *176*, 1059–1088.

Turelli, M., and Orr, H.A. (2000). Dominance, Epistasis, and the Genetics of Postzygotic Isolation. Genetics *154*, 1663–1679.

Wolf, J.B., Oakey, R.J., and Feil, R. (2014). Imprinted gene expression in hybrids: perturbed mechanisms and evolutionary implications. Heredity *113*, 167–175.

Bibliography

Allen, W.R. (1969). Factors influencing pregnant mare serum gonadotrophin production. Nature *223*, 64–65.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T.T., Mast, J., Sunayama-Morita, T., and Stern, D.L. (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res. *21*, 610–617.

Arends, D., Prins, P., Jansen, R.C., and Broman, K.W. (2010). R/qtl: high-throughput multiple QTL mapping. Bioinformatics *26*, 2990–2992.

Assogba, B.S., Djogbenou, L., Saizonou, J., Diabate, A., Dabire, A., Moiroux, N., Gilles, J.R., Makoutode, M., and Baldet, T. (2014). Characterization of swarming and mating behavior between Anopheles coluzzii and Anopheles melas in a sympatry are of Benin. Acta Trop. *132*, S53–S63.

Atlan, A., Capillon, C., Derome, N., Couvet, D., and Montchamp-Moreau, C. (2003). The evolution of autosomal suppressors of sex-ratio drive in Drosophila simulans. Genetica *117*, 47–58.

Ávila, V., Marion de Procé, S., Campos, J.L., Borthwick, H., Charlesworth, B., and Betancourt, A.J. (2014). Faster-X Effects in Two Drosophila Lineages. Genome Biol. Evol. *6*, 2968–2982.

Baker, D.A., Nolan, T., Fischer, B., Pinder, A., Crisanti, A., and Russell, S. (2011). A comprehensive gene expression atlas of sex-and tissue-specificity in the malaria vector, Anopheles gambiae. BMC Genomics *12*, 296.

Barbash, D.A., Siino, D.F., Tarone, A.M., and Roote, J. (2003). A rapidly evolving MYB-related protein causes species isolation in Drosophila. Proc. Natl. Acad. Sci. *100*, 5302–5307.

Barbash, D.A., Awadalla, P., and Tarone, A.M. (2004). Functional Divergence Caused by Ancient Positive Selection of a Drosophila Hybrid Incompatibility Locus. PLoS Biol. *2*, e142.

Bateson, W. (1909). Heredity and variation in modern lights. In Darwin and Modern Science, (Cambridge: Cambridge University Press), pp. 85–101.

Benedict, M., Knols, B., Bossin, H., Howell, P., Mialhe, E., Caceres, C., and Robinson, A. (2009). Colonisation and mass rearing: learning from others. Malar. J. *8*, S2–S4.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. 289–300.

Bi, Y.-M., Meyer, A., Downs, G.S., Shi, X., El-Kereamy, A., Lukens, L., and Rothstein, S.J. (2014). High throughput RNA sequencing of a hybrid maize and its parents shows different mechanisms responsive to nitrogen limitation. BMC Genomics *15*, 77.

Biryukova, I., and Ye, T. (2015). Endogenous siRNAs and piRNAs derived from transposable elements and genes in the malaria vector mosquito Anopheles gambiae. BMC Genomics *16*.

Bock, I.R. (1984). Interspecific hybridization in the genus Drosophila. Evol. Biol. 41–70.

Bolnick, D.I., and Near, T.J. (2005). Tempo of hybrid inviability in centrarchid fishes (Teleostei: Centrarchidae). Evolution *59*, 1754–1767.

Bolnick, D.I., Turelli, M., Lopez-Fernandez, H., Wainwright, P.C., and Near, T.J. (2008). Accelerated Mitochondrial Evolution and “Darwin’s Corollary”: Asymmetric Viability of Reciprocal F1 Hybrids in Centrarchid Fishes. Genetics *178*, 1037–1048.

Bongiorni, S., and Prantera, G. (2003). Imprinted facultative heterochromatization in mealybugs. Genetica *117*, 271–279.

Bregliano, J., Picard, G., Bucheton, A., Pelisson, A., Lavige, J., and L’Heritier, P. (1980). Hybrid dysgenesis in Drosophila melanogaster. Science *207*, 606–611.

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.

Broman, K.W., Wu, H., Sen, S., and Churchill, G.A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics *19*, 889–890.

Brothers, A.N., and Delph, L.F. (2010). Haldane’s rule is extended to plants with sex chromosomes. Evolution *64*, 3643–3648.

Bryan, J.H. (1979). Observations on the member species of the Anopheles gambiae complex in The Gambia, West Africa. Trans. R. Soc. Trop. Med. Hyg. *73*, 463–466.

Cabrero, J., Teruel, M., Carmona, F., Jimenez, R., and Camacho, J. (2007). Histone H3 lysine 9 acetylation pattern suggests that X and B chromosomes are silenced during entire male meiosis in a grasshopper. Cytogenet. Genome Res. *119*, 135–142.

Campbell, P., Good, J.M., and Nachman, M.W. (2013). Meiotic sex chromosome inactivation is disrupted in sterile hybrid male house mice. Genetics *193*, 819–828.

de la Casa-Esperon, E., and Sapienza, C. (2003). Natural selection and the evolution of genome imprinting. Annu. Rev. Genet. *37*, 349–370.

Castellano, L., Rizzi, E., Krell, J., Di Cristina, M., Galizi, R., Mori, A., Tam, J., De Bellis, G., Stebbing, J., Crisanti, A., et al. (2015). The germline of the malaria mosquito

produces abundant miRNAs, endo-siRNAs, piRNAs and 29-nt small RNAs. BMC Genomics *16*, 100.

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

Cocquet, J., Ellis, P.J.I., Yamauchi, Y., Mahadevaiah, S.K., Affara, N.A., Ward, M.A., and Burgoyne, P.S. (2009). The Multicopy Gene Sly Represses the Sex Chromosomes in the Male Mouse Germline after Meiosis. PLoS Biol. *7*, e1000244.

Coetzee, M., Hunt, R.H., Wilkerson, R., Della Torre, A., Coulibaly, M.B., and Besansky, N.J. (2013). Anopheles coluzzii and Anopheles amharicus, new members of the Anopheles gambiae complex. Zootaxa *3619*, 246–274.

Coluzzi, M., Sabatini, A., Petrarca, V., and Di Deco, M.A. (1979). Chromosomal differentiation and adaptation to human environments in the Anopheles gambiae complex. Trans. R. Soc. Trop. Med. Hyg. *73*, 483–497.

Coluzzi, M., Sabatini, A., Della Torre, A., Di Deco, M.A., and Petrarca, V. (2002). A polytene chromosome analysis of the Anopheles gambiae species complex. Science *298*, 1415–1418.

Coluzzii, M., Sabatini, A., Petrarca, V., and Di Deco, M.A. (1979). Chromosomal differentiation and adaptation to human environments in the Anopheles gambiae complex. Trans. R. Soc. Trop. Med. Hyg. *73*, 483–497.

Coste, B., Xiao, B., Santos, J.S., Syeda, R., Grandl, J., Spencer, K.S., Kim, S.E., Schmidt, M., Mathur, J., Dubin, A.E., et al. (2012). Piezo proteins are pore-forming subunits of mechanically activated channels. Nature *483*, 176–181.

Counterman, B.A., Ortiz-Barrientos, D., and Noor, M.A. (2004). Using comparative genomic data to test for fast-X evolution. Evolution *58*, 656–660.

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 Charlesworth, B. (1989). Genetic analysis of X-linked sterility in hybrids between three sibling species of Drosophila. Heredity *62*, 97–106.

Coyne, J.A., and Orr, H.A. (1989). Two rules of speciation. Speciat. Its Consequences 180–207.

Coyne, J.A., and Orr, H.A. (1997). "Patterns of Speciation in Drosophila" Revisited. Evolution *51*, 295.

Coyne, J.A., and Orr, H.A. (2004). Speciation (Sinauer Associates, Inc.).

Coyne, J.A., Charlesworth, B., and Orr, H.A. (1991). Haldane's rule revisited. Evolution 1710–1714.

Craft, W.A. (1938). The sex ratio in mules and other hybrid mammals. Q. Rev. Biol. *13*, 19–40.

Crouse, H.V. (1960). The controlling element in sex chromosome behavior in Sciara. Genetics *45*, 1429–1443.

Cuamba, N., and Mendis, C. (2009). The role of Anopheles merus in malaria transmission in an area of southern Mozambique. J. Vector Borne Dis. *46*, 157–159.

Curtis, C.F. (1982). the mechanism of hybrid male sterility from crosses in the Anopheles gambiae and Glossina morsitans complexes. In Recent Developments of Insect Disease Vectors, (New York: Stipes Publishing), pp. 290–312.

Darwin, C. (1859). On the Origin of Species by Means of Natural Selection, Or, the Preservation of Favoured Races in the Struggle for Life (London: J. Murray).

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., and Blaxter, M.L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. *12*, 499–510.

Davidson, G. (1964a). Anopheles gambiae, a complex of species. Bull. World Health Organ. *31*, 625.

Davidson, G. (1964b). The Five Mating-Types in the Anopheles Gambiae Complex. Riv Malariol *43*, 167–183.

Davies, B., Hatton, E., Altemose, N., Hussin, J.G., Pratto, F., Zhang, G., Hinch, A.G., Moralli, D., Biggs, D., Diaz, R., et al. (2016). Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. Nature *530*, 171–176.

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. Mol. Biol. Evol. *32*, 2534–2546.

Dhana, K. (2016). Identify, describe, plot, and remove the outliers from the dataset.

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

Dobzhansky, T. (1937). Genetics and the Origin of Species (New York: Columbia University Press).

Dobzhansky, T. (1970). Genetics of the evolutionary process (Columbia University Press).

Donnelly, M.J., Pinto, J., Girod, R., Besansky, N.J., and Lehmann, T. (2004). Revisiting the role of introgression vs. shared ancestral polymorphisms as key processes shaping genetic diversity in the recently separated sibling species of the Anopheles gambiae complex. Heredity *92*, 61–68.

Duselis, A.R., and Vrana, P.B. (2010). Aberrant growth and pattern formation in Peromyscus hybrid placental development. Biol Reprod *83*, 988–996.

Ellegren, H. (2011). Emergence of male-biased genes on the chicken Z-chromosome: sex-chromosome contrasts between male and female heterogametic systems. Genome Res. *21*, 2082–2086.

Elliott, R.J., Aggoun, L., and Moore, J.B. (1994). Hidden Markov Models (Springer).

Ellis, P.J.I., Clemente, E.J., Ball, P., Toure, A., Ferguson, L., Turner, J.M.A., Loveland, K.L., Affara, N.A., and Burgoyne, P.S. (2005). Deletions on mouse Yq lead to upregulation of multiple X- and Y-linked transcripts in spermatids. Hum. Mol. Genet. *14*, 2705–2715.

Ferree, P.M., and Barbash, D.A. (2007). Distorted Sex Ratios: A Window into RNAi-Mediated Silencing. PLoS Biol. *5*, e303.

Fitch, K.R., Yasuda, G.K., Owens, K.N., and Wakimoto, B.T. (1998). Paternal effects in Drosophila: implications for mechanisms of early development. Curr Top. Dev Biol *38*, 1–34.

Fontaine, M.C., Pease, J.B., Steele, A., Waterhouse, R.M., Neafsey, D.E., Sharakhov, I.V., Jiang, X., Hall, A.B., Catteruccia, F., Kakani, E., et al. (2015). Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science *347*, 1258524–1258524.

Fontenot, B.E. (2009). Natural Hybridization And Speciation In Toads Of The Anaxyrus americanus Group.

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

George, P., Jensen, S., Pogorelcnik, R., Lee, J., Xing, Y., Brasset, E., Vaury, C., and Sharakhov, I.V. (2015). Increased production of piRNAs from euchromatic clusters and genes in Anopheles gambiae compared with Drosophila melanogaster. Epigenetics Chromatin *8*.

Goday, C., and Esteban, M.R. (2001). Chromosome elimination in sciarid flies. Bioessays *23*, 242–250.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. *11*, R86.

Goldschmidt, R. (1940). The material basis of evolution (New Haven, Conn.: Yale University Press).

Gomes, S., and Civetta, A. (2014). Misregulation of spermatogenesis genes in Drosophila hybrids is lineage-specific and driven by the combined effects of sterility and fast male regulatory divergence. J. Evol. Biol.

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.

Good, J.M., Giger, T., Dean, M.D., and Nachman, M.W. (2010). Widespread over-expression of the X chromosome in sterile F1 hybrid mice. PLoS Genet. *6*, e1001148.

Govere, J., Durrheim, D., Coetzee, M., Hunt, R.H., and La Grange, J. (2000). Captures of mosquitoes of the Anopheles gambiae complex (Diptera: Culicidae) in the Lowveld region of Mpumalanga Province, South Africa. Afr. Entomol. *8*, 91–99.

Gray, A.P. (1958). Bird hybrids. A check-list with bibliography.

Gray, A.P. (1972). Mammalian hybrids. A check-list with bibliography. In Technical Communications, (Commonwealth Bureau of Animal Breeding and Genetics), p.

Guioli, S., Lovell-Badge, R., and Turner, J.M.A. (2012). Error-Prone ZW Pairing and No Evidence for Meiotic Sex Chromosome Inactivation in the Chicken Germ Line. PLoS Genet. *8*, e1002560.

Haig, D., and Grafen, A. (1991). Genetic scrambling as a defence against meiotic drive. J. Theor. Biol. *153*, 531–558.

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

Hall, A.B., Qi, Y., Timoshevskiy, V., Sharakhova, M.V., Sharakhov, I.V., and Tu, Z. (2013). Six novel Y chromosome genes in Anopheles mosquitoes discovered by independently sequencing males and females. BMC Genomics *14*, 1–13.

Hall, A.B., Papathanos, P.-A., Sharma, A., Cheng, C., Akbari, O.S., Assour, L., Bergman, N.H., Cagnetti, A., Crisanti, A., Dottorini, T., et al. (2016). Radical remodeling of the Y chromosome in a recent radiation of malaria mosquitoes. Proc. Natl. Acad. Sci. *113*, E2114–E2123.

Hamilton, W.D. (1967). Extraordinary Sex Ratios. Science *156*, 477–488.

Hauser, F. (2006). Identifying neuropeptide and protein hormone receptors in Drosophila melanogaster by exploiting genomic data. Brief. Funct. Genomic. Proteomic. *4*, 321–330.

Hillers, K.J. (2004). Crossover interference. Curr. Biol. *14*, R1036–R1037.

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

Hunt, R.H., Coetzee, M., and Fettene, M. (1998). The Anopheles gambiae complex: a new species from Ethiopia. Trans. R. Soc. Trop. Med. Hyg. *92*, 231–235.

Hurst, L.D., and Pomiankowski, A. (1991). Causes of Sex Ratio Bias May Account for Unisexual Sterility in Hybrids: A New Explanation of Haldane's Rule and Related Phenomena. Genetics *128*, 841–858.

Iordanou, E., Chandran, R.R., Blackstone, N., and Jiang, L. (2011). RNAi Interference by dsRNA Injection into Drosophila Embryos. J. Vis. Exp.

Ishikawa, R., and Kinoshita, T. (2009). Epigenetic programming: the challenge to species hybridization. Mol Plant *2*, 589–599.

Jaenike, J. (2001). Sex chromosome meiotic drive. Annu. Rev. Ecol. Syst. 25–49.

Johnson, N.A. (2010). Hybrid incompatibility genes: remnants of a genomic battlefield? Trends Genet. *26*, 317–325.

Josefsson, C., Dilkes, B., and Comai, L. (2006). Parent-Dependent Loss of Gene Silencing during Interspecies Hybridization. Curr. Biol. *16*, 1322–1328.

Kamali, M., Xia, A., Tu, Z., and Sharakhov, I.V. (2012). A new chromosomal phylogeny supports the repeated origin of vectorial capacity in malaria mosquitoes of the Anopheles gambiae complex. Plos Pathog. *8*, e1002960.

Kelly, W.G., Schaner, C.E., Dernburg, A.F., Lee, M.-H., Kim, S.K., Villeneuve, A.M., and Reinke, V. (2002). X-chromosome silencing in the germline of C.elegans. Development *129*, 479–492.

Kelsey, G., and Feil, R. (2013). New insights into establishment and maintenance of DNA methylation imprints in mammals. Philos. Trans. R. Soc. B Biol. Sci. *368*, 20110336.

Kirkpatrick, M. (2006). Chromosome Inversions, Local Adaptation and Speciation. Genetics *173*, 419–434.

Krzywinska, E., Dennison, N.J., Lycett, G.J., and Krzywinski, J. (2016). A maleness gene in the malaria mosquito Anopheles gambiae. Science *353*, 67–69.

Labrador, M., Farré, M., Utzet, F., and Fontdevila, A. (1999). Interspecific hybridization increases transposition rates of Osvaldo. Mol. Biol. Evol. *16*, 931–937.

Laschiavo, M., Nguyen, Q.K., Duelis, A.R., and Vrana, P.B. (2007). Maping and identification of candidate loci responsible for Peromyscus hybrid overgrowth. Mamm. Genome *18*, 75–85.

Laurie, C.C. (1997). the weaker sex is heterogametic: 75 years of Haldane's rule. Genetics *147*, 937–951.

Lees, R.S., Knols, B., Bellini, R., Benedict, M.Q., Bheecarry, A., Bossin, H.C., Chadee, D.D., Charlwood, J., Dabiré, R.K., Djogbenou, L., et al. (2014). Review: Improving our knowledge of male mosquito biology in relation to genetic control programmes. Acta Trop. *132*, S2–S11.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.

Lifschytz, E., and Lindsley, D.L. (1972). The Role of X-Chromosome Inactivation during Spermatogenesis. Proc. Natl. Acad. Sci. *69*, 182–186.

Linnaeus, C. (1735). Systema Naturae (Netherlands: Johann Friedrich Gmelin).

Liu, B.H. (1997). Statistical genomics: linkage, mapping, and QTL analysis (CRC Press).

Llopart, A. (2012). The rapid evolution of X-linked male-biased gene expression and the large-X effect in Drosophila yakuba, D. santomea, and their hybrids. Mol. Biol. Evol. *29*, 3873–3886.

Lloyd, V. (2000). Parental imprinting in Drosophila. Genetica *109*, 35–44.

Loppin, B., Bonnefoy, E., Anselme, C., Laurencon, A., Karr, T.L., and Couble, P. (2005). The histone H3.3 chaperone HIRA is essential for chromatin assembly in the male pronucleus. Nature *437*, 1386–1390.

Macias, V., Coleman, J., Bonizzoni, M., and James, A.A. (2014). piRNA pathway gene expression in the malaria vector mosquito *Anopheles stephensi*: piRNA genes in *Anopheles stephensi*. Insect Mol. Biol. *23*, 579–586.

Magnusson, K., Lycett, G.J., Mendes, A.M., Lynd, A., Papathanos, P.-A., Crisanti, A., and Windbichler, N. (2012). Demasculinization of the Anopheles gambiae X chromosome. BMC Evol. Biol. *12*, 1–13.

Maheshwari, S., and Barbash, D.A. (2011). The Genetics of Hybrid Incompatibilities. Annu. Rev. Genet. *45*, 331–355.

Malone, J.H., and Michalak, P. (2008). Physiological Sex Predicts Hybrid Sterility Regardless of Genotype. Science *319*, 59.

Marchand, R.P. (1984). Field Observations on Swarming and Mating in Anopheles gambiae Mosquitoes in Tanzania. Neth J Zool *34*, 367–387.

Martinez, O., and Curnow, R.N. (1992). Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. *85*, 480–488.

Masly, J.P., and Presgraves, D.C. (2007). High-resolution genome-wide dissection of the two rule of speciation in Drosophila. PLOS Biol. *5*, e243.

Mason, L.J., Pashley, D.P., and Johnson, S.J. (1987). The Laboratory as an Altered Habitat: Phenotypic and Genetic Consequences of Colonization. Fla. Entomol. *70*, 49.

Mattingly, P.F. (1977). Names for the Anopheles gambiae complex. Mosq. Syst. *9*, 323–328.

Mayr, E. (1963). Animal species and Evolution (Cambridge: Belknap Press of Harvard University Press).

McDermott, S.R., and Noor, M.A. (2010). The role of meiotic drive in hybrid male sterility. Philos. Trans. R. Soc. B Biol. Sci. *365*, 1265–1272.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

Megy, K., Emrich, S.J., Lawson, D., Campbell, D., Dialynas, E., Hughes, D.S.T., Koscielny, G., Louis, C., MacCallum, R.M., Redmond, S.N., et al. (2012). VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. Nucleic Acids Res. *40*, D729–D734.

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

Meiklejohn, C.D., Landeen, E.L., Cook, J.M., Kingan, S.B., and Presgraves, D.C. (2011). Sex Chromosome-Specific Regulation in the Drosophila Male Germline But Little Evidence for Chromosomal Dosage Compensation or Meiotic Inactivation. PLoS Biol. *9*, e1001126.

Meisel, R.P., and Connallon, T. (2013). The faster-X effect: integrating theory and data. Trends Genet. *29*, 537–544.

Metcalfe, C.J., Bulazel, K.V., Ferreri, G.C., Schroeder-Reiter, E., Wanner, G., Rens, W., Obergfell, C., Eldridge, M.D.B., and O'Neill, R.J. (2007). Genomic instability within centromeres of interspecific marsupial hybrids. Genetics *177*, 2507–2517.

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

Moehring, A.J., Teeter, K.C., and Noor, M.A. (2007). Genome-wide patterns of expression in Drosophila pure species and hybrid males II Examination of multiple-species hybridizations, platforms, and life cycle stages. Mol. Biol. Evol. *24*, 137–145.

Moore, T., and Haig, D. (1991). Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet. *7*, 45–49.

Moran, P.A., Ritchie, M.G., and Bailey, N.W. (2017). A rare exception to Haldane's rule: Are X chromosomes key to hybrid incompatibilities? Heredity.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

Moyle, L.C. (2006). Genome-Wide Associations Between Hybrid Sterility QTL and Marker Transmission Ratio Distortion. Mol. Biol. Evol. *23*, 973–980.

Muller, H. (1942). Isolating mechanisms, evolution and temperature. Biol Symp 71–125.

Mwangangi, J.M., Mbogo, C., Nzovu, J.G., Githure, J.I., Yan, G., and Beier, J.C. (2003). Blood-meal analysis for anopheline mosquitoes sampled along the Kenyan coast. J. Am. Mosq. Control Assoc. *19*, 371–375.

Nolan, T., Papathanos, P., Windbichler, N., Magnusson, K., Benton, J., Catteruccia, F., and Crisanti, A. (2011). Developing transgenic Anopheles mosquitoes for the sterile insect technique. Genetica *139*, 33–39.

Noor, M.A., Grams, K.L., Bertucci, L.A., and Reiland, J. (2001a). Chromosomal inversions and the reproductive isolation of species. Proc. Natl. Acad. Sci. *98*, 12084–12088.

Noor, M.A., Cunningham, A.L., and Larkin, J.C. (2001b). Consequences of recombination rate variation on quantitative trait locus mapping studies: simulations based on the Drosophila melanogaster genome. Genetics *159*, 581–588.

Nosil, P., and Schluter, D. (2011). The genes underlying the process of speciation. Trends Ecol. Evol. *26*, 160–167.

Oliver, P.L., Goodstadt, L., Bayes, J.J., Birtle, Z., Roach, K.C., Phadnis, N., Beatson, S.A., Lunter, G., Malik, H.S., and Ponting, C.P. (2009). Accelerated Evolution of the Prdm9 Speciation Gene across Diverse Metazoan Taxa. PLoS Genet. *5*, e1000753.

O'Neil, R.J.W., O'Neil, M.J., and Graves, J.A.M. (1998). Undermethylation associated with retroelement activation and chromosome remodeling in a interspecific mammalian hybrid. Nature *393*, 68–72.

Orr, H.A. (1993). Haldane's rule has multiple genetic causes. Nature *361*, 532–533.

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.

Orr, H.A., and Turelli, M. Dominance and Haldane's rule. Genetics *143*, 613.

Park, J.-S., Lee, S.-H., Na, H.-J., Pyo, J.-H., Kim, Y.-S., and Yoo, M.-A. (2012). Age- and oxidative stress-induced DNA damage in Drosophila intestinal stem cells as marked by Gamma-H2AX. Exp. Gerontol. *47*, 401–405.

Payseur, B.A., Krenz, J.G., and Nachman, M.W. (2004). Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. Evolution *58*, 2064–2078.

Perez, D.E., and Wu, C.-I. (1995). Further Characterization of the Odysseus Locus of Hybrid Sterility in Drosophila: One Gene is Not Enough. Genetics *140*, 201–206.

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

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., and Hoekstra, H.E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE *7*, e37135.

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

Phadnis, N., Baker, E.P., Cooper, J.C., Frizzell, K.A., Hsieh, E., De La Cruz, A.F.A., Shendure, J., Kitzman, J.O., and Malik, H.S. (2015). An essential cell cycle regulation gene causes hybrid inviability in Drosophila. Science *350*, 1552–1555.

Pock Tsy, J.-M.L., Duchemin, J.-B., Marrama, L., Rabarison, P., Le Goff, G., Rajaonarivelo, V., and Robert, V. (2003). Distribution of the species of the Anopheles gambiae complex and the first evidence of Anopheles merus as a malaria vector in Madagascar. Malar. J. *2*, 1–7.

Pombi, M., Stump, A.D., Della Torre, A., and Besansky, N.J. (2006). Variation in recombination rate across the X chromosome of Anopheles gambiae. Am J Trop Med Hyg *75*, 901–903.

Presgraves, D.C. (2002). Patterns of postzygotic isolation in Lepidoptera. Evolution *56*, 1168–1183.

Presgraves, D.C. (2008). Sex chromosomes and speciation in Drosophila. Trends Genet. *24*, 336–343.

Presgraves, D.C. (2010). The molecular evolutionary basis of species formation. Nat. Rev. Genet. *11*, 175–180.

Presgraves, D.C., and et al. (2003). Adaptive evolution drives divergence of a hybrid inviability gene between two species of Drosophila. Nature *423*, 715–719.

Presgraves, D.C., and Orr, H.A. (1998). Haldane's rule in taxa lacking a hemizygous X. Science *282*, 952–954.

Ranson, H., N'Guessan, R., Lines, J., Moiroux, N., Nkuni, Z., and Corbel, V. (2011). Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? Trends Parasitol. *27*, 91–98.

Ray, J. (1686). Historia Plantarum (London: Clark).

Reddy, M.R., Overgaard, H.J., Abaga, S., Reddy, V.P., Caccone, A., Kiszewski, A.E., and Slotman, M.A. (2011). Outdoor host seeking behaviour of Anopheles gambiae mosquitoes following initiation of malaria vector control on Bioko Island, Equatorial Guinea. Malar. J. *10*, 184.

Renaut, S., Nolte, A.W., and Bernatchez, L. (2009). Gene Expression Divergence and Hybrid Misexpression between Lake Whitefish Species Pairs (Coregonus spp. Salmonidae). Mol. Biol. Evol. *26*, 925–936.

Sawamura, K. (1996). Maternal effect as a cause of exceptions for Haldane's rule. Genetics *143*, 609.

Sawamura, K., Taira, T., and Watanabe, T.K. (1993). Hybrid lethal systems in the Drosophila melanogaster species complex. I. the maternal hybrid rescue (mhr) gene of Drosophlia simulans. Genetics *133*, 299–305.

Schilthuizen, M., Giesbers, M., and Beukeboom, L.W. (2011). Haldane's rule in the 21st century. Heredity *107*, 95–102.

Schoenmakers, S., Wassenaar, E., Hoogerbrugge, J.W., Laven, J.S.E., Grootegoed, J.A., and Baarends, W.M. (2009). Female Meiotic Sex Chromosome Inactivation in Chicken. PLoS Genet. *5*, e1000466.

Schoenmakers, S., Wassenaar, E., Laven, J.S.E., Grootegoed, J.A., and Baarends, W.M. (2010). Meiotic silencing and fragmentation of the male germline restricted chromosome in zebra finch. Chromosoma *119*, 311–324.

Schulz, R., Proudhon, C., Bestor, T.H., Woodfine, K., Lin, C.-S., Lin, S.-P., Prissette, M., Oakey, R.J., and Bourc'his, D. (2010). The Parental Non-Equivalence of Imprinting Control Regions during Mammalian Development and Evolution. PLoS Genet. *6*, e1001214.

Seaton, G., Haley, C.S., Knott, S.A., Kearsey, M., and Visscher, P.M. (2002). QTL Express: mapping quantitative trait loci in simple and complex pedigrees. Bioinformatics *18*, 339–340.

Sharakhova, M.V., and Sharakhov, I.V. (2010). Organization and evolution of heterochromatin in malaria mosquitoes. Russ. J. Genet. *46*, 1250–1253.

Singh, N.D., Larracuente, A.M., and Clark, A.G. (2008). Contrasting the efficacy of selection on the X and autosomes in Drosophila. Mol. Biol. Evol. *25*, 454–467.

Sinka, M.E., Bangs, M.J., Manguin, S., Coetzee, M., Mbogo, C.M., Hemingway, J., Patil, A.P., Temperley, W.H., Gething, P.W., Kabaria, C.W., et al. (2010). The dominant Anopheles vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis.

Slotman, M., Della Torre, A., and Powell, J.R. (2004). The genetics of inviability and male sterility in hybrids between Anopheles gambiae and An. arabiensis. Genetics *167*, 275–287.

Slotman, M.A., Reimer, L.J., Thiemann, T., Dolo, G., Fondjo, E., and Lanzaro, G.C. (2006). Reduced Recombination Rate and Genetic Differentiation Between the M and S Forms of Anopheles gambiae s.s. Genetics *174*, 2081–2093.

Snoek, M., Teuscher, C., and van Vugt, H. (1998). Molecular analysis of the major MHC recombinational hot spot located within the G7c gene of the murine class III region that is involved in disease susceptibility. J. Immunol. *160*, 266–272.

Städler, T., Florez-Rueda, A.M., and Paris, M. (2012). Testing for “Snowballing” Hybrid Incompatibilities in Solanum: Impact of Ancestral Polymorphism and Divergence Estimates. Mol. Biol. Evol. *29*, 31–34.

Stanley, S.M. (1979). Macroevolution, Pattern and Process (San Francisco: W.H. Freeman).

Storchova, R., and Divina, P. (2006). Nonrandom representation of sex-biased genes on chicken Z chromosome. J. Mol. Evol. *63*, 676–681.

Stump, A.D., Pombi, M., Goeddel, L., Ribeiro, J.M.C., Wilder, J.A., Torre, A.D., and Besansky, N.J. (2007). Genetic exchange in 2La inversion heterokaryotypes of Anopheles gambiae. Insect Mol. Biol. *16*, 703–709.

Stupar, R.M., Hermanson, P.J., and Springer, N.M. (2007). Nonadditive Expression and Parent-of-Origin Effects Identified by Microarray and Allele-Specific Expression Profiling of Maize Endosperm. PLANT Physiol. *145*, 411–425.

Stupar, R.M., Gardiner, J.M., Oldre, A.G., Haun, W.J., Chandler, V.L., and Springer, N.M. (2008). Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. BMC Plant Biol. *8*, 33.

Sun, S., Ting, C.-T., and Wu, C.-I. (2004). the normal function of a speciation gene, Odysseus, and its hybrid sterility effect. Science *305*, 81–83.

Tang, S., and Presgraves, D.C. (2015). Lineage-Specific Evolution of the Complex Nup160 Hybrid Incompatibility Between Drosophila melanogaster and Its Sister Species. Genetics *200*, 1245–1254.

Tao, Y., Chen, S., Hartl, D.L., and Laurie, C.C. (2003). Genetic dissection of hybrid incompatibilities between Drosophila simulans and D. mauritiana. I. Differential accumulation of hybrid male sterility effects on the X and autosomes. Genetics *164*, 1383–1398.

Tao, Y., Masly, J.P., Ke, Y., and Hartl, D.L. (2007b). a sex-ratio meiotic drive system in Drosophila simulans I: An autosomal suppressor. PLOS Biol. *5*, e292.

Tao, Y., Malsy, J.P., Araripe, L., Ke, Y., and Hartl, D. (2007a). a sex-ratio meiotic drive system in Drosophila simulans II: an X-linked distorter. PLOS Biol. *5*, e293.

Teeter, K.C., Payseur, B.A., Harris, L.W., Bakewell, M.A., Thibodeau, L.M., O'Brien, J.E., Krenz, J.G., Sans-Fuentes, M.A., Nachman, M.W., and Tucker, P.K. (2007). Genome-wide patterns of gene flow across a house mouse hybrid zone. Genome Res. *18*, 67–76.

Temu, E.A., Minjas, J.N., Coetzee, M., Hunt, R.H., and Shift, C.J. (1998). The role of four anopheline species (Diptera: Culicidae) in malaria transmission in coastal Tanzania. Trans. R. Soc. Trop. Med. Hyg. *92*, 152–158.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. *13*, 2129–2141.

Tiffin, P., Olson, M.S., and Moyle, L.C. (2001). Asymmetric crossing barriers in angiosperms. Proc. R. Soc. B *268*, 861–867.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

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

Tulchinsky, A.Y., Johnson, N.A., and Porter, A.H. (2014). Hybrid incompatibility despite pleiotropic constraint in a sequence-based bioenergetic model of transcription factor binding. Genetics *198*, 1645–1654.

Turelli, M., and Moyle, L.C. (2006). Asymmetric Postmating Isolation: Darwin's Corollary to Haldane's Rule. Genetics *176*, 1059–1088.

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

Turelli, M., and Orr, H.A. (2000). Dominance, Epistasis, and the Genetics of Postzygotic Isolation. Genetics *154*, 1663–1679.

Turner, J.M.A. (2007). Meiotic sex chromosome inactivation. Development *134*, 1823–1831.

Volff, J.N. (2005). Genome evolution and biodiversity in teleost fish. Heredity *94*, 280–294.

Vrana, P.B., Guan, X.-J., Ingram, R.S., and Tilghman, S.M. (1998). Genomic imprinting is disrupted in interspecific Peromyscus hybrids. Nat. Genet. *20*, 362–365.

Wang, H., and Dey, S.K. (2006). Roadmap to embryo implantation: clues from mouse models. Nat. Rev. Genet. *7*, 185–199.

Wang, S., and Jacobs-Lorena, M. (2013). Genetic approaches to interfere with malaria transmission by vector mosquitoes. Trends Biotechnol. *31*, 185–193.

Wang, D., Liu, F., Wang, L., Huang, S., and Yu, J. (2011). Nonsynonymous substitution rate (Ka) is a relatively consistent parameter for the defining fast-evolving and slow-evolving protein-coding genes. Biol. Direct *6*, 1–17.

Watson, E.T., and Demuth, J.P. (2012). Haldane's Rule in Marsupials: What Happens When Both Sexes Are Functionally Hemizygous? J. Hered. *103*, 453–458.

Weetman, D., Steen, K., Rippon, E.J., Mawejje, H.D., Donnelly, M.J., and Wilding, C.S. (2014). Contemporary gene flow between wild An. gambiae ss and An. arabiensis. Parasit. Vectors *7*, 345.

White, G.B. (1971). Chromosomal evidence for natural interspecific hybridization by mosquitoes of the Anopheles gambiae complex. Nature *231*, 184–185.

White, G.B. (1973). Comparative Studies on Sibling Species of Anopheles-Gambiae Giles Complex (Dipt Culicidae) 3. Distribution, Ecology, Behavior and Vectorial Importance of Species-D in Bwamba-County, Uganda, with an Analysis of Biological, Ecological, Morphological and Cytogenetical Relationships of Ugandan Species-D. Bull. Entomol. Res. *63*, 65–97.

White, B.J., Collins, F.H., and Besansky, N.J. (2011). Evolution of Anopheles gambiae in Relation to Humans and Malaria. Annu. Rev. Ecol. Syst. *42*, 111–132.

White, B.J., Kundert, P.N., Turissini, D.A., Van Ekeris, L., Linser, P.J., and Besansky, N.J. (2013). Dose and developmental responses of Anopheles merus larvae to salinity. J. Exp. Biol. *216*, 3433–3441.

Wolf, J.B., Oakey, R.J., and Feil, R. (2014). Imprinted gene expression in hybrids: perturbed mechanisms and evolutionary implications. Heredity *113*, 167–175.

Wright, K.M., Lloyd, D., Lowry, D.B., Macnair, M.R., and Willis, J.H. (2013). Indirect evolution of hybrid lethality due to linkage with selected locus in Mimulus guttatus. PLOS Biol. *11*, e1001497.

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

Wu, C.-I., and Ting, C.-T. (2004). Genes and speciation. Nat. Rev. Genet. *5*, 114–122.

Wu, C.-I., Johnson, N.A., and Palopoli, M.F. (1996). Haldane's rule and its legacy: Why are there so many sterile males? Trends Ecol. Evol. *11*, 281–284.

Yan, Z., Zhang, W., He, Y., Gorczyca, D., Xiang, Y., Cheng, L.E., Meltzer, S., Jan, L.Y., and Jan, Y.N. (2012). Drosophila NOMPC is a mechanotransduction channel subunit for gentle-touch sensation. Nature *493*, 221–225.

Zheng, L., Benedict, M., Cornel, A.J., Collins, F.H., and Kafatos, F.C. (1996). An integrated genetic map of the African human malaria vector mosquito, Anopheles gambiae. Genetics *143*, 941–952.

Zouros, E., Lofdahl, K., and Martin, P.A. (1988). Male Hybrid Sterility in Drosophila: Interactions between Autosomes and Sex Chromosomes in Crosses of D. mojavensis and D. arizonensis. Evolution *42*, 1321.

Zufall, R.A., and Rausher, M.D. (2004). Genetic changes associated with floral adaptation restrict future evolutionary potential. Nature *428*, 847–850.

(2005). World Malaria Report.

(2015). World Malaria Report.