SEX CHROMOSOME EVOLUTION IN BLOW FLIES

by

Anne Amarila Andere

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Biology at IUPUI

Indianapolis, Indiana

August 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL

STATEMENT OF COMMITTEE APPROVAL

Dr. Christine Picard, Chair

Biology Department, IUPUI

Dr. Matthew Hahn

Biology Department, IU Bloomington

Dr. Nicolas Berbari

Biology Department, IUPUI

Dr. Stephen Randall

Biology Department, IUPUI

Dr. Yunlong Liu

IUSM, Department of Medical and Molecular Genetics, IUPUI

Approved by:

Dr. Theodore Cummins

Dedicated to my family

ACKNOWLEDGMENTS

I would like to take this opportunity to heartily thank everyone who made this journey possible. First and foremost, I would like to thank Dr. Christine Picard for accepting me in her lab as a master's student and entrusting me with a project that would lead me to pursue my doctorate degree in the same lab. With her constant encouragement, I received numerous opportunities to present my work both locally and nationally, for which I am forever grateful. To my committee members – Dr. Yunlong Liu, Dr. Matthew Hahn, Dr. Nicolas Berbari, Dr. Stephen Randall – thank you for your support, patience and guidance throughout my whole PhD journey. To the biology department, I am grateful to have had the opportunity to be a part of this extremely welcoming department. A special thank you to Dr. Katherine Marrs and Dr. Randall Roper for their mentorship and friendship throughout my PhD journey. Thank you to Dr. Ann Grens for encouraging me to target my goals as an undergrad, and a big thank you to Dr. Murli Nair for having faith in me and introducing me to the field of computational biology.

I would also like to thank all the members of the Picard Lab who have become my extended family. I was blessed to be a part of a lab which offered a lot of compassion from everyone who passed through it. Dr. Charity Owings – thank you for being a wonderful friend, and a support system. You have stood up for me in countless occasions, sometimes short of engaging in fights to defend me. Only a true friend can step up like that and I am forever grateful! Laura Doll and Khadija Moctar thank you for making sure I took breaks in between my marathon working hours and always ready to lend an ear and a hand. Dr. Gina Dembinski and John Whale – I am extremely grateful we crossed paths and are still good friends. May the friendship last forever. To my other labmates: Dr. Ti Eriksson, Sarah Lewis, Ioanna Koltsidou, Kevin Smolar, Dr. Abeer Mohsen and all the undergrads I mentored (Rabiah Amjad, Christina Martin, Grace LeFevre, Grace King, Thalles Lopes) – thank you for the encouragement and never-ending support. I would also like to thank the other members of the SLonely biology labs (2nd floor SL biology labs) for their constant support. All the members of Lata's lab (especially Dr. Lata Balakrishnan, Dr. Catherine Njeri, Yeka Ononye, and Dr. Sneha Surendran) thank you for your moral support and encouragement. Catherine – the world is definitely small! Never in a million years would I have predicted us being high school desk mates as teenagers back in Kenya, then miraculously bumping into each other in good ole Indiana after more than ten years apart, and ending up working in the same department,

same floor, and labs next to each other. God has a unique sense of humor ☺. Thank you for all your support, friendship and constant reminder to always target my goals. To the members of the Kuczmierczyk lab (especially Dr. Dilraj Panfair, Dr. Lindsay Hammack, Kyle Firestone, Aishwarya Ramamurthy) – thank you for your unwavering friendship. To my good friend Susan Agwata, living in different continents has not derailed our friendship and daily communication. Thank you for your patience, kindness, and friendship, and always being available despite the distance.

Last but not least, I would like to thank my family who have been my rock! I thank God for all of you. You have patiently been by my side throughout my entire academic journey – through thick and thin, good days and bad days, highs and lows. Thank you for stepping in as guardians after mom and dad passed away. Betty, Mary, Lois, Jada, Linda, Cindy, Alice – you are all amazing, and I couldn't ask for any better siblings than you all. Aggy, I wish you were here but I know you have been rooting for me from above. Mom and dad, I finally became a doctor – not the kind you had in mind, but a different kind of doctor ☺.

Thank you to everyone (friends, family, mentors) whom I did not have room to mention, there are so many of you who have impacted my life.

Thank you, God, for seeing me through and leading me each step of the way – Phillipians 4:13.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Chromosomal mechanisms of sex determination vary greatly in phylogenetically closely related species, indicative of rapid evolutionary rates. Sex chromosome karyotypes are generally conserved within families; however, many species have derived sex chromosome configurations. Insects display a plethora of sex chromosome systems due to rapid diversification caused by changes in evolutionary processes within and between species. A good example of such a system are insects in the blow fly family Calliphoridae. While cytogenetic studies observe that the karyotype in blow flies is highly conserved (five pairs of autosomal chromosomes and one pair sex chromosome), there is variation in sex determining mechanisms and sex chromosome structure within closely related species in blow flies. The evolutionary history of sex chromosomes in blow fly species have not been fully explored. Therefore, the objective of this research was to characterize the sex chromosome structures in four species of blow flies and investigate the selective forces which have played a role in shaping the diverse sex chromosome system observed in blow flies. The blow fly species used in this study are *Phormia regina, Lucilia cuprina, Chrysomya rufifacies* and *Chrysomya albiceps. Phormia regina,*and *Lucilia cuprina* have a heteromorphic sex chromosome system and are amphogenic (females produce both male and female offspring in equal ratio). In contrast, *Chrysomya rufifacies* and *Chrysomya albiceps*, have a homomorphic sex chromosome system, are monogenic (females produce unisexual progeny), have two types of females (arrhenogenic females – male producers and thelygenic females – female producers), and sex of the offspring is determined by the maternal genotype.

To accomplish these tasks, a total of nine male and female individual draft genomes for each of the four species (including three individual draft genomes of *Chrysomya rufifacies* – male, and the two females) were sequenced and assembled providing genomic data to explore sex chromosome evolution in blow flies. Whole genome analysis was utilized to characterize and identify putative sex chromosomal sequences of the four blow fly species. Genomic evidence confirmed the presence of genetically differentiated sex chromosomes in *P. regina* and *L. cuprina;* and genetically undifferentiated sex chromosomes in *C. rufifacies* and *C. albiceps.* Furthermore, comparative analysis of the ancestral Dipteran sex chromosome (Muller element F in *Drosophila)* was determined to be X-linked in *P. regina* and *L. cuprina* contributing to sex chromosome differentiation but not sex-linked in *C. rufifacies* and *C. albiceps*. Evolutionary pressures are often

quantified by the ratio of substitution rates at non-synonymous (dN) and synonymous (dS) sites. Substitution rate ratio analysis (dN/dS) of homologous genes indicated a weaker purifying selection may have contributed to the loss of sex-linked genes in Muller element F genes of the undifferentiated sex chromosome as compared to the differentiated sex chromosome system. Overall, the results presented herein greatly expands our knowledge in sex chromosome evolution within blow flies and will reinforce the study of sex chromosome evolution in other species with diverse sex chromosome systems.

CHAPTER 1. INTRODUCTION

1.1 Origin and Evolution of Sex Chromosomes

Sex chromosomes are one of the main biological components that drive and facilitate sexual differentiation and dimorphism between sexes in eukaryotes [1, 2]. In a given species, sex chromosomes pairs differ between the sexes and contribute to the divergent phenotypes observed between males and females in behavior, morphology, and physiology [3]. For most eukaryotic organisms, sex is determined by the presence of heteromorphic sex chromosomes [4, 5]. The sex chromosomes can be XY for species with male heterogamety (for example humans), where males are XY and females are XX (using an XX/XY sex-determining system) or ZW for species with female heterogamety (for example birds [6], and butterflies [7]) where females are ZW and males are ZZ (using a ZZ/ZW sex-determining system) [4, 8-10]. Sex chromosome evolution postulates that heteromorphic sex chromosomes originated from a pair of ordinary homologous autosomes, upon which a sex-determining gene was acquired on one of the homologous chromosomes [9, 11, 12]. Sexually antagonistic mutations which are mutations beneficial for one sex but harmful for the other, begin to accumulate in the region close to the newly acquired sex-determining gene resulting in selective pressures that favor the suppression or elimination of recombination between the new proto-X/Y or Z/W chromosomes [11, 12]. Onset of reduction of recombination (restriction of recombination) results in the accumulation of deleterious mutations in the sex-limited chromosome (Y/W) which begin to allow the sex chromosomes to diverge both functionally and morphologically. Distinct differences begin to form between the X (Z) and Y(W), with Y(W) degeneration due to lose of active genes, thus evolving into heteromorphic and differentiated sex chromosomes [12-15]. In some extreme cases, there is a continual degeneration of the Y(W), leading to a loss of functional genes on the Y(W) chromosome. This results in an ultimate loss of the Y(W) chromosome resulting to XX/X0 (ZZ//Z0) sex system (for example some mole voles [16], and some grasshoppers species [17]).

The underlying biology and traditional view of sex chromosome evolution has mostly stemmed from a few notable well–studied model organisms, specifically humans and *Drosophila melanogaster* (the common fruit fly) [11]. The standard default theme here is the presence of X(Z) and Y(W) heteromorphic sex chromosomes with a differentiated Y(W) are characterized as old

sex chromosomes [9]. In humans, the X and Y chromosome originated ~200 – 300 million years ago (mya) in eutherian mammals [11, 12] while *Drosophila* originated ~60 mya [18]. This 'simple' design XX/XY (ZZ/ZW) does not, however, reflect the variety of sex chromosome systems present. The existence of homomorphic sex chromosomes proves how labile sex chromosome evolution is. Sex chromosomes which display low levels of differentiation are usually considered to be evolutionarily young and at the initial stages of evolution [19, 20]. The expectation is that they will eventually differentiate. For example, the homomorphic sex chromosomes of *Drosophila miranda* evolved ~1mya; meaning they are still recombining [20], and the neo-Y (chromosome transforming into sex chromosome) has undergone some degeneration over time and its sex chromosomes are in the process of differentiating [20-22].

Examples of organisms which display homomorphic sex chromosomes include some birds [23, 24], snakes [25], and insects (mosquitoes of the family Culicidae) [26]. Some organisms, (for example snakes) are unique in that they possess variation in the level of sex chromosome heteromorphism [25, 27]. Snakes exhibit female heterogametic sex chromosomes (ZZ males/ZW females). Snakes in the families Pythonidae (pythons) and Boidae (boas) have homomorphic sex chromosomes where the Z and W chromosomes appear undifferentiated at the cytological level, while those within the family Elapidae and Viperidae have a highly degenerated W [25, 27]. In contrast, the Colubridae family appear to be have moderately differentiated Z and W chromosome karyotypes and are suggested to be at an intermediate stage of sex chromosome evolution [25, 27]. Mosquitoes follow a similar system. Mosquitoes in the genus *Aedes* and *Culex* (subfamily Culicinae) have a male determining locus located in a morphologically undifferentiated (homomorphic) sex chromosome [28, 29]. In contrast mosquitoes within the Anophelinae subfamily have fully morphologically differentiated (heteromorphic) sex chromosomes (XY) [28, 29].

As with most of nature, not all homomorphic sex chromosomes are young. Some bird species (emus [23]) and mosquito species (*Aedes* [29]) have old homomorphic sex chromosomes, which have not diverged from their ancestral state. Mapping studies show that emus sex chromosomes are largely homologous with only a small differentiated region [30, 31]. In contrast, *Aedes* mosquito has genetically differentiated homomorphic sex chromosomes. The differentiated region occurs over a much larger region than the sex determining region (~40% of the whole sex chromosome) [28]. It is not quite clear why there is a lack of differentiation in some lineages;

however, some evolutionary models have been put forth to explain possible causes. Suppression of recombination is one of the strategies to resolve conflict caused by the presence of sexually antagonistic mutations which favor either sex (sex-biased genes). It does this in order to eliminate deleterious effects of a sexually antagonistic allele which may be harmful to the sex it does not benefit [32, 33]. However, in old homomorphic systems, the selective pressure to abolish recombination may be eliminated due to the evolution of sex biased expression at sexually antagonistic alleles along the protosex chromosomes (newly evolved sex chromosome) [23].

Other examples of the evolutionary lability of sex chromosome evolution involves the formation of new sex chromosomes by the translocation or fusion of a differentiated sex chromosome with autosomes to create neo-sex chromosomes [20, 34], or the transposition of a dominant sex determining gene to an autosome initiating the cycle of sex chromosome differentiation, from one autosome to another [1]. *Drosophila miranda* has recently formed new sex chromosomes pair (neo-sex chromosomes) due to a fusion of an autosome to the Y chromosome [20, 21]. The neo-Y chromosome is therefore in transition from an ordinary autosome into a degenerate Y chromosome. This is evidenced by the presence of non-functional genes containing frame-shift mutations or stop codons and other signs of degeneration such as elevated rates of amino acid substitutions [20, 21]. In *Drosophila pseudoobscura,* its X chromosome consists of two chromosomal arms resulting from a fusion of the ancestral X chromosome to an autosome which occurred ~18mya [15, 35]. Genes which were originally on the ancestral Y chromosome translocated to another chromosome and became autosomal by purging much of the repetitive DNA that had accumulated on the ancestral Y chromosome [15, 35].

1.2 Sex Chromosome Structures

Sex chromosomes in most eukaryotic organisms are typically morphologically and genetically distinct between males and females [9, 11, 36, 37]. Differentiated heteromorphic sex chromosomes are usually distinguishable as they differ in size, gene and repetitive sequence content [4, 11, 36, 38]. In both the XX/XY and ZZ/ZW sex chromosome systems, the X (Z) is typically large and euchromatic while the Y (W) is small and heterochromatic [9, 15]. The X chromosome is more often than not expected to contain a large number of genes as compared to the Y on the basis of its large size and is typically organized like autosomes. While the Y is expected to have a small number of genes due to its reduced size caused by suppressed

recombination [9, 15]. However, recombination still continues to take place in females sex chromosome (XX) resulting to the same gene density as autosomes, and a retention and maintenance of its original genes [15]. One of the key features that differentiates sex chromosomes from autosomes in a heterogametic sex chromosome system is the presence of a sex determining region on the heterogametic sex. In most species, the sex determining region may occur over some or most of the length of the heterogametic sex [11, 15, 38]. This region is characterized by a reduction of recombination; and also a male-limited transmission of the non-recombining segment [12]**.** The non-recombining Y, especially in eukaryotes with male heterogamety and highly differentiated sex chromosomes (a good example – the Y chromosome in *Drosophila* and in humans), has undergone genetic degeneration and lost many of its original genes which were present on the ancestral X [8, 9, 12].

1.2.1 Degeneration on the Y (W) Chromosome Due to Reduced Recombination

Suppression of recombination is a necessary condition for genetic divergence to begin between the X and Y sex chromosome [8, 9]. Lack of recombination on the Y(W) chromosome leads to the accumulation of deleterious mutations in most of its ancestral original genes [20, 21] [39]. Accumulation of deleterious mutations over a long evolutionary time subsequently accelerates the rate of degeneration leading to gene loss, which is usually associated with a simultaneous accumulation of repetitive DNA elements on the Y(W) chromosome [40]. Repetitive sequences such as retrotransposons, microsatellite repeats, and ribosomal DNAs has been reported in abundance on sex chromosomes of numerous animal and plant species [41, 42]. Richness in repetitive DNA sequences is involved in the heterochromatization of the Y(W) chromosome [12, 41, 43-45] with a higher density of transposons in heterochromatic DNA vs DNA present in euchromatin [46].

Analysis of DNA sequences from a young and newly evolved sex chromosome in *D. miranda* showed an enrichment of DNA insertions on the neo-Y chromosome [47]. The insertions were mainly transposable elements, specifically retrotransposons [20, 47, 48]. A similar accumulation of transposable elements has also been observed in other recently evolving Y(W) chromosomes systems such as plants (*Silene latifolia* – white campion [42, 49], *Marchantia polymorpha* –liverwort [50, 51], papaya [52], and animals (*Oryzias latipes* – medaka fish [53], *Gasterosteus aculeatus* – threespine stickleback [54]). This provides compelling evidence that

transposable elements are an early invader of newly formed Y chromosomes and are involved in the first step in Y chromosome degeneration. The accumulation of repetitive elements hence represents the early process of shaping the Y chromosome even before genes begin to degenerate [55].

Transposable elements are initiators of mutations: active transposons can lead to deleterious mutations which may disrupt the reading frame of functional genes or introduce premature stop codon inactivating the gene. Inactivation of genes leads to gene decay and gene loss which in evolutionary time results to the reduction in size of the neo-Y(W) chromosome. Additionally, they can also trigger the inactivation of adjacent genes on the neo-Y [20]. Indirectly, transposable elements also interfere with gene expression by producing antisense transcripts of adjacent genes or by altering the chromatin structure of the Y chromosome [20, 56, 57].

1.2.2 Pseudoautosomal Region (PAR)

Sex chromosomes in most organisms with differentiated heteromorphic sex chromosome systems differ from autosomal chromosomes (as mentioned earlier) [9, 58, 59]. They show difference in gene content, size and structure. However, in some taxa, recombination persists in one or more regions of the differentiated X and Y. These recombining regions are known as pseudoautosomal regions (PAR) [60]. Recombination and homologous pairing in this region aids in the maintenance of sequence homology between X and Y sex chromosome causing the PAR to exhibit autosomal inheritance [60-62]. Mapping studies in horse, chimpanzee, human and mice indicate that PAR varies in size and gene content in evolutionarily close as well as distantly related species [63].

The PAR region has been maintained for over 140 million years in eutherian mammals. The genes in this region escape X inactivation and are inherited in an autosomal manner more so than sex-linked fashion [64]. In humans, other great apes, and some other mammals such as mice, PAR is required for initiating and maintaining effective pairing of the X and Y chromosomes during segregation in male meiosis, and recombination [64]. It is thought that the X-Y pairing serves a critical function in spermatogenesis in mouse and humans [65, 66]. A deletion of the homologous region in X chromosome which mediates pairing with the Y chromosome produced infertile male mice [65]. A similar event was observed in humans when a male with complete failure of sex chromosome pairing due to a deletion in the PAR region of the X chromosome

caused spermatogenic development arrested at the metaphase stage in meiosis [66]. In other mammals, such as the marsupials, the PAR is absent and the X and Y sex chromosomes do not undergo homologous pairing. Absence of homologous pairing of XY chromosomes in this species do not appear to disrupt segregation at meiosis [64] and three homologs of the human PAR genes are autosomal in marsupials. It is unclear what has replaced homologous pairing recombination in marsupials [61, 64].

1.2.3 Dosage Compensation

When species have gene-rich X chromosomes, it is important that there be regulatory mechanisms in place to account for dosage differences between females with two copies and males with one copy of the relevant genes. In highly evolved sex chromosomes, some special adaptations such as dosage compensation [67, 68] or meiotic sex chromosome inactivation (MSCI) [69] occur in order to prevent the reversal of a heteromorphic chromosome back to autosomes [12]. Dosage compensation is a regulatory mechanism evolved to balance the expression of X-linked and autosomal genes in the heterogametic sex [68, 70]. The degeneration of the Y chromosome and subsequent loss of gene functions causes the X chromosome to appear haploid in males differing from the rest of the diploid genome. This results in a detrimental imbalance for dosage sensitive genes between males (XY) and females (XX). Dosage compensation therefore evolved to compensate for dosage imbalance on the X in the heterogametic sex (XY or XO) [68, 70]. This is to ensure that females with two X chromosomes and males' with only one X would have equal levels of gene products. In *D. melanogaster,* the Y chromosome is male specific however the X chromosome has some reproductive genes as well as other housekeeping genes essential for basic cellular and developmental pathways [71]. Since the level of gene products produced on the X is the same in the two sexes, dosage compensation process kicks in to equalize levels of gene products. In *Drosophila*, dosage compensation of X-linked genes begins early in embryogenesis by responding to the number of X chromosomes in the nucleus early in embryonic development [72]. Four genes are involved in the regulation of dosage compensation in *Drosophila* males: maleless (mle), and male-specific lethal-1, -2, -3 (msl-1, msl-2, msl-3) [68, 73]. The four genes are termed as the MSL complex and are essential in elevating transcription of the X chromosomes in males. This complex uses histone acetyltransferase MOF to cause a global hyperacetylation of the X-linked chromatin in order to increase gene expression by hypertranscription of genes in the

single X in males [73]. In females, the presence of two X chromosomes represses translation of the MSL preventing inappropriate dosage compensation in females. Another chromosome specific protein within *Drosophila*, *Painting of fourth (Pof),* has also been found to be involved in dosage compensation within the *Drosophila* lineage. It is chromosome specific to the dot chromosome (Muller element F) which is the Dipteran ancestral sex chromosome but subsequently reverted to autosome in *Drosophila. Pof* in *D. melanogaster,* colocalizes with the dosage compensation protein msl-3, and in *D. busckii,* it mediates the hypertranscription of the entire X chromosome exclusively in males suggesting a relationship to dosage compensation [74, 75]. A *Pof* ortholog in *L. cuprina*, the gene *no blokes* (*nbl*), was found to be important for X chromosome dosage compensation in *L. cuprina* as it is required for male viability and normal levels of gene expression of most X-linked genes. In humans, where females have two copies of a gene-rich X chromosome, and males have one X and a gene-poor Y chromosome, an imbalance in the expression of genes and can turn lethal if dosage compensation does not occur [76]. Dosage compensation involves silencing and inactivation of genes on one female X early in development [76]. The inactivation is then compensated by twofold up-regulation of genes on the active X chromosome maintaining a balanced dose of genes in both sexes.

1.3 Forces That Contribute to X and Y Differentiation

A number of molecular processes and changes occur during Y chromosome degeneration resulting into an accumulation of deleterious mutations which lead to gene loss and chromosome decay [9]. Some population genetic models have been proposed to predict decay via a variety of mechanisms, most of which only operate in nonrecombining regions of the chromosome [77, 78]. Some of these models are Muller's ratchet, background selection, and genetic hitchhiking.

1.3.1 Muller's Ratchet

Muller's ratchet has been proposed as a potential explanation for the degeneration of Y chromosomes [79, 80]. Some conditions which favor Muller's ratchet are the absence of recombination, presence of a finite population size, absence of back mutation, high mutation rates of slightly deleterious mutations and a weak purifying selection to remove new deleterious mutations [21, 81]. Many weakly deleterious mutations occur repeatedly in a population of

individuals. Muller's ratchet proposes that due to lack of recombination, the accumulation of weakly harmful DNA polymorphisms in a population over many generations can lead to the extinction of the species [82, 83]. This would be prominent in asexual populations which do not undergo recombination and are doomed to accumulate deleterious mutations due to genetic drift and mutation which accompanies a small population [82, 84]. Muller's ratchet has been used to explain the process of degeneration on the Y chromosome in *D. miranda* [21]. It suggests that a non-recombining, gene-rich Y chromosome rapidly degenerates in the initial stages but eventually slows down in gene decay over time after a threshold number of Y-linked genes is reached. Empirical evidence from *D. miranda* and its Y chromosome gene content seems to show consistency with this model [77].

1.3.2 Background Selection

Background selection is another mechanism which leads to the accumulation of mildly deleterious mutations [15]. In this model, only Y chromosomes which are free of strongly deleterious mutations survive and contribute to future generations [85]. In the absence of recombination, Y chromosomes with strongly deleterious mutations are doomed and do not persist in the population. Once the deleterious mutations are removed, the population size of the Y is greatly reduced due to a lack of recombination. A reduced population size therefore reduces variation in the population and increases the rate of fixation of the mildly deleterious mutations, upon which over evolutionary time, leads to the degeneration of the Y chromosome [15, 85]. Background selection depends on high rates of deleterious mutations and predicts that regions of high recombination will preserve more variation. Background selection has been used to explain the existence of reduced variability on the Y chromosome and the dot chromosome (fourth chromosome) in *Drosophila* [86].

1.3.3 Genetic Hitchhiking

Genetic hitchhiking is the process by which a gene or a mutation may increase in frequency by virtue of being linked to a gene which is positively selected [87]. In X chromosomes, where recombination is present, beneficial alleles can be fixed without dragging deleterious mutations. However, on the Y chromosome, which is non-recombining, fixation of beneficial mutations

simultaneously fixes deleterious mutations which are linked. One of the factors that lead to gene decay and ultimately degeneration of the Y chromosome is the presence of deleterious mutations. Deleterious mutations on the Y chromosome can be dragged to fixation and increase in frequency if they are linked with other beneficial genes [88], for example, a male-determining locus. The hitchhiking effect is more pronounced if it is in combination with the Muller's ratchet mechanism [88].

1.4 Sex Chromosomes in Insects

Insects are diverse and ubiquitous – they can be found in both terrestrial and freshwater habitats [89]. Resultantly, their sex chromosome system varies considerably, as evidenced from a study of 37 species in the order Diptera which showed 12 distinct sex chromosomes configurations despite a relatively homogenous karyotype [1]. Most insects reproduce sexually and are gonochoristic – they are either male or female throughout their life [37]. However, different sex chromosomal systems are in existence in insects to differentiate the two sexes. For example, most insects of the order Diptera exhibit the most familiar form of sex chromosome system the XX/XY sex chromosome system [90, 91] where the male is heterogametic XY and female homogametic XX. (for example, the common house fly *Musca domestica* [92], species within the genus *Drosophila* [93, 94]). In contrast, most insects in the order Lepidoptera (for example moths and butterflies[7, 94]) exhibit the ZZ/ZW sex determining pathway with the female being heterogametic ZW and the male homogametic ZZ. Interestingly, the ZW system has also been observed in the family Tephritidae within Diptera however not much is known about their sex chromosome evolution [1, 90]. Others, such as grasshoppers, locusts [95, 96] (order Orthoptera), and cockroaches [97] (order Blatella), an XO sex chromosome system is observed in some species where the original Y chromosome of the males is completely lost and males carry a single X while females are XX [90]. Therefore, this diversity observed in insects provides a good model system for the study of sex chromosomes evolution and can be used as a good indicator of understanding the fluidity observed in genome evolution.

1.4.1 Sex Chromosomes in Diptera

Diptera are among the most diverse insect orders. They are ubiquitous and can be found in nearly every habitat and all continents. They have approximately 150,000 described species. Diptera flies are generally divided into two groups – lower Diptera (Nematocera) and higher Diptera (suborder Brachycera). Brachycera includes *Drosophila* species and most families of Diptera [1]**.** Most of these species appear to lack chiasmata in males and a majority of the species possess the XX/XY sex chromosome system. However, some diverge from this 'norm'. For example, some species within the family Tephritidae (which is one of two families referred to as fruit flies, the other being Drosophilidae) have evolved the ZZ/ZW sex chromosome systems diverging from the common XX/XY sex chromosome system [1, 90].

The level of sex chromosome differentiation varies widely among species within Diptera [1, 98]. The most common sex system is the presence of heteromorphic sex chromosome [99], however, several lineages within Diptera lack heteromorphic sex chromosomes and possess homomorphic sex chromosomes [1]. A good example is of *Megaselia scalaris* (laboratory fly) of the family Phoridae, whose X and Y sex chromosome cannot be differentiated under the microscope [100]. The sex chromosomes of some species within the family Calliphoridae (*Chrysoma rufifacies, Chrysomya albiceps, Calliphora erythrocephala)* also exhibit homomorphic characteristics [1, 59, 101, 102].

Sex chromosomes play a role in sex determination in various ways within Diptera. In some Dipteran species exhibiting heteromorphic sex chromosomes, sex determination is usually under the control of a dominant male determiner (M) which is fixed and located on the heteromorphic pair (the Y chromosome) [98]. Its presence on the Y chromosome represses female development and promotes the male phenotype, for example in the Mediterranean fruit fly *Ceratitis capitata*[103] , and the blow fly *Lucilia cuprina* [104]. In other Dipteran species, the dominant male determiner (M) behaves like a transposable element and is variously located in different chromosomes (linkage groups). The species *M. scalaris*, displays this system. Both of its sexes carry 3 homomorphic chromosome pairs, however the dominant male determiner M behaves like a transposable element and moves between the three chromosomes [100]. The presence of the M factor within a chromosome therefore blocks female development pathway. *Musca domestica* is another species with a unique sex system. In a standard strain, the male determiner is carried on the Y chromosomes of a heteromorphic pair [105] while other strains the M is located on the other

five autosomes as well as the X chromosome [105, 106]. In the autosomal strains, both sexes exhibit homomorphic sex chromosomes. Other Dipteran species determine sex by the ratio of the X sex chromosomes to the autosome for example in the *Drosophila* genus [107]. In *D. melanogaster*, sex is determined by the dose of the X-linked gene *sex lethal* (*Sxl*) whereby diploid XX embryos develop into females and haploid XY embryos develop into males [1, 107].

1.4.2 Ancestral Sex Chromosome in Diptera (Muller Element F)

Cytogenetic studies show that chromosomal gene elements among Diptera are highly conserved [108] and represent the basic ancestral karyotype of Diptera. The basic ancestral karyotype in most higher Diptera is composed of six pairs of chromosomes, five large euchromatic 'rod' chromosomes which are the autosomal pairs, and one small 'dot' chromosome which is the sex chromosome [1]. In *Drosophila*, the autosomal chromosomes contain approximately 2000 genes while the smaller heterochromatic dot contains ~100 genes [109, 110]. The chromosomal gene elements of each of the chromosomes (the five large rods and one small dot) are named Muller elements A through E and F, respectively [1, 108, 110]. Numerous studies have shown that Muller element F, which is the dot chromosome or the fourth chromosome in *Drosophila melanogaster* is the ancestral sex chromosome in Diptera [1, 108, 110, 111]. It has been sex-linked for over 200 million years (MY) of evolution in many higher fly families [1, 108]. However, within *D. melanogaster*, Muller element A segregates as the X chromosome [107]. Muller element F reverted back to autosomal in *D. melanogaster*, but it still features characteristics which are similar to an X chromosome. Increased dosage of the dot chromosome shifts intersex individuals in *D. melanogaster* towards female development Additionally, genes located on Muller F show a higher expression in female embryos during early development as compared to males, a characteristic similar to X-linked genes [107]. In some other Dipteran species, additional chromosomal elements became assimilated into the ancestral sex chromosome. *Glossina morsitans* (tse tse fly) is one example. The X chromosome of *G. morsitans* resulted from the incorporation of two additional chromosomal elements (elements A and D) to the ancestral X chromosome (element F) [1]. A similar event is seen in the robber fly (*Holcocephala fusca*) where part of muller element B is added to the ancestral sex chromosome Muller F.

1.4.3 Is Muller Element F X-linked in all Dipteran Species?

In most species belonging to higher Diptera families, most of the genes located on Muller element F – which maps as the X chromosome – are X-linked [1, 112]. This was determined via a whole-genome sequencing study which was performed to identify X-linked genes in several species of Diptera families [1]. Many of the genes characterized as X-linked were located on Muller element F indicating that element F is part of a heteromorphic sex chromosome pair in several Brachycera with a degenerate Y chromosome, supporting the hypothesis that it is the ancestral X chromosome in higher Diptera [1]. However not all genes characterized as belonging to the ancestral sex chromosome are X-linked. Using sequencing technology, an X-linked gene would have twice as much DNA in the female as compared to the male. In contrast, the autosomal genes would have the same copy number in both male and female DNA. Some of the homologous genes in *Lucilia cuprina* which are characterized as Muller element F in *Drosophila* are X-linked [112, 113]. Others were found to exhibit autosomal read coverage properties suggesting that these genes may have reverted back to autosomes [112]. Insects with homomorphic sex chromosomes did not show any X-linked properties within Muller element F [114]. For example, in the blow fly species *Calliphora erythrocephala* the ancestral sex chromosome (Muller F) is not X-linked which supports the hypothesis that it possesses homomorphic sex chromosomes [1].

1.5 Sex Chromosomes in Insects

Sex chromosome systems vary greatly in phylogenetically closely-related species and even within a species suggesting a rapid evolutionary rate [26], a phenomenon also observed within blow flies. It is hypothesized that the rapid diversification of the Calliphoridae family may be due to peculiar geological changes and evolutionary processes which created new niches that triggered an adaptive radiation [115]. The last common ancestor for blow flies (Calliphoridae) is estimated to be around 22 mya which was followed by a rapid radiation of the subfamily Chrysomyniae (~17 mya) and Luciliinae sister-lineages (~16 mya) [115]. A diverse presence of sex chromosome systems is displayed within the family Calliphoridae. Similar to other insects within Brachycera, Calliphoridae karyotype has been shown to be highly conserved, and composed of five autosomal pairs of chromosomes and one pair of sex chromosomes [1, 104, 116-119], with Muller element F as its ancestral sex chromosome [120].

Several blow fly species in the subfamily Chrysomyinae such as *Cochliomyia hominivorax, Cochliomyia macellaria,, Protophormia terranova*, *Phormia regina*, *Chrysomya megacephala* [59, 117, 119] display a heteromorphic sex chromosome with a male heterogametic sex chromosome system where the male is XY (with a degenerated and differentiated Y chromosome); and the female is XX. Sex development in most blowflies is controlled via a dominant male-determining factor on the Y chromosome [1, 104, 121]. In *L. cuprina*, sex is determined by a Y-linked male-determining gene (*M*) which is centered near the Y chromosome centromere [104, 116]. *Chrysomya rufifacies* and its sister species *Chrysomya albiceps*, despite belonging to the same subfamily Chrysomyinae as the blow flies listed above, exhibits homomorphic sex chromosomes with an X and Y which appear identical in morphology [59, 102, 122]. Furthermore, they have two types of females: thelygenic females (produces only female offspring) and arrhenogenic females (produces only male offspring). Sex determination of their offspring is independent of environmental factors such as temperature or diet [123, 124] and is determined by the maternal genotype. Mating studies hypothesize that thelygenic females are heterozygous for a dominant female-determiner (*F/f*) while both arrhenogenic females and males are homozygous (*f/f*) at this same locus [123]. The nature of this *f* gene is unknown [121]. Restriction of recombination in blow fly species with heteromorphic chromosomes (such as *Cochliomyia. macellaria*, *L. cuprina* and *Phormia regina*) may have taken place so that sexually antagonistic alleles are restricted to one sex. As for those that display homomorphic sex chromosomes (e.g *C. rufifacies*), a different evolutionary route may have been taken to minimize the deleterious effects introduced by sexually antagonistic variants thereby preserving recombination. This proposes that sexually antagonistic variants are thus confined to the sex it benefits [23].

1.6 Research Aims

Cytological karyotypes show a highly conserved blow fly karyotype among species. However, it also shows the presence of both homomorphic and heteromorphic sex chromosomes within the same family of blow flies suggesting the occurrence of a rapid evolutionary event in species within the same family. Apart from a few studies on the sex chromosome of the blow fly *Lucilia cuprina,* sex chromosome evolution in blow flies remains relatively understudied and has not been fully explored. Therefore, the goal of my dissertation is to characterize the evolutionary forces which contribute to the diversification of sex chromosomes in blow flies, by analyzing their genomic

composition. The results obtained herein will reinforce the understanding of sex chromosome evolution within the Calliphoridae family as well as in other dipteran species.

A comparative genomic approach is utilized to characterize and investigate the sex chromosomes of four blow fly species – *Phormia regina* and *Lucilia cuprina* which possess the heteromorphic system, and *Chrysomya rufifacies* and *Chrysomya albiceps* with the homomorphic system and monogenic females. The first chapter is a general outline of the process of sex chromosome evolution describing how evolutionarily labile sex chromosomes are even within species of the same family. The second chapter presents genomic sequences and assembled genomes of the three sex types observed within the blow fly *Chrysomya rufifacies* (male, arrhenogenic female, and thelygenic female). The third chapter provides genomic composition and annotation of the putative sex chromosomes of the four blow fly species, and a characterization of the dipteran ancestral sex chromosome (Muller F). Lastly, the fourth chapter provides evolutionary inference on the forces that contribute to the divergence and differentiation of sex chromosomes in blow fly species.

CHAPTER 2. THE GENOMES OF A MONOGENIC FLY: VIEWS OF PRIMITIVE SEX CHROMOSOMES

The work presented in this chapter was from a collaborative effort with other authors and has been submitted for publication (AA Andere, ML Pimsler, AM Tarone, CJ Picard. The genomes of a monogenic fly: Views of primitive sex chromosomes. *Under review*). My contribution to this study involved performing all the bioinformatic analysis, generation of figures and tables, writing the initial drafts of the manuscript and contributing to subsequent editing thereafter.

2.1 Abstract

The production of male and female offspring is often determined by the presence of specific sex chromosomes which control sex-specific expression, and sex chromosomes evolve through reduced recombination and specialized gene content. Here we present the genomes of *Chrysomya rufifacies*, a monogenic blow fly (females produce female or male offspring, exclusively) by separately sequencing and assembling each type of female and the male. The genomes (>25X coverage) do not appear to have any sex-linked Muller F elements (typical for many Diptera) and exhibit little differentiation between groups supporting the morphological assessments of *C. rufifacies* homomorphic chromosomes. Males in this species are associated with a unimodal coverage distribution while females exhibit bimodal coverage distributions, suggesting a potential difference in genomic architecture. The presence of the individual-sex draft genomes herein provides new clues regarding the origination and evolution of the diverse sex-determining mechanisms observed within Diptera. Additional genomic analysis of sex chromosomes and sex-determining genes of other blow flies will allow a refined evolutionary understanding of how flies with a typical X/Y heterogametic amphogeny (male and female offspring in similar ratios) sex determination systems evolved into one with a dominant factor that results in single sex progeny in a chromosomally monomorphic system.

2.2 Introduction

Animals and plants exhibit typical patterns of sex chromosomes evolution in heteromorphic chromosomal systems [38]. An autosome first begins to differentiate following the

acquisition of a sex-determining locus and this differentiation is maintained via reduced recombination. This can lead to initial expansion and eventual degeneration of the Y chromosome in X/Y systems, and similar processes happening in Z/W systems [14, 38, 125-128]. Evolutionary theory postulates that differentiated sex chromosomes trace their ancestry to an undifferentiated autosomal pair where one of the autosomal homologs acquired a sex-determining gene and consequently sexually antagonistic mutations arose causing reduced or eliminated recombination between the pair [9, 12]. Restricted recombination led to the emergence of the sex-limited chromosome, in this case the Y chromosome. The newly evolved sex chromosomes therefore diverged functionally and morphologically resulting in heteromorphic chromosomes [9, 11, 12]. In general, Y chromosomes contain very little genic material and the chromosome is mostly heterochromatic, typically due to the result of mutations, insertions and deletions, and transposable element activity. Of course, with every rule come the exceptions. In Diptera, the model species *Drosophila melanogaster* has heteromorphic sex chromosomes, however the ancestral Dipteran sex chromosome thought to be the dot or 4th chromosome, is an autosome in *D. melanogaster*. Furthermore, the *D. melanogaster* mode of sex determination does not depend on the presence of a male-determining locus on the Y chromosome, but rather dosage differences of genes on the X chromosome results in alternatively spliced transcripts driving the development towards either a male or female fate. Furthermore, fundamental differences in sex determination processes vary across Diptera (for a review see: [129]). For example, the mosquito *Aedes aegypti*, the house fly (*Musca domestica*) and Mediterranean fruit fly (*Ceratitis capitata*) all harbor a male determining factor present on the Y chromosome, following typical, non-Drosophilid tradition [130]. In contrast, sex determination in sciarid flies, such as *Sciara ocellaris*, relies upon dosage compensation affected by temperature-dependent paternally donated X-chromosome destruction [131, 132].

Signatures in the genome left behind from multiple evolutionary events can be used to decode the mystery of sex-determining systems in many living organisms [1, 133-135]. Transitions of sex determination mechanisms have been found to be frequent in nature among species which display homomorphic sex chromosomes in both sexes [136]. For example, in amphibians and reptilians the turnover rate of sex-determining genes and sex chromosomes is high. Approximately 96% of amphibian species possess homomorphic sex chromosomes with a sex-determining gene that is easily and rapidly replaced by another gene of a different chromosome across their

phylogeny [137-139]. Epigenetic and environmental factors such as temperature can also play a role in sex determination [140]. In comparison, species with heteromorphic sex chromosomes (XY and ZW systems) are presumed to be highly differentiated and have reached an evolutionary end point with the sex-determining gene in the sex chromosome limited sex [9, 58].

In Calyptrate flies, the most common sex chromosome system is the XX/XY system [1, 90] with a homogametic female XX and a heterogametic male XY. Heteromorphic sex chromosomes are observed in a majority of blow fly species (Diptera: Calliphoridae), with highly differentiated X and Y sex chromosomes in both morphology and sequence, and a Y-linked male-determining factor M [1, 59, 99, 121, 141]. In some *Lucilia* species the difference in genome sizes between the sexes can be > 50 Mb, representing > 7% of female genomic content [142]. However, the blow fly *Chrysomya rufifacies*, and its closest sister species *Chrysomya albiceps*, employ homomorphic (undifferentiated) sex chromosomes and both sexes have the same size genomes [59, 101, 143-145]. Furthermore, in these monogenic species, the females produce either all female or all male progeny [145-147] (Figure 2.1) — a divergence from the heteromorphic (differentiated) and amphogenic sex chromosome system observed in other Calliphoridae [59, 148]. The genetic basis of monogeny in *C. rufifacies* has been hypothesized from mating studies, ovary and pole-cell transplantation and patterns of protein expression [102, 122, 146, 147]: female producing flies (thelygenic females) are heterozygous for a dominant female-determiner (F/f) with predetermined sex-determining properties while the male producing females (arrhenogenic females) and males are homozygous for the recessive allele (f/f) at this same locus. Sex determination in *C. rufifacies* is largely genetic and independent of environmental factors such as diet, season and temperature [145]. However, the molecular nature of the primary sex-determining gene(s) or locus in *C. rufifacies* remains unknown.

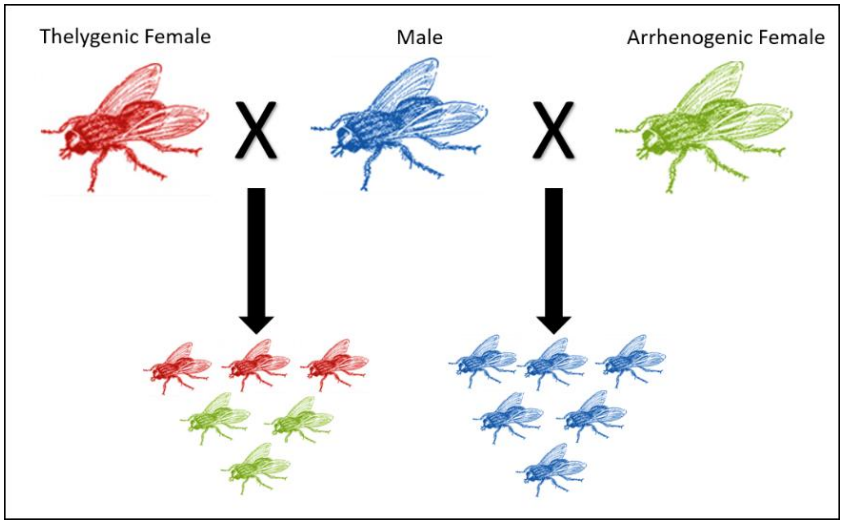


Figure 2.1. Sex determination of *C. rufifacies* offspring is determined by the maternal genotype. Thelygenic females produce only female offpsring while arrhenogenic females produce only male offspring.

In this study, we present the genomic sequences and the assembled genomes of male, thelygenic female, and arrhenogenic female *C. rufifacies* for the first time. We characterize putative sex chromosomes and document candidate sequences which belong to the dipteran ancestral sex chromosome (Muller F). We also show genomic evidence that these putative sex chromosomes appear to be undifferentiated, unless differentiation occurs through copy number. These results will allow for a greater depth of evolutionary study on sex chromosomes across the Calliphorid species and give insight into the unique sex-determining mechanism of a monogenic fly.

2.3 Methods

2.3.1 DNA Library Preparation and Sequencing

Pooled genomic DNA was extracted from the heads of five male-producing females (arrhenogenic), five female-producing females (thelygenic) and five male flies originating from a lab colony of *Chrysomya rufifacies* (see [37] for colony foundation, maintenance, and sample collection procedures) using the DNeasy Blood and Tissue DNA Extraction kit following the manufacturer's instructions (Qiagen Inc., Valencia, CA, USA). Each extract was quantified using a Qubit fluorimeter (Thermo Fisher Scientific, Waltham, MA, USA) so that a total of 1 µg of

genomic DNA was sent to a facility for library preparation. Libraries (N = 3) were constructed following the TruSeq DNA Sample Preparation Guide by Illumina (Catalog #PE-940-2001. Part # 15005180 Rev. A, November 2010). Sequencing was performed on the three paired-end libraries using the Illumina HiSeq2000 sequencing platform (Illumina Inc, San Diego, CA, USA) with a read length of 2 x 100 bp. Both of the library preparation and sequencing was completed by the Purdue University Genomics Core Facility (West Lafayette, IN, USA). The three libraries were multiplexed on a single lane. All sequencing data produced in this study have been deposited in the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) and can be accessed under the BioProject ID PRJNA575047 and SRA accession number SRP238163.

2.3.2 Pre-processing and Quality Trimming

Raw reads were trimmed to eliminate low quality reads (Phred score < 20) and adapter sequences. On a per-library basis, overlapping pairs of reads were merged into a single sequence read creating longer and higher quality reads. Mismatch cost was set to 2, gap cost was set to 3, and the minimum score required for an alignment to be accepted for merging was set to 8. Both read trimming and merging were analyzed using the software CLC Genomics Workbench (CLC-GWB v9) (Qiagen Inc.). Extraneous or contaminating DNA were filtered out by mapping the merged and trimmed reads to 3,006 phage (www.phantome.org, v2016-04-01) and 49,290 bacterial genomes (www.ncbi.nlm.nih.org, downloaded on 05/2016 and 03/2017). Mitochondrial reads were subsequently removed by mapping the reads on to the mitochondrial genome of *C. rufifacies* (NC_019634.1). The resulting unmapped reads were thereafter used in the *de novo* assembly step.

2.3.3 Genome Assemblies, Scaffolding and Evaluation

De novo genome assemblies were performed on each of the three processed and quality filtered libraries (male, arrhenogenic female and thelygenic female) using the CLC-GWB v9 assembler. Several iterations of the de novo assemblies were carried out with k-mer sizes ranging between 24 – 50 nucleotide, and bubble sizes ranging from 100 – 1000; with the intention of selecting the ideal assembly with optimal parameters to be used in downstream analysis. Optimal k-mer sizes for all three sets of libraries was determined to be 32 bp. Additionally a transcriptome

of the thelygenic female was also assembled for scaffolding purposes only, using a k-mer size of 32 bp. For all the assemblies, a mismatch cost of 2, insertion cost of 3 and deletion cost of 3 was selected. Mapping parameters were set such that 50% of each read needed to have at least 90% identity to be included in the final mapping. Contigs from each of the three assembled draft genomes were scaffolded with the assistance of the assembled thelygenic transcriptome using the scaffolding program LRNA scaffolder [149]. This program uses transcriptome contigs to orient and combine genomic fragments. Calculations of the assembly statistics was done by CLC-GWB v9 and the genome assessment tool QUAST v3.1 [150]. Coverage mapping and subsequent variant detection was done by mapping reads to the assembled genomes ignoring positions with coverage >100,000 and ignore broken paired reads. Data were visualized using Microsoft Excel using frequency distributions. Universal single copy orthologs (USCOs) was used to assess completeness and contiguity of the assembled genomes using the Benchmarking Universal Single-Copy Orthologs (BUSCO) v2.0.1 [151]. BUSCO measures the fraction of genes highly conserved in related species by mapping and identifying them using a database of orthologs (OrthoDB) from eukaryotes, diptera, arthropods and insects.

2.3.4 Gene Prediction, Annotation and Ontology

Ab initio prediction of gene and protein sequences for each of the three sex types was performed by the gene predicting program Maker [152] on the three draft genomes. The flag option 'always_complete' in the maker_opts.ctl file was set to 1, the rest of the parameter were left at default settings. To infer gene predictions, expressed sequence tag (EST) evidence for gene transcription was obtained from the assembled thelygenic transcriptome and alternate EST evidence from *D. melanogaster* gene sequences (GCF_000001215.4_Release_6_plus_ISO1_MT_rna). Additional evidence was obtained from protein sequences of *L. cuprina* (GCA_001187945.1_ASM118794v1_protein), *D. melanogaster* (GCF_000001215.4_Release_6_plus_ISO1_MT_protein) and arrhenogenic female protein sequence (from a previous gene prediction run not published). Gene sequences which encoded peptide sequences ≥30 amino acids in length were filtered and preserved. RNA-seq reads from the thelygenic female (accession number SRX149675) were mapped onto the gene sequences predicted for each of the three sex types following the same mapping parameters used in the genome assembly process. Annotation was performed using a non-redundant arthropoda protein

BLAST database (BLASTp v2.2.28+) with an E-value cutoff of ≤ 1E-5 [77]. Functional categorization of the BLAST results was conducted using Gene Ontology (GO) via Blast2GO v3.3.5 [153] and summarized at level 2 into the 3 main GO categories – biological processes, cellular component and molecular function, using the number of sequences in each category for the two females and the male. KEGG map pathways [154] were extracted from the annotated protein sequences using default parameters in Blast2GO v3.3.5. The web platform OrthoVenn [155] was used to identify overlap among orthologous clusters from the predicted protein sequences of the two females and the males in a genome wide perspective. The predicted protein sequences for the thelygenic female, arrhenogenic female and the male were uploaded onto OrthoVenn independently in fasta format and default parameters were used to run the analysis. Orthologous clusters that were unique to each sex type, shared between the two females, shared between each of the females and the male, and common in all three were grouped together. The cluster classification was done according to sequence analysis data, protein similarity comparisons, and phylogenetic relationships [155]. OrthoVenn deduced the putative function of each orthologous cluster by performing a protein BLAST search against a non-redundant protein database in UniProt. Top hits with an e-value of <1E-5 were defined as the putative function of each cluster [155].

2.3.5 Sex Chromosome Characterization

Putative X and Y chromosome sequences were characterized using the chromosome quotient approach [156] which utilizes read coverage ratios of alignment to differentiate X, Y and autosomal sequences. The chromosome quotient program [156] was used to align male and female reads onto each other's genome (male reads independently mapped to male genome and to each of the female genomes, and vice versa). A stringent aligning criterion requiring a whole read to map onto the reference contigs with zero mismatch was done in order to reduce the number of false positives that may be caused by the highly repetitive sequences from Y chromosomes with closely related sequences on the autosomal or X chromosomes due to duplication events [26, 156]. Chromosome quotients were calculated by comparing the number of alignments from female sequence data to male sequence data. Ideally, putative X sequences were expected to have a CQ ratio of 2 with X sequences characterized as those with twice as many female reads aligned as male, while putative Y sequences a CQ ratio of 0. Due to the presence of the two types of females

(thelygenic and arrhenogenic), the CQ approach was implemented on each female independently resulting in two sets of X and Y sequences. Male contig sequences with a CQ of less than 0.3X were grouped as putative Y chromosomes to accommodate repetitive Y sequences that may be present in both the male and female. A total of 2,195 contigs (~2 Mb from male and arrhenogenic female comparison) and 4,031 contigs (~4 Mb from male and thelygenic female comparison) were identified as putative Y chromosomal sequences (Table 2.3). The two predicted sets of putative Y sequences were compared to determine the proportion of overlap shared between them. Female contig sequences with a CQ ranging between 1.6X and 2.5X were grouped as putative X sequences. This CQ interval was selected to reduce false positives. A total of 23,624 contigs (~64 Mb) and 7,448 contigs (~15 Mb) from the arrhenogenic and thelygenic female respectively, were categorized as putative X chromosomes. A comparative analysis of both sets of putative X chromosomes was performed by CD-HIT-2D-EST v4.5.6 [157, 158], to isolate a representative set of *C. rufifacies* chromosome X sequences characterized by both females, using a length difference cutoff and a sequence identity cutoff both of 80%. A nucleotide BLAST (BLASTn v2.6.0+, E-value cutoff ≤ 1E-5) was performed on the characterized sex chromosome sequences using a non-redundant nucleotide database [159]. Resulting BLAST results were functionally characterized using default parameters on Blast2GO v5.1.13 [153] and gene ontology (GO) terms assigned to the BLAST results. The functional categories were simplified using the GO slim functionality in Blast2GO and enrichment analysis using Fisher's exact test performed on them. The enriched GO terms and their corresponding FDR values were summarized and categorized to the three GO domains: biological processes, cellular component and molecular functions; and visualized using default settings of the REViGO web server [160].

2.3.6 X-linked Muller Elements

Coding sequences of the chromosomal gene contents (Muller elements A-F) from *Drosophila melanogaster* were downloaded from GenBank. The longest isoforms were selected for each gene resulting to a total of 10,488 coding sequences. They were thereafter queried against the assembled genomes of the male and the two females using a translated nucleotide and database (tBLASTx v2.6.0+, E-value cutoff ≤ 1E-5) to identify orthologous contig sequences within the genomes. Orthologous contig sequences were assigned as belonging to the respective Muller elements they segregated with. To determine which Muller elements were X-linked in *C. rufifacies*,

male and female sequence reads were aligned to the identified orthologous contig sequences using the CLC-GWB v9 read mapper, and the read coverages compared. To reduce false positives, stringent mapping parameters were used such that 100% of each read needed to have at least 80% identity to be included in the final mapping. The program DESeq [161] was used to identify any differential read coverages observed within the orthologous Muller elements to identify sequences with a twofold higher abundance in females than males, by calculating a Log2(M/F) coverage ratio. Contig sequences with a Log2 (M/F) coverage ratio within the range of -0.6 and -1.3 were considered to be X-linked.

2.3.7 Repeat Sequence Analysis

A library of all known Diptera repetitive elements was used to identify repetitive elements in each of the 3 genomes and the putatively characterized X and Y chromosomes using the program RepeatMasker v4.0.7 in default mode [162].

2.4 Results and Discussion

2.4.1 Sequencing and *De novo G*enome Assembly

Three separate genomes (male: M, thelygenic female: TF, and arrhenogenic female: AF) were paired-end sequenced resulting in an average read length of 100 bp and average quality score of 37 following adapter sequence trimming, low quality read filtering and overlapping pairs merged. Approximately 0.07% (M), 0.06% (TF) and 0.11% (AF) of reads were removed as they were identified as either non-fly or mitochondrial reads, resulting in 8.5 X 10^7 (M), 1.02 X 10^8 (TF), and 1.34 X 10^8 (AF) high-quality reads used to assemble three genomes. Initial draft genomes were further scaffolded using the TF *C. rufifacies* transcriptome as a guide [163], resulting in 107,111 TF contigs; 114,048 AF contigs; and 109,341 M contigs (assembled metrics summarized in Table 2.1). To evaluate single-base accuracy of the assembled genomes and determine the percentage of reads used in the contig construction, each set of the processed reads were mapped back to their respective assembled genome. Approximately 95% of the reads from each sex type mapped back to the contigs with an average coverage range of 27 – 42X reads suggesting that most of the reads were utilized in the genome construction (Table 2.1). From a set of 1,066 Arthropoda and 1,658 Insecta single copy gene orthologs, approximately 93% and 91%, respectively, were

present in the three draft genomes (Table 2.1, Table A1). Notably, the assemblies were smaller in size than expected [142]; however, read mappings and BUSCO results signify largely complete and high quality (albeit fragmented) genome assemblies. A complete BUSCO report is detailed in Table A1. The assembled genomes and raw reads have been deposited in GenBank and the SRA (BioProject ID PRJNA575047 and SRP238163, respectively).

Table 2.1. Summary of de novo genome assemblies of the AF, TF and M genomes, read mapping statistics, BUSCO completeness assessment results, number of predicted genes and the percent of repetitive elements detected in each genome

	Arrhenogenic Female (AF)	**Thelygenic Female (TF)**	**Male (M)**
No. of processed reads	134,541,815	102,695,597	85,597,908
N50 (bp)	4,101	3,889	4,164
Mean contig (bp)	2,588	2,606	2,638
No. of contigs	114,048	107,111	109,341
No. of mapped reads (%)	127,991,372 (95.13%)	98,025,293 (95.45%)	82,377,083 (96.24%)
Mean coverage	43X	34X	28X
Estimated genome size (Expected: 426 Mb [[164])	295,268,734 bases 295 Mb	279,238,173 bases 279 Mb	288,503,435 bases 289 Mb
BUSCO - complete*	993 (93.1%)	989 (92.8%)	994 (93.3%)
Repetitive elements %	6.84%	6.61%	6.89%
No. of predicted genes (mean length, bp)	13,910 3,345	13,590 3,271	13,798 3,233
Predicted genes with BLAST hits (%)	93.45%	94.42%	94.22%

** Arthropoda database of 1066 BUSCO groups. For full list, see Table A1.*

There are discordances between the expected genome sizes (424 Mbp, [142]) and the assembled genome sizes. This is not unusual if the genome has a large proportion of repetitive sequences or unsequenced heterochromatic regions. The sequenced-based estimate of the *Drosophila melanogaster*, a relatively small and repeat depauperate genome, have been supported through follow-up work [165]. Results diverge in species with larger genomes though; a previously assembled blow fly genome (*Phormia regina*, [148]) assembled close to the expected size (assembled larger at 550/534 Mbp vs. ~529/518 Mbp expected for the female and male respectively). In another example, the assembled genome size of *Lucilia cuprina* was 458 Mbp [166], smaller than the expected 665/568 Mbp for the female and male, however, a large proportion (57.8%) was attributed to the repetitive landscape of the genome.

For *Chrysomya rufifacies*, the expected genome sizes were the same for the two sexes at 425Mbp [142], yet the assembled genome sizes were 295, 279 and 289 Mbp for arrhenogenic

female, thelygenic female, and male, respectively. This amounts to roughly 150 Mbp of missing assembled genome. Such discrepancies are not uncommon when sequence- or molecular-based estimates are compared to cytometric estimates of genomes size. The genome of *Arabidopsis thaliana* was originally underestimated to be roughly 115 Mbp [167] vs. a revised/accepted genome size of 157 Mbp [168] based on flow cytometry. Similarly, the *Lucilia cuprina* genome size as estimated by genome sequencing was considerably smaller than the fly cytometry estimate [169]. A possible explanation would be the challenges of assembling a genome with a high proportion of repetitive elements. For example, GC content of the *Arabidopsis* genome sequence did not agree with the known GC content of the species, suggesting that GC rich repeat regions did not assemble well. However, this does not appear to be the case with *C. rufifacies*, as only <7% of the genome is attributed to the repetitive landscape (see results below). Another potential explanation for the discrepancies in genome sizes is large duplicated chromosomal segments [170, 171]. If a/some chromosome(s) has/have duplicated, one would expect to see parts of the genome containing twice as much coverage as the unduplicated portions. We generated frequency distributions of coverage across each genome and visualized this data in Figure 2.2 with data represented in Table 2.2. For both the female genomes, it was obvious there were two distributions of data from the genome, and visually inserting a coverage cutoff, each side of the distribution was analyzed for coverage statistics as well as for the number of variants (Table 2.2). When considering each side of the distribution, it is apparent that the right skewed distribution (> coverage levels) are roughly 2X the coverage of the left side. Considering duplication theory, if the left side represents 1X and the right 2X, the approximate genome sizes would be 469 Mbp and 434 Mbp for arrhenogenic females and thelygenic females respectively. Another potential explanation for this pattern may be if there is polyploidy or underreplication in the tissues used to produce the genomic sequence data (for a review, see [172]). This study used heads, which are typically considered to lack tissues with these features [173]. It is interesting to note that each sex/type exhibited a different pattern of major to minor peak heights, which may be a clue in deciphering the sex chromosomal dynamics of the species. The assembly results reported here indicate that clues to the sex determination system are largely limited to the non-repetitive portions of the genome in this study.

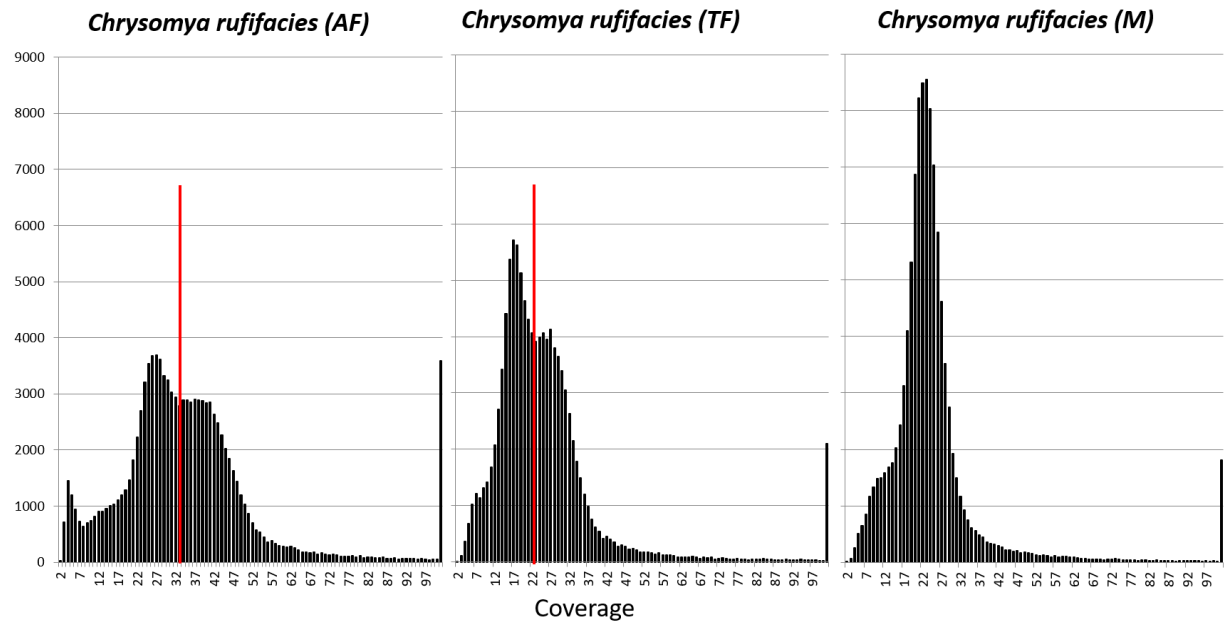


Figure 2.2. Coverage distributions for the different genomic assemblies with coverage (x-axis) vs. the number of assembled contigs at each coverage. There is a clear bimodal distribution of the main component of the coverage distribution in females and the different types of females exhibit different ratios of major and minor peak heights.

Table 2.2. Data associated with read mapping statistics for the distributions in Figure 2.2

	Cruf AF		*Cruf* TF		*Cruf* M
Mean Coverage (SE)	52.50 (2.14)		36.86 (1.81)		34.79 (1.28)
Median Coverage	32.07		21.80		22.17
Total # variants	2,826,180		2,924,285		2,198,938
Red line delineation	<32X	>32X	<23X	>23X	n/a
# contigs	54,772	55,176	60,304	50,965	109,341
Length of contigs (Mbp)	99	185	108	163	289
# reads mapped	25,506,380	129,383,978	19,531,466	78,785,565	85,597,908
# variants	1,194,952	2,631,228	838,346	3,085,939	2,198,938
# variants/1000bp	12.0	14.0	7.7	12.7	7.8
% SNV	81.0	78.6	81.7	78.9	79.6
% MNV	3.8	3.6	3.4	3.5	3.6
% Ins	6.7	8.0	6.3	7.8	7.4
% Del	7.8	8.9	7.9	9.0	8.6
% Replacement	0.8	0.9	0.6	0.8	0.8

2.4.2 Gene Prediction and Functional Annotation

A total of 13,590 TF, 13,910 AF and 13,798 M genes were predicted based on the nucleotide sequences, with approximately 80% of the predicted genes supported by TF RNA-seq reads (79.71% arrhenogenic female, 80.17% thelygenic female and 80.52% male). Approximately

94% of the predicted amino acid sequences demonstrated homology to arthropod sequences in GenBank with E-values less than 1E-5 (12,831 TF, 12,999 AF and 13,000 M genes). Therefore, approximately 6% of the predicted genes (759 TF, 911 AF and 798 M) were unique to *C. rufifacies.* As expected, calyptrate flies species were among the top hit species present in the BLASTp results for the three genomes, supporting phylogenetic relationships among these species. Included in the top hit species list were the sheep blow fly (*Lucilia cuprina*), the stable fly (*Stomoxys calcitrans*) the common house fly (*Musca domestica*), and the black blow fly (*Phormia regina*) [148]. In fact, the majority of the proteins resulted in hits to *L. cuprina*, *C. rufifacies*' most closely related species in the database (Figure B1). The number of protein sequences annotated to the different GO terms at hierarchical level 2 for all three GO categories is shown in Figure 2.3. GO analysis of the sequences with BLAST results show that the abundant GO biological processes were cellular process and metabolic process; within cellular component, cell and cell part were most frequent, while binding and catalytic activity were most prevalent molecular functions. The distribution and classification of the GO terms in each of the three categories is comparable to those of other dipteran species such as the blow fly *P. regina* [148], the common house fly *M. domestica* [92] and the fruit fly *D. melanogaster* [165].

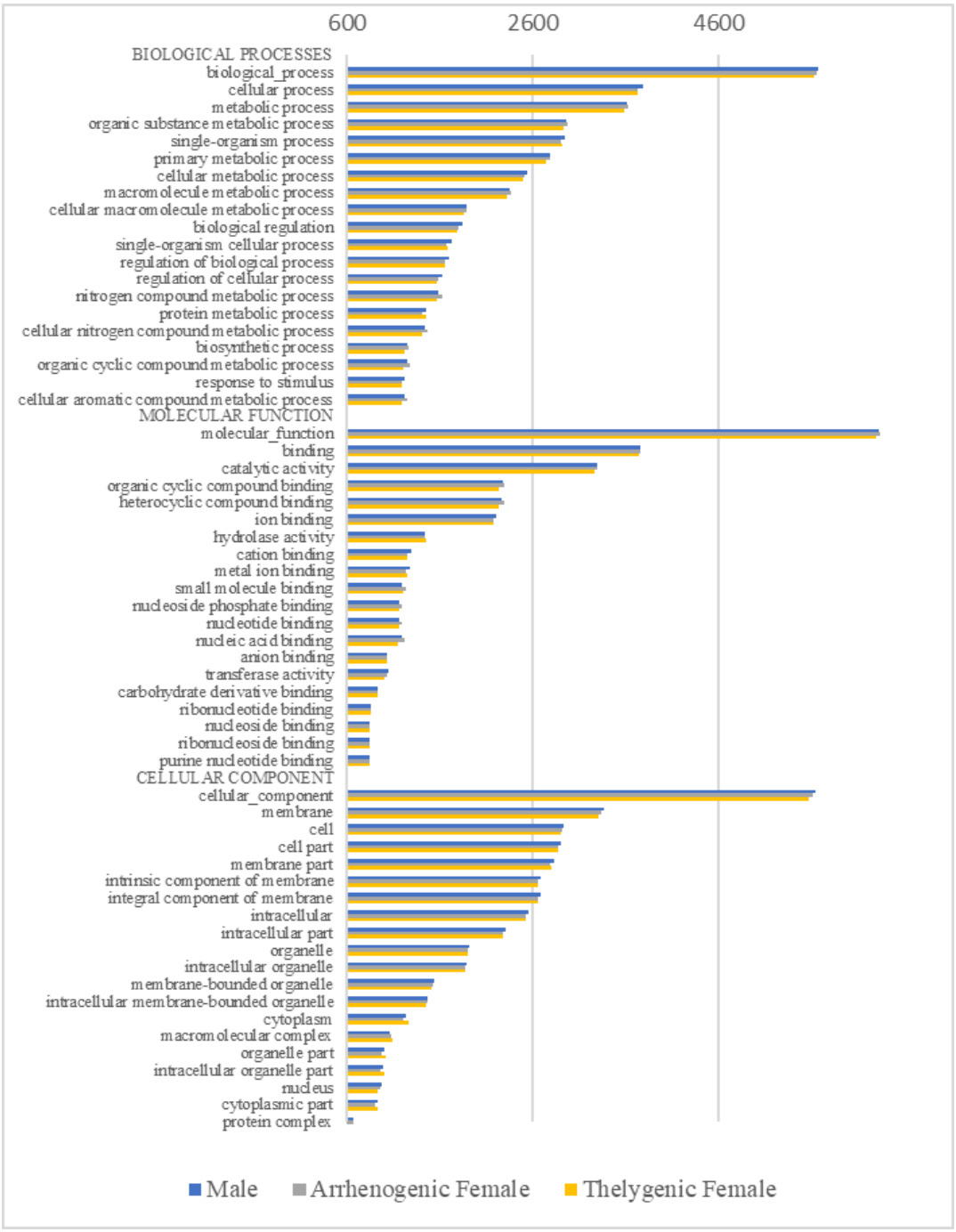


Figure 2.3. Functional categorization using GO terms (biological processes, molecular function and cellular components) from the predicted protein sequences in the thelygenic female, arrhenogenic female and the male *C. rufifacies*

The top KEGG pathways [154] in all the three sex types were purine metabolism, thiamine metabolism and biosynthesis of antibiotics (**Error! Reference source not found.**), a pattern also o bserved in the black blow fly [148].

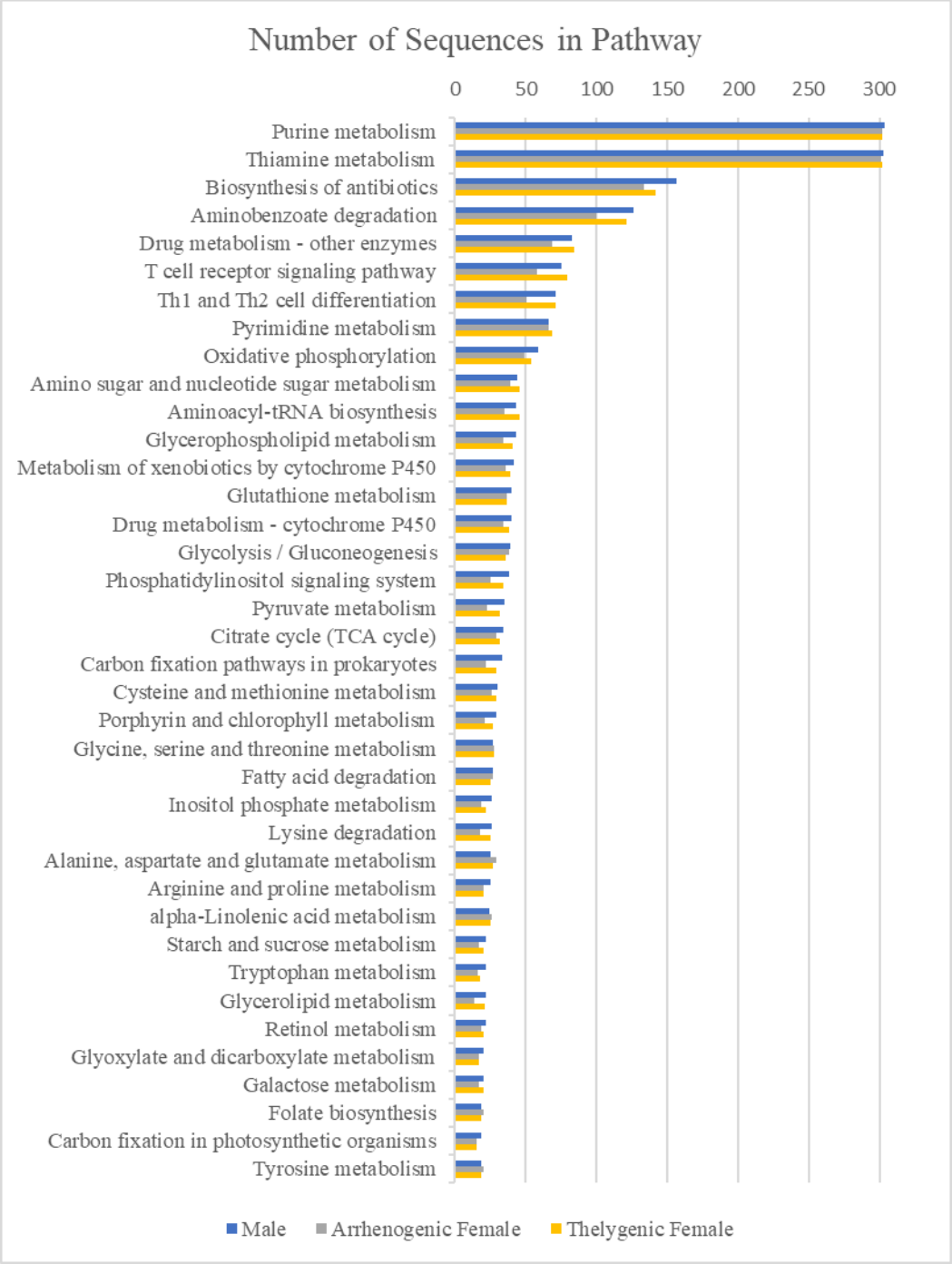


Figure 2.4. KEGG biological pathways of the thelygenic female, arrhenogenic female and the male extracted from the KEGG component in Blast2GO [153]

2.4.3 Comparative Analysis of Predicted Genes

Orthologous protein sequence clusters were identified and annotated using OrthoVenn [155] as seen in Figure 2.5. A total of 10,354 orthologous clusters were shared among the two females and the male totaling to 15,596 protein sequences shared among the three sexes with average lengths of ~425 amino acids/protein. Generally, paired groups shared similar clusters (AF-M: 732 clusters; TF-M: 774 clusters, and AF-TF: 644 clusters), with a small number of unique clusters (TF: 17 clusters, AF: 30 clusters, M: 20 clusters; Figure 2.5,Table A2). In all three genomes, the average lengths of the unique protein sequences were ~160 amino acids and are therefore are most likely sequencing and assembly artifacts. These unique clusters were analyzed for enriched GO-terms (p-value < 0.05;). Unsurprisingly, the shared orthologous protein sequences between the two females show five clusters annotated as yolk protein genes, which is described as the major yolk protein of eggs used as a food source during embryogenesis in *Drosophila* [174], and typically found on X chromosome in *Drosophila* [175]. Due to its absence in the male genome, it is possible that these genes are part of a region which has differentiated from the "Y" chromosome, or perhaps in a region that did not assemble well, though it is unclear if these are just linked to a causal factor or the causal factor themselves.

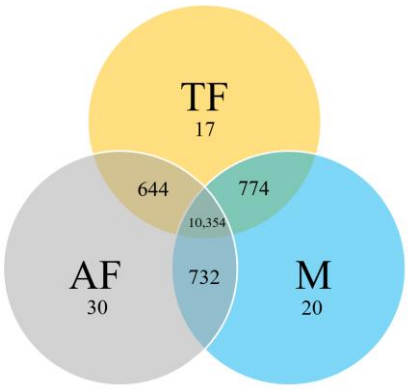


Figure 2.5. A Venn diagram produced by OrthoVenn [155] displaying the number of orthologous clusters of the predicted protein sequences (i) shared among the three sexes, (ii) shared between any two sexes and (iii) those uniquely found in each group. Cluster classification was done according to sequence analysis data, protein similarity comparisons, and phylogenetic relationships

2.4.4 Sex Chromosome Genomic Characterization

Using read coverage ratios (chromosome quotient, CQ) to compare the male and female genomes and their associated reads, it is possible to isolate genomic regions that are characterized as differentiated, such as would be the case with sex chromosomes [1, 26, 156]. Based on flow cytometry measurements of genome size differences in male and females (= no difference) [164], it was not expected that a large portion of the genomes would be isolated using the CQ approach. In fact, only ~3.3 Mb and ~1.5 Mb were isolated as putative X and Y chromosome sequences respectively (Table 2.3), suggesting largely undifferentiated sex chromosomes. Assuming the isolated genomic regions are a part of a differentiated region on putative sex chromosomes, their annotations via BLASTn hits (E-value cutoff ≤ 1E-5) resulted in 86% of the putative X sequences and 29% of the putative Y sequences being annotated.

Table 2.3. Characterization of candidate X and Y sequences using the chromosome quotient (CQ) approach.

Putative X chromosomes		**Putative Y chromosomes**	
No. of contigs	**Size (bp)**	**No. of contigs**	**Size (bp)**
650	3,305,692	1,590	1,515,034

A significant portion of the sequences with BLASTn results (42.4% in the X chromosome, and 30.8% in the Y chromosome) corresponded to repetitive sequences. This included BAC sequences from *Calliphora vicina* achaete-scute complex, AS-C (accession numbers LN877230-LN877235), and microsatellite clone sequences from both *Chrysomya albiceps* (accession numbers DQ478598, DQ478605) and *Haematobia irritans* (accession number EF629377). In *Ca. vicina,* the *AS-C* gene complex is flanked by repeats and transposable elements [176]. Additionally, within Diptera, the AS-C gene complex (which is made up of the genes *achaete*, *scute*, *lethal of scute,* and *asense*) is located on the X chromosomes in *Drosophila* and is involved in the sex-determining pathway wherein *scute* is an X chromosome signaling element [177].

The remaining portion of putative X sequences included 16 sequences with hits on yolk protein genes (*L. cuprina* yolk protein D (*ypD*), yolk protein A (*ypA*) and yolk protein B (*ypB*) genes, accession number GU109181, and one from *Calliphora erythrocephala* yolk protein 3, accession number X7079), two sequences with a hit to the *no bloke* (*nbl*) gene (accession number MH173327), nine sequences corresponding to *HSP70* gene (accession number HQ609501) and 2 sequences with hits on paired box protein Pax-6-like (*eyeless* in *Drosophila*) gene (accession

numbers XM_023446990 and XM_023450490). Within higher Diptera, yolk protein accumulates in oocytes to be used during embryogenesis and development [175, 178]. Genetic and molecular studies in *D. melanogaster* and *L. cuprina* have shown that *yp* genes are specifically expressed in females [175, 179, 180] though in *Drosophila* (where there has been more work on the topic), there is evidence of low *yp* expression in males [181-183] and sperm [184]. Binding sites belonging to the sex-determining gene *doublesex* (*dsx*) have been found on *yp* genes signifying its role in sex specific regulation [175, 180, 185]. The presence of homologous *yp* sequences in *C. rufifacies* putative chromosome X sequences indicates that these genes are also female specific or female biased in *C. rufifacies* and possibly maintained on a small neo-X region of a chromosome. The gene *no bloke* (*nbl)* in *L. cuprina* [112], a homolog of *D. melanogaster's protein of fourth* (*pof*) gene [74, 75] (an RNA binding protein involved in dosage compensation by targeting the ancestral dipteran sex chromosome (chromosome 4) and chromosome X in *D. melanogaster*) was one of the BLAST hits on 2 putative chromosome Y sequences. In both *L. cuprina* and *D. melanogaster,* this gene has been found to be essential in both male and female viability and fertility [74, 112].

Homologous sequences of *L. cuprina's* heat shock protein *hsp70* were found in 9 sequences in putative chromosome Y. The promoter region of the *hsp70* gene has been used in sterile insect technique (SIT) studies to develop molecular conditional female lethal genetic modifications [186]. In mammals, *hsp70-Sox9* interactions have been implicated in sex determination with a complex formed at sites where SOX9 binds DNA [187]. A member of the family is reported as testis enriched in an eel [188]. In the putative Y chromosome contigs, 12.3% (57 sequences) of the BLASTn results had hits to the bacteria *Serratia marcescens* (NZ_HG326223, NZ_ALOV00000000, NZ_ATOH00000000). The presence of homologous sequences in *C. rufifacies* to these set of genes from the BLAST results in both the male and female putative sex sequences raises the possibility that a microbial genome may be involved in sex determination and differentiation in *C. rufifacies* as is seen in the isopod *Armadillidium vulgare* (Crustacea, Isopoda), where a chromosomal insertion of a *Wolbachia* genome drives sex determination [189], though it may also be possible that these are just symbiont sequences that escaped computational filters. Additionally, the signal could be a consequence of the sex determination system as has been observed in *C. elegans,* where lineages that self-fertilize are more sensitive to *S. marcescens* than those that outcross [190].

2.4.5 Muller Element F is not X-linked in *C. rufifacies*

Chromosomal gene contents, commonly known as Muller elements A to F in the genus *Drosophila* [191], are thought to be highly conserved across Diptera [1, 191]. Muller element F, the dot/fourth chromosome in *Drosophila*, is thought to be the ancestral X chromosome found in many major fly lineages [1, 107, 191]. Whole genomes of some non-drosophilid insect species which exhibit stable X-Y differentiated sex chromosomes were analyzed and it was determined that genes located on the *Drosophila* dot chromosome are X-linked in these species [1]. In *Drosophila* however, Muller F reverted back to an autosome more than 60 million years ago but has maintained many characteristics similar to a former X chromosome [107, 109]. Muller element F in most Calliphoridae segregates as the sex chromosome, and a dominant male determiner factor located on the Y chromosome directs differential expression of sex determining genes down the male path, leading to distinct structural differences [59, 119, 192]. In species in which the Muller element is sex-linked, one would expect to observe half as many sequencing reads to map to the reference sequences in males compared to females. When mapping male and female reads (both AF and TF) to each Muller element (A-F), less than 5% of the orthologous contig sequences segregated as X-linked to Muller elements (including Muller element F) (Figure 2.6, Table A3), confirming the high likelihood of undifferentiated sex chromosomes in *C. rufifacies* and introducing a lineage within Calliphoridae in which Muller element F is not the predominant sex-linked element.

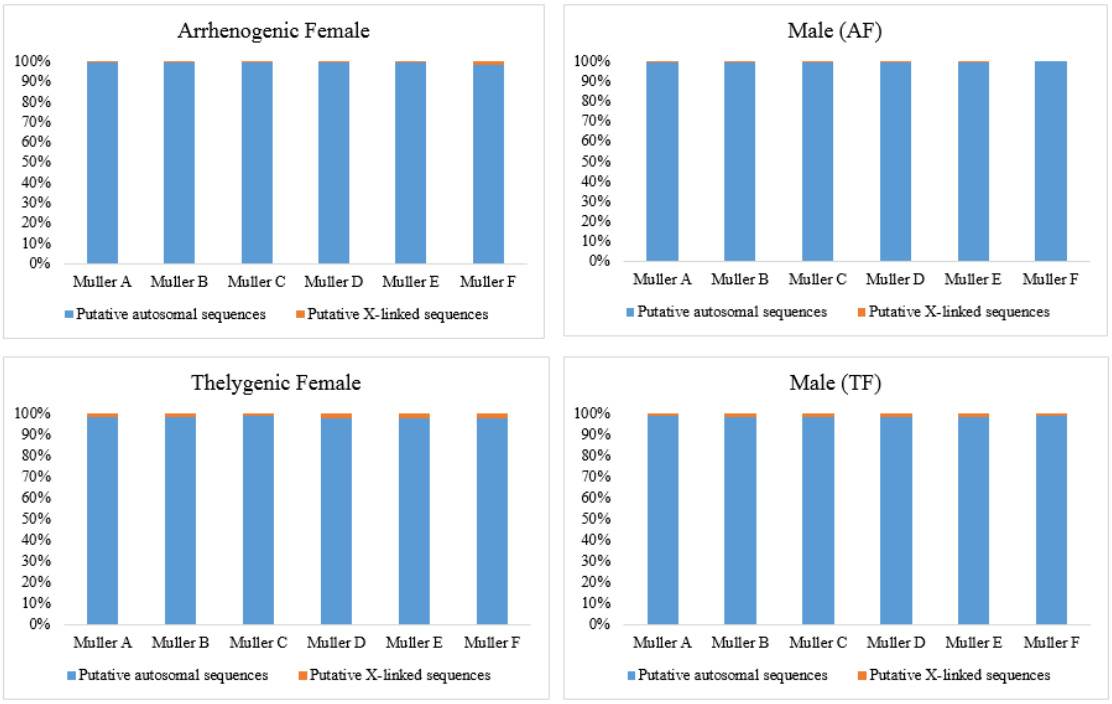


Figure 2.6. Sequence coverage analysis of *C. rufifacies's* Muller element orthologs from *D. melanogaster*. The two male panels are for the male sequences mapped to each of the female genomes. As expected in a homomorphic sex-determining system, gene sequences in the Muller elements are not X-linked. Instead an autosomal characteristic sequence coverage distribution is observed in all the elements including the Dipteran ancestral sex chromosome Muller F (AF = arrhenogenic female, TF = thelygenic female).
These results imply that the ancestral genes found in the Dipteran ancestral X chromosome (Muller element F) are not predominantly X-linked in *C. rufifacies*. They also suggest that the sex determination region may be a small region of the genome not easily detectable using coverage differentiation of euchromatic regions of the genome.

2.4.6 Repetitive Landscape

Repeat sequences have recently been found to be important precursors and contributors to eukaryotic genome's architecture, stability, evolution and environmental adaptation [193, 194]. In *Stomoxys calcitrans*, the Muller element suspected as the sex chromosome seems to exhibit a distinct repeat element pattern [195]. The repetitive landscape can cause alteration of a gene's function or may act as raw materials for new genes [196, 197]. The amount of repetitive DNA among insect species varies greatly [148, 198-200]. Some insects have greater than 50% of their genome occupied by repetitive elements (American cockroach, *Periplaneta americana* [200]) while others have less than 10% (*Phormia regina*, black blow fly [148]). The assembled portion of the *C. rufifiacies* genome has a small proportion of repetitive elements in the assembly,

accounting for 6.61% (18 Mb), 6.84% (20 Mb) and 6.89% (19 Mb) of the TF, AF and M assembled genomes, respectively (Table 2.1, Table A4). The predominant repetitive elements were simple repeats, which occupy approximately 4.3% (~12.5 Mb) of the *C. rufifacies* genomes. The remainder of the repetitive landscape comprised of ~0.5% of DNA retrotransposons (LTRs, LINEs and SINEs), ~0.2% DNA transposons (hAT, CMC, Maverick, Kolobok, Mule, P, PIF, PiggyBac, Sola, TcMar, Zator), ~0.7 rolling circle, ~1% low complexity regions and ~0.06% unknown repetitive sequences (Figure 2.7). In the characterized putative sex chromosomes, 6.17% of the X chromosome (~204 kb) and 2.77% of the Y chromosome (41.90 kb) were repetitive elements. As is the case observed in the whole genomes, simple repeats represented the highest set of repetitive elements in the sex chromosomes (Table A4).

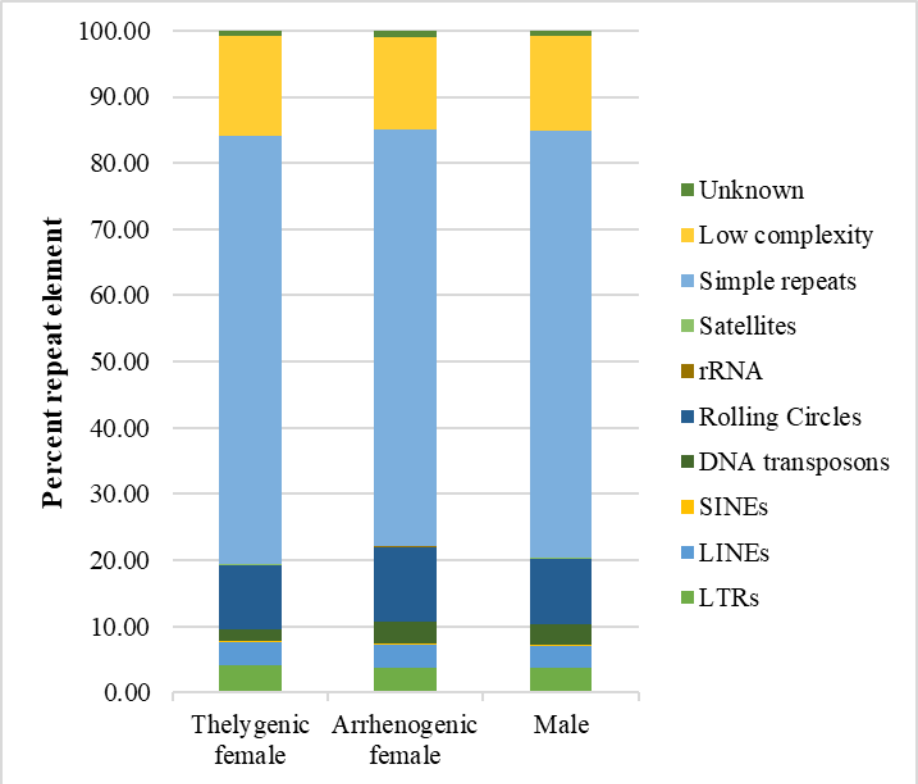


Figure 2.7. The graph shows the percentage of repeat elements composing the repetitive landscape in each sex type of *C. rufifacies*. Retrotransposons composed of SINEs, LINEs and LTRs occupied approximately 7% of the total repeatome, while DNA transposons occupied approximately 3% of the repeatome in the male and male producing females and ~2% in the female producing females. Satellites and rRNA can barely be seen on the graph as they occupied only 0.07% and 0.05% of the repeatome respectively. Simple repeats were the predominant repetitive element occupying almost 65% of the whole repetitive landscape.

2.5 Conclusion

Rapid diversification caused by changes in evolutionary processes has introduced variation in sex-determining mechanisms between and within species [1, 115, 201]. The family Calliphoridae is an excellent model for evaluation of sex chromosome evolution as both homomorphic (*C. rufifacies, C. albiceps* [101, 144]), and heteromorphic (*L. cuprina, P.regina* [143]) sex chromosomes are observed among closely related species. Additionally, while a majority of blow flies are amphogenic (females produce an equal ratio of male and female progeny), others, such as *C. rufifacies* and its sister species *C. albiceps* possess a distinct monogenic (females produce unisexual progeny [101, 122]) system, have two type of females (arrhenogenic and thelygenic [101, 122]) and the sex of the offspring is determined by the maternal genotype [122]. This may in fact be in response to selective pressures with respect to inbreeding – producing unisexual offspring guarantees full siblings will not mate with each other, thus resulting in a genetically robust population even when population numbers begin to decline. For example, parasitic wasps will alter the sex ratios of their offspring through a process of arrhenotoky (selective fertilized eggs or unfertilized eggs resulting in female or male offspring respectively [202]), allowing females to control the sex ratio of their offspring. Gall midges [203], Hessian flies [204], and certain populations of *Musca domestica* [205] have monogenic life histories, all of which is likely related to controlling for inbreeding depression, not uncommon when resources are scarce and unpredictable. Therefore, the presence of the individual sex draft genomes herein will facilitate addressing questions on the origination and evolution of the diversity of sex-determining mechanisms observed within Calliphoridae. Additional genomic analysis of sex chromosomes and sex-determining genes of other blow flies will allow a refined evolutionary history of not only the Calliphoridae sex system, but to more accurately predict and understand other insect systems in which the dominant factor producing one sex over another remains elusive.

As calliphorids are decomposers and filth flies [206], many of this group's adaptations have also resulted in their classification as agricultural pests [207] and their utility in forensic entomology investigations [208]. For instance, ectoparasitism emerged 6.95 million years ago in the obligatory parasite *Cochliomyia hominivorax* (screwworm fly) and 2.28 million years ago for the facultative parasite *L. cuprina* (Australian sheep blow fly) [115], with little to no understanding for the genetic basis of this adaptation. As for forensic utility, many of these species have evolved a keen sense of odor detection, useful for estimating postmortem intervals as they are capable of

rapidly colonize remains [209]. Genomic resources are thought to provide the basis for improving forensic estimates [148, 210]. The function of many Calliphoridae as decomposers of animal remains also means they are important nutrient recyclers [211, 212], which are becoming of greater interest in decomposition ecology as most previous research has been focused on autotrophic biomass [213]. Genetic information from such species will enhance efforts to assess evolution in these systems. Therefore, the addition of these draft genomes and the predicted protein-encoding genes will expand the taxonomic breadth of study organisms and provide unique insights into the molecular biology, ecology, and evolution of blow flies. This, in cooperation with genomic evaluations of other dipteran species, will contribute in the exploration and provision of new targets for pest control strategies based on controlling specific sexes. Currently, the sterile insect technique is still in use to control the primary screwworm fly (*Co. hominivorax*) in which males are irradiated and released into the environment. However, these mass production facilities must rear male and female offspring due to the reproductive biology of this species and difficulty differentiating between the sexes in the immature stages, resulting in production of a sex that is not even used and is thus discarded. Understanding the mechanism in which a single sex is produced, and being able to genetically modify other calliphorid species to include this switch, could provide both economic and agricultural benefits [214, 215].

In conclusion, this new genome consisting of three draft genomes of two females types and males represent additional genomic resources of a calliphorid fly with economic, agricultural, forensic and medical importance. The genomes identify an important link in the study of evolution and diversification of sex-determining systems. We provide evidence for a loss of sex chromosomes, or the movement of very small components of ancestral sex chromosomes to autosomes, as there is little evidence for sex chromosomes in the genome (though some contigs identified do align with traditional sex-determining chromosomes) and no obvious pattern in Muller element allocation of such sites. Several interesting hypotheses regarding the sex-determination mechanism of this species arise from this work including the role of the *no blokes* / *painted on the fourth*, *scute*, *yolk proteins*, and potentially inserted *Serratia marcescens* genes in this unique monogenic sex-determination system with seemingly no (or very small and possibly neo) sex chromosomes. Interestingly, canonical sex determination genes (*transformer* and *Musca domestica male determiner*) either produced truncated proteins when annotated (*tra*) or did not align (*Mdmd*) with our genomic scan for sex-determining elements. These results are similar to a

previous chromosomal staining experiments in the species that only found evidence for *daughterless* near a suspected sex-determining translocation [216], though it is worth noting that a full accounting of *Drosophila* sex determination loci was lacking at the time of that experiment. It is also worth noting that *daughterless* and *scute* (identified as a putative X chromosomal sequence here) interact in *Drosophila* [217], providing (along with the *no blokes* location on a putative Y chromosomal contig) some evidence that a dosage compensation-like molecular function [218] may be important in *C. rufifacies* sex determination. This hypothesized role of dosage compensation coincides with the observed differences in genomic coverage between males and females, where females exhibit two peaks in coverage and males exhibit one. Furthermore, *eyeless* (*Pax-6*) is known to interact with both *daughterless* [219] and there is some support for it interacting with *doublesex* [220] in *Drosophila*; deepening the original support for a role of *daughterless* in the *Ch. rufifiacies* sex determination system. Additional connections of identified targets include *hsp70-Sox9* regulation of sex in some systems [187] and common co-regulation by Pax/Sox genes in a variety of systems [221]. Additional work on the annotation of the sex-determining cascade of genes, as well as the identification of the master switch in *Ch. rufifacies*, will lead to invaluable and potentially wide-ranging implications across evolutionary biology. Although these genomes have some limitations (mostly fragmented genomes), the genomes and identified targets here are ideal starting points for more detailed dissections of this sex determination mechanism and sex chromosome evolution.

CHAPTER 3. SEX CHROMOSOME STRUCTURE IN FOUR SPECIES OF BLOW FLIES (DIPTERA: CALLIPHORIDAE)

3.1 Abstract

Sex chromosomes are evolutionarily labile and have repeatedly evolved in numerous eukaryotic species. The accepted evolutionary theory states that sex chromosomes were derived from a pair of homologous autosomes (homomorphic) which began to differentiate after the emergence of a sex determining gene on one of the pairs thus suppressing recombination. However, rapid evolution of sex chromosome systems has caused some closely related species to exhibit heteromorphic sex systems while others retaining or reverting back to the ancestral state of homomorphic sex chromosomes. One of the ways to understand the mechanisms involved in the rapid evolution of sex chromosomes is by characterizing and analyzing sex chromosome structures of closely related species which display both types of sex chromosomes. Blow flies are as such a good model to study sex chromosome evolutions as they have a generally conserved cytological karyotype but a diverse sex chromosome system. This study therefore uses genomic sequence data to identify and characterize putative sex chromosomes in four blow fly species: *Phormia regina* and *Lucilia cuprina* which exhibit heteromorphic sex chromosomes and *Chrysomya rufifacies* and *Chrysomya albiceps* which have homomorphic sex chromosomes and are also monogenic with two types of females (arrhenogenic and theylgenic females). This study presents deduced sex chromosomes sequences of the four blow fly species. Genomic evidence that *P. regina* and *L. cuprina* have differentiated sex chromosomes while *C. rufifacies* and *C. albiceps* have undifferentiated sex chromosomes is presented. Moreover, the ancestral Dipteran sex chromosome (Muller element F in *Drosophila)* is highly conserved and has remained as the sex chromosome in some species of blowflies.

3.2 Introduction

In most eukaryotic organisms, sex determination is governed by diverse mechanisms which rapidly evolve [222]. One of these mechanisms is the sex chromosome system where a specific set of chromosomes carrying genes related to sexual characteristics contribute to the differentiation and determination of sex [2, 9, 223]. Sex chromosome systems vary considerably. Dimorphic sex

chromosomes are observed in diverse multicultural eukaryotes such as mammals [62, 63], insects [7, 10, 100], fish [41, 224, 225], birds [6, 30, 34] and even plants [44, 50, 78]. The most common and familiar sex chromosome system is the XX/XY sex system, which involves a homogametic female (XX) and a heterogametic male (XY) [9, 11]. An accepted theory of sex chromosome evolution postulates that sex chromosomes evolved from a pair of ordinary homologous autosomes after the acquisition of a sex determining region in one of the autosomes (Y), suppressing recombination [9]. A degenerated Y chromosome thereby emerged giving rise to morphologically and functionally distinct X and Y sex chromosome [9, 226]**.** A lack of recombination on the Y chromosome led to a loss of active genes, and an accumulation of sexually antagonistic variants on the X chromosome [227]. Based on evidence obtained from cytological karyotype analysis, highly differentiated sex chromosomes display an X and Y chromosome which differ in shape and size (large X chromosome and a small Y chromosome) [70, 226]. These chromosomes exhibit few signatures of their evolutionary history. Therian mammals, including humans, are an example - males possess a degenerated Y chromosome which contains few genes, has male specific functions and a sex determining gene (SRY) responsible for the male phenotype [228]. In contrast, the X chromosome typically resembles autosomes in gene density as it mostly maintains its original size and most of its ancestral gene contents [15]. However, not all species exhibit heteromorphic sex chromosomes. In amphibians, approximately 96% of its species possess homomorphic sex chromosomes where the X and Y chromosomes are almost identical with a few differences in gene content [138]. Homomorphic sex chromosomes are also observed in some snake species [25, 43], some insect species [28, 114], and even some avian species [23, 30]. Sex chromosomes which display low levels of differentiation are usually considered to be evolutionarily young and at the initial stages of evolution [19, 20] as is observed in the insect *Drosophila miranda* [20, 21]. The expectation is that they will eventually differentiate as they age. However, this theory does not always hold true in all the species with homomorphic sex chromosomes. Emus and the mosquitos in the genus *Aedes* possess old homomorphic sex chromosomes which have not diverged from their ancestral state [23, 28-31].

The basic ancestral karyotype in many higher Diptera is composed of five large euchromatic chromosomes and a small heterochromatic dot chromosome [229]. The gene content of these chromosomes are highly conserved in Diptera and are hypothesized as syntenic in numerous Dipteran species [191, 229-231]. A nomenclature system proposed by Muller [229],

refers to the conserved chromosomal gene elements as Muller elements A – F representing each of the chromosomal arms [108, 229]. Despite a similar chromosomal architecture, the nature of sex chromosomes varies among closely related Dipteran species [91]. *Musca domestica,* the common housefly, reflects this diversity [232]. There are strains within *M. domestica* which follow the standard norm of a homogametic XX female and a heterogametic XY male containing a male determining gene on the Y. While other strains possess homomorphic sex chromosomes and the male determining gene can be located in any of the five chromosomes [105, 114].

Muller element F which is the small dot chromosome or chromosome 4 in *D. melanogaster*, is considered as the ancestral sex chromosome in higher Diptera [1, 233]. However, it reverted back to autosome in the lineage leading to *Drosophila* [107, 234] but has maintained characteristics similar to an X chromosome such as its feminizing effect [107]. The size of Muller element F differs even across *Drosophila* species. In *D. melanogaster,* it is ~5.2 Mb while in *D. ananassae* it is >18.7 Mb [235]. The difference in size has been attributed to the presence of transposons – specifically long terminal repeats (LTRs) and LINE retrotransposons [235]. Despite the presence of a high level of repetitive elements, Muller F contains the same gene density as the other chromosomes [109, 110, 235], with genes located on the distal portion of its long arm which contains euchromatic DNA [110]. Orthologs of genes located on Muller F in *D. melanogaster* are X-linked in some non-drosophilid insect species [1, 107, 234].

Blow flies (Calliphoridae) are ubiquitous and distributed worldwide. They are a diverse species and accordingly, display a diverse sex chromosome system. Their last common ancestor is estimated to be around 22 mya which was followed by a rapid radiation of the subfamily Chrysomyniae (~17 mya) and Luciliinae sister-lineages (~16 mya) [115]. Insects within the Calliphoridae family show a remarkably uniform karyotype composed of five pairs of large euchromatic autosomes and a pair of heteromorphic XX/XY sex chromosomes [59, 119, 236]. However, there are some blow fly species which possess homomorphic sex chromosomes (e.g *C. rufifacies* and *C. albiceps*) [101, 144]. It has been shown that the dot chromosome segregates as the X chromosome in most blow fly species as is also observed in other insects within Brachycera [59, 120, 192]. Rapid diversification within Calliphoridae has produced a diverse presence of sex chromosome systems observed even within the same subfamily [115]. For example, in the subfamily Chrysomyinae, a majority of the blow fly species are amphogenic, which means females produce both male and female offspring in the same ratio. However, there are some species which

are monogenic whereby two types of females exist which only produce offspring of one sex – arrhenogenic females (produces male offspring) and thelygenic females (produces female offspring) [123, 124].

Phormia regina, *C. rufifacies* and *C. albiceps* belong to the Chrysomyinae family while *L. cuprina* belong to the Luciliinae family [237]. *Phormia regina* is one of the most common blow fly species across North America and prefers cooler temperatures [238]. It has the most 'normal' life history with no specialized parasitic adaptations or known unique sex determination strategies. *Lucilia cuprina*, *C. rufifacies* and *C. albiceps* prefer slightly warmer climates, albeit being found all around the world [239]. In the United States, they are predominantly found in the southern half of the country [240, 241]. Blow flies within the *Lucilia* species have undergone adaptive phenotypic change on a rapid scale whereby depending on their geographic location they either feed on live tissue (ectoparasites) or on dead tissues (carrion) [198, 242]. In the United States, *L. cuprina* is one of the common blow flies which are the primary colonizers of human remains and is typically used as a forensic indicator species in estimating postmortem intervals [243]. Interestingly, in Australia and New Zealand, *L. cuprina* is almost exclusively a facultative ectoparasite which mainly feeds on living tissue from sheep and causes fly strike [243, 244]**.** While in South Africa, it has been known to harbor both characteristics – as a sheep parasite and as a carrion colonizer [239, 242]. It is hypothesized that the rapid diversification of the Calliphoridae family may be due to peculiar geographical changes and evolutionary processes which created new niches that triggered an adaptive radiation [245].

Identification and characterization of sex chromosomes in a diverse array of non-model species has been made possible by new bioinformatic methods [246]. The recent explosion of next generation sequencing (NGS) data has made it possible to access whole genome sequences from a variety of organisms including non-model organisms lacking a well annotated reference genomes [23, 25, 156, 225]. Next generation sequencing (NGS) data has made it easier to develop and access less expensive strategies for identifying sex chromosome sequences. Some of the ways sex chromosomes are identified include a genomic coverage approach using next generation sequencing data to exploit differences in sex chromosome ploidy between males and females [107]; leveraging sex differences in gene expression to identify sex-limited transcripts from the Y chromosome [247], and also, utilizing restriction site-associated DNA sequencing (RAD-seq) to identify sex-limited loci [248]. The chromosome quotient (CQ) approach (as described by Hall *et*

al. [156]), which is one of the methods that employ genome coverage to distinguish sex chromosomes from autosomes, was used to characterize sex chromosomes and sex specific sequences in the four blow fly species used in this study. It has successfully been used to identify sex specific sequences and genes in the *Anopheles* [156] and *Aedes* [26] mosquitos. The CQ approach allows for the differentiation of X, Y and autosomal chromosomes from each other using read coverage ratios of alignments and is useful in characterizing differentiated sex chromosomes [1].

Sex chromosomes in blow flies has not been fully explored. A few studies have been performed on the sex chromosome of *L. cuprina* [104, 112, 141, 236] however not many have delved into the genomic composition and characterization of blow fly sex chromosome. Therefore, a comparative genomic approach was utilized to analyze the genomic composition and characterization of the sex chromosomes of four blow fly species: *P. regina, L. cuprina* (both with heteromorphic sex chromosomes), and *C. rufifacies* and *C. albiceps* (sister species with homomorphic sex chromosomes and monogenic females). Additionally, the Dipteran ancestral sex chromosome, Muller element F, was identified and characterized in the four blow fly species.

3.3 Methods

3.3.1 Characterizing Putative Sex Chromosomes Using the Chromosome Quotient Approach and Nucleotide Blast (BLASTn)

The chromosome quotient (CQ) approach as described by Hall *et al*. [156] was used to characterize the putative sex chromosomes in the four blow fly species. The CQ method discovered isolates putative sex chromosome sequences by using a stringent alignment criterion of the male and female reads onto each other's genomes. The stringent settings required a whole read to map onto the reference contigs with zero mismatches [156]. This was done in order to reduce the number of false positives which may be caused by repetitive sequences from Y chromosomes with closely related sequences on the autosomal or X chromosomes [156]. CQ uses the assumptions that (i) males and females share the same complement of autosomal sequences and are present in the same quantities in both sexes, (ii) females have two X chromosomes while males have only one, and (iii) Y sequences are only found in males [156]. This means that X chromosome sequences in females should be roughly twice as frequent as in males and unique Y sequences

should only be present in males. To accommodate the difference in read coverage between the male and female sequence data and minimize bias, CQ values were normalized to the median CQ of autosomal sequences, which are assumed to be one. Using a somewhat subjective approach, male contigs with a CQ ratio of less than 0.3 were characterized as putative Y sex chromosomes to accommodate repetitive Y sequences that may be present in both the male and female, while female contigs with a CQ ratio ranging between the arbitrary 1.6 and 2.5 were grouped as putative X chromosomes. This CQ interval was selected to reduce false positives. In *Chrysomya rufifacies*, the CQ approach was implemented on the data sets of the two types of females (theylgenic and arrhenogenic – each female independently), resulting to two sets of X and Y sequences. The two predicted sets of putative Y sequences were compared to determine the proportion of overlap shared between them. A total of 23,624 contigs (~64 Mb) and 7,448 contigs (~15 Mb) from the arrhenogenic and thelygenic female respectively, were categorized as putative X chromosomes. A comparative analysis of both sets of putative X chromosomes was performed by CD-HIT-2D-EST v4.5.6 [157, 158], to isolate a representative set of *C. rufifacies* chromosome X sequences characterized by both females, using a length difference cutoff and a sequence identity cutoff both of 80%. A nucleotide BLAST was performed on the isolated sex chromosome sequences for all the four blow fly species against a non-redundant nucleotide database (BLASTn v2.6.0+, E-value cutoff ≤ 1E-5 [159]). Gene ontology terms were extracted from the annotated sequences using Blast2GO v5.2.5. Enrichment analysis of the GO terms was performed using the Fishers exact test plugin on Blast2GO, and they were filtered using a FDR value of ≥ 1E-10. The webserver REViGO [160] was used to summarize and present a representative subset of the enriched terms using a clustering algorithm. The allowed similarity was set to 0.7 (medium) and the database of GO terms selected was from *D. melanogaster*.

3.3.2 Muller Elements Identification (Muller F)

Drosophila melanogaster's chromosomal gene contents (Muller elements A-F coding sequences) were downloaded from GenBank (www.ncbi.nlm.nih.org, downloaded on 08/2017, version NC_004354.4). A total of 10,448 coding sequences of the longest isoforms representative of each of the gene contents were isolated and used to query both the male and female genomes of the four blow fly species (tBLASTx v2.6.0+, E-value ≤ 1E-5). The longest isoform was selected because it was the most comprehensive isoform with the largest total coding exon size.

Orthologous sequences of the Muller elements for each sex of the four blow fly species were identified. Orthologous contig sequences were assigned to the respective Muller elements. To determine which Muller elements were X-linked in *C. rufifacies*, male and female sequence reads were aligned to the identified orthologous contig sequences using the CLC-GWB v9 read mapper, and the read coverages compared. To reduce false positives, stringent mapping parameters were used such that 100% of each read needed to have at least 80% identity to be included in the final mapping. The program DESeq [161] was used to identify any differential read coverages observed within the orthologous Muller elements to identify sequences with a twofold higher abundance in females than males, by calculating a Log2(M/F) coverage ratio. Contig sequences with a Log2 (M/F) coverage ratio within the range of -0.6 and -1.3 were considered to be X-linked.

3.3.3 Repetitive Element Annotation

A library of all known Dipteran repetitive elements was used to identify repetitive elements in each of the 3 genomes and the putatively characterized X and Y chromosomes using the program RepeatMasker v4.0.7 in default mode [162].

3.3.4 Data Availability

The *P. regina* genomes used in this study are available from NCBI using the accessions GCA_001735545.1 and GCA_001735585.1 for the female and male respectively [148]. *Phormia regina's* DNA sequenced reads are available from accessions SRR4047460 and SRR4047244. *Lucilia cuprina* and *C. albiceps* male and female genomes were locally assembled (unpublished) using the CLC genomics workbench following the protocol described in Chapter 2 which was used to assemble the *C. rufifacies* genome. DNA sequenced reads of the female *L. cuprina* can be accessed from NCBI accession number SRR1200273; those from *C. rufifacies* can be accessed from BioProject ID PRJNA575047 and SRP238163, while DNA sequenced reads for *C. albiceps* came from Brazil (unpublished).

3.4 Results and Discussion

3.4.1 Characterization and Annotation of Putative Sex Chromosomes in Blow Flies

Most blow flies with heteromorphic sex chromosomes exhibit dimorphic genomes sizes, with the female having a slightly larger genome than the male [148, 164]. However, some blow fly species with homomorphic sex chromosomes have a minimal difference in the genome sizes between the male and female, resultantly proposing their genomes are not dimorphic [164]. The genome size differences (or lack of) between male and female flies may be attributed to the sizes of the X and Y chromosome. Within heteromorphic sex chromosomes, the X is slightly larger than the Y due to a degenerated Y [9, 11, 59], while in homomorphic sex chromosomes X and Y sex chromosomes appear to be identical in size and exhibit few differences from each other in gene content. They are consequently difficult to distinguish by solely using karyotype data and read coverage information [1, 59, 101, 122, 144].

Independent estimates of genome sizes in blow flies from both flow cytometry and genome assembly studies confirm the presence of dimorphic genome sizes in the blow fly *P. regina* [148, 164] and *L. cuprina* [164] indicating the presence of heteromorphic (differentiated) sex chromosomes. Additionally, almost nearly identical genome sizes in *C. rufifacies* [164] were observed confirming the presence of homomorphic (undifferentiated) sex chromosomes. A difference of ~12 Mb is observed from flow cytometry measurements in the female *P. regina* (529.3 Mb) and the male (517.7 Mb), while in *L. cuprina*, a difference of ~100 Mb (female = 665.4 Mb, male = 567.9 Mb) is observed between the female and male. *Chrysomya rufifacies* only shows a difference of 0.2 Mb between the female (425.8 Mb) and the male (426.0 Mb) genomes is observed, however, this study did not differentiate between the two types of females [164]. Flow cytometry studies have not yet been performed on *C. albiceps* however as it is closely related to *C. rufifacies* (sister species) and sometimes are misidentified as the same species [249] as they share very similar character states (homomorphic sex chromosomes, and unique sex determination system – monogeny, or presence of two types of females – arrhenogenic females and thelygenic females [101, 237, 249]), it can safely be assumed to contain similar genomic characteristics as *C. rufifacies.*

Table 3.1. Putative sex chromosomes of the four blow fly species characterized using the chromosome quotient approach.

Species	**Putative X chromosomes**		**Putative Y chromosomes**	
	No. of contigs	**Size (bp)**	**No. of contigs**	**Size (bp)**
Chrysomya rufifacies	650	3,305,692	1,590	1,515,034
Chrysomya albiceps	14,677	22,969,545	43,899	75,420,199
Phormia regina	32,307	58,974,505	32,709	30,067,802
Lucilia cuprina	6,428	16,807,791	19,531	71,063,377

Putative sex chromosomes of the blow fly *P. regina, L. cuprina, C. rufifacies, and C. albiceps* were characterized using the chromosome quotient (CQ) approach. The CQ method identified and characterized putative X and Y sex chromosomes using read coverage ratios of male and female alignments as described in the methods section. CQ ratios distributed ~1 was characterized as autosomal sequences in both sexes, ~2 for putative X chromosome sequences, and ~0 for putative Y sequences. A total of 32,307 (~59 Mb) and 32,709 (~30 Mb) contig sequences (Table 3.1) were characterized as putative chromosome X and Y, respectively for *P. regina,* a characteristic consistent with differentiated sex chromosomes due to the presence of a larger X than Y chromosome. To the contrary, a total of 6,428 (~16 Mb) and 19,531 (~71 Mb) contig sequences in *L. cuprina* were characterized as putatively belonging to the X and Y chromosomes respectively (Table 3.1). The characterized Y chromosomal sequences are unexpectedly larger than the X chromosome which was not expected as *L. cuprina* exhibits differentiated sex chromosomes. The male genome used in this study was locally assembled (see methods**)** using reads sequenced from a male *L. cuprina* collected from Florida, USA. However, the female genome was assembled from reads sequenced from an Australian *L. cuprina* species. The differences in genome sizes between the male and female assembled *L. cuprina* genomes is ~274 Mb, with a bigger male genome compared to the female (Table 3.1). A comparison of the genome qualities of the published *L. cuprina* [198] and our locally assembled male and female genomes show they are similar and have relatively good assembly qualities (Table 3.2). Therefore, a plausible explanation for the different sex chromosome sizes would be that the two flies are genotypically different due to spatial or adaptive divergence, which may have introduced bias in the characterization step and likely accounting for these differences. Since Australian *L. cuprina* are ectoparasites and the North American strain is primarily a carrion colonizer, adaptations in their respective geographical niches may have influenced their genetic makeup in order to successfully exploit their environments, as has been observed in cichlid fish [250-252]. It is

therefore a possibility that the two are different strains of *L. cuprina* and thus comparing their sex chromosomes may not give accurate data. Furthermore, the chromosome quotient approach uses a stringent mapping criterion, where 100% of the read has to align to the reference for it to be used in the female/male read quotient calculation. This may suggest that some female reads may not have mapped onto the male reference sequence, and characteristically those sequences improperly identified as male specific (putative Y sequences) contributing to the large set of sequences characterized as male specific. Additionally, there is an ~ 280 Mb genome size difference between the assembled genomes (Table 2.3), where the male genome is ~611 Mb as compared to the female genome of ~337 Mb. A bigger male genome may also have contributed to an improper characterization of putative chromosome Y sequences.

In *Chrysomya rufifacies,* the CQ approach was implemented on the data sets of the two types of females (theylgenic and arrhenogenic – each female independently), resulting to two sets of X and Y sequences. A total of 2,195 contigs (~2 Mb from male and arrhenogenic female comparison) and 4,031 contigs (~4 Mb from male and thelygenic female comparison) were identified as putative Y chromosomal sequences. The two predicted sets of putative Y sequences were compared to determine the proportion of overlap shared between them. A total of 23,624 contigs (~64 Mb) and 7,448 contigs (~15 Mb) from the arrhenogenic and thelygenic female respectively, were initially categorized as putative X chromosomes. A comparative analysis of both sets of putative X chromosomes was performed to isolate a representative set of *C. rufifacies* chromosome X sequences characterized by both females. Therefore, a total of 650 contig sequences (~3.3Mb) and 1,590 contig sequences (~1.5Mb) were characterized as putative X and Y sequences for *C. rufifacies,* respectively. As expected, only ~1.6 Mb of the sequences between the male and female genome was differentiated which is consistent with the presence of undifferentiated sex chromosomes (Table 3.1)**.** However, in *C. albiceps,* a total of 14,677 (~22Mb) and 43,899 (~75Mb) contig sequences were identified as putative X and Y sequences for *C. albiceps,* showing a larger portion of Y sequences as compared to the X. Comparing the two genome sizes, male *C. albiceps* genome is ~21 Mb larger than the female Table 3.2**)**. A possible explanation of the differences observed in the characterized sex specific sequences may be due to the presence of the sequenced reads obtained from different populations. Sequenced reads used in the assembly of the male *C. albiceps* genome came from two flies, one from the wild (a natural park) and the other a lab strain. While the sequenced reads used to assemble the female *C. albiceps*

genome came from two wild type female flies (a farm). Additionally, it is unknown what type of female was sequenced. It is not known whether they were both arrhenogenic, thelygenic or one of each. Ideally, sequenced data used for genome assemblies is preferred to be obtained from inbred homozygous individuals in order to avoid data complexity created by heterozygosity and allelic variations in sequenced individuals [253]. This may result in reads being left out of the read alignment step as CQ utilizes stringent mapping parameters only allowing a read to be mapped if it has zero mismatches with the reference contigs.

Table 3.2. Genome assembly statistics of the four blow fly species. This includes both locally assembled draft genomes (*C. rufifacies* and *C. albiceps*, *L. cuprina)* and the published Australian *L. cuprina* (*, [198]) and *P. regina* male and female genomes [148].

	Lucilia cuprina			***Phormia regina***		***Chrysomya rufifacies***			***Chrysomya albiceps***	
	***Unknown sex**	**Female**	**Male**	**Female**	**Male**	**Arrh. Female**	**Thel. Female**	**Male**	**Female**	**Male**
Assembled genome size	458 Mb (GCA_001187945.1)	337 Mb	611 Mb	550 Mb (GCA_001735545.1)	534 Mb (GCA_001735585.1)	295 Mb	279 Mb	289 Mb	293 Mb	315 Mb
Flow cytometry genome size	-	665.4 Mb	567.9 Mb	529.3 Mb	517.7 Mb	425.8 Mb		426 Mb	-	-
#Contigs	74,043	130,917	88,495	192,662	187,700	114,048	107,111	109,341	172,109	159,974
N50 (bp)	744,413	3,866	24,771	7,918	7,177	4,101	3,889	4,164	2,084	
GC content	29.3%	29.0%	29.3%	28.0%	28.0%	27.3%	27.3%	27.2%	27.6%	27.4%
Repetitive Sequence	57.80%	-	10.6%	8.7%	7.3%	6.8%	6.6%	6.9%	6.9%	7.1%
# Predicted Genes	14,554		13,061	8,312	9,490	13,910	13,590	13,798	9,161	8,461

- Denotes unavailable information.

Flow cytometry genome sizes (Mb) data retrieved from [164]

Overall, more than half (~57%, ~85%, and ~86%) of the characterized putative X sequences in *P. regina, L. cuprina,* and *C. rufifacies* respectively had homology to sequences in GenBank (E-values less than 1E-5), as compared to the Y sequences (Table 3.3). However, *C. albiceps* had relatively almost the same proportion of sequences with BLASTn results in both the characterized X and Y sequences (~62% and ~66.8%) (Table 3.3). The gene *painting of fourth* (*Pof*), which codes for an RNA-binding protein that binds to chromosome 4 (dot chromosomes) in *Drosophila*, is a remnant of chromosome dosage compensation of the ancestral *Drosophila species* which had chromosome 4 as the X chromosome [74, 75] and is important for expression of *Drosophila* chromosome 4 genes. *No blokes* (*nbl*), an ortholog of *Pof* could be important for X chromosome dosage compensation in *L. cuprina* as it is required for normal levels of gene expression of most X-linked genes [112]. Nucleotide blast hits of *nbl* were found on the sequences characterized as sex chromosomes. Approximately 0.13% (76,495 bp), and 0.19% (325,520 bp), of the characterized putative X chromosome in *P. regina* and *L. cuprina* respectively had hits to the *L. cuprina* isolate LA07 no blokes (*nbl*) gene (accession number MH173327). Approximately 0.06% (19,000 bp) and 0.42% (295,882 bp) of the characterized Y sequences from *P. regina* and *L. cuprina*, respectively, had hits to *nbl*. In the homomorphic species, characterized putative X chromosomes of *C. rufifacies* did not have any hits to the *nbl* gene. In the Y sequences 0.1% of the characterized Y sequences (1,518 bp) had homology to the *nbl* gene. While in *C. albiceps*, both the X and Y sequences had ~0.06% of sequences with homology to *nbl.* This shows that dosage compensation mechanism may be actively involved in *P. regina* and *L. cuprina* to ensure the same expression of X-linked genes on the heterogametic (XY) sex as compared to the homogametic (XX) sex. In *C. rufifacies* and *C. albiceps*, a small proportion of the sequences were homologous to the *nbl* gene suggesting the two species may be lacking an active dosage compensation system. Other nucleotide blast results of interest included yolk proteins (*L. cuprina yolk protein* D, A and B (*ypD, ypA, ypB*) – accession number GU109181) which are female specific proteins found in the oocytes of Dipteran species. Major yolk proteins are derived from vitellogenins and they provide nutrients for developing embryos [178, 254].

Table 3.3. Nucleotide Blast results (BLASTn) of the characterized putative X and Y sex chromosomes

	Putative X chromosomes			Putative Y chromosomes		
Blow fly species	**No. of contigs**	**# contigs BLASTn hits**	**%BLASTn hits**	**No. of contigs**	**# contigs BLASTn hits**	**%BLASTn hits**
Phormia regina	32,307	18,254	56.5	32,709	14,237	43.5
Lucilia cuprina	6,428	5,478	85.2	19,531	15,191	77.8
Chrysomya rufifacies	650	559	86.0	1,590	465	29.2
Chrysomya albiceps	14,677	9,030	61.5	43,899	29,306	66.8

Gene ontology (GO) enrichment analysis showed that 11, 4 and 6 enriched GO terms within biological processes (BP), cellular component (CC) and molecular function (MF), respectively of the X sequences, were shared in all four species (FDR ≥ 1E-10, Table A5). Some of these enriched GO terms include reproduction, development process, response to stimuli, protein folding, and transporter activity. Of the Y sequences**,** 14 and 19 enriched GO terms within cellular component (CC) and molecular function (MF) (Table A6.), respectively, were shared in all four species. Some of the enriched GO terms include endomembrane system, translation regulator activity, molecular function regulator, transporter activity. A full list of the GO terms can be found on Table A5 and Table A6.

3.4.1.1 Repetitive Element Analysis of Characterized X and Y Sequences

Repetitive analysis of the predicted putative sex chromosomes showed that ~7%, ~7%, ~9% and ~8% of the predicted X sequences in *C. rufifacies, C. albiceps, P. regina and L. cuprina,* respectively, were occupied by repetitive sequences. Additionally, ~3%, ~8%, ~6% and ~13% in the same order were respectively occupied by repetitive sequences in the predicted Y sequences (Table A7). In both the putative X and Y sequences, simple repeats occupied the largest percentage of the total repeatome – between 30-70% (Figure 3.1). A study on 136 insect genomes showed that simple repeats (microsatellites) were mostly located in intergenic regions [255] and were ubiquitously distributed across the genomes which would explain their presence in the putative sex chromosomes. Total interspersed repeats were also in abundance (Figure 3.1). Studies have shown that most interspersed repeats appear to be derived from transposable elements, which are major components of eukaryotic genomes as they affect genome structure and evolution [256]. There was an almost complete absence of small RNA and satellites from all four blow fly species

(Figure 3.1). Within sex chromosomes, amplification of repeat sequences facilitates suppression of recombination which leads to genetic degradation of Y chromosomes and also, the evolution of heteromorphic sex chromosomes [11, 226]. Therefore, it is expected that the putative Y sex chromosomes display would contain a higher percentage of repetitive sequences than the X, especially in the heteromorphic species. However, the putative X and Y sex chromosomes in both the homomorphic and heteromorphic blow fly species, exhibited an almost equal distribution of the repetitive elements between the X and Y. One explanation would be due to the difficulty assembling repetitive sequences, which are therefore overlooked and left out of the assembly [257].

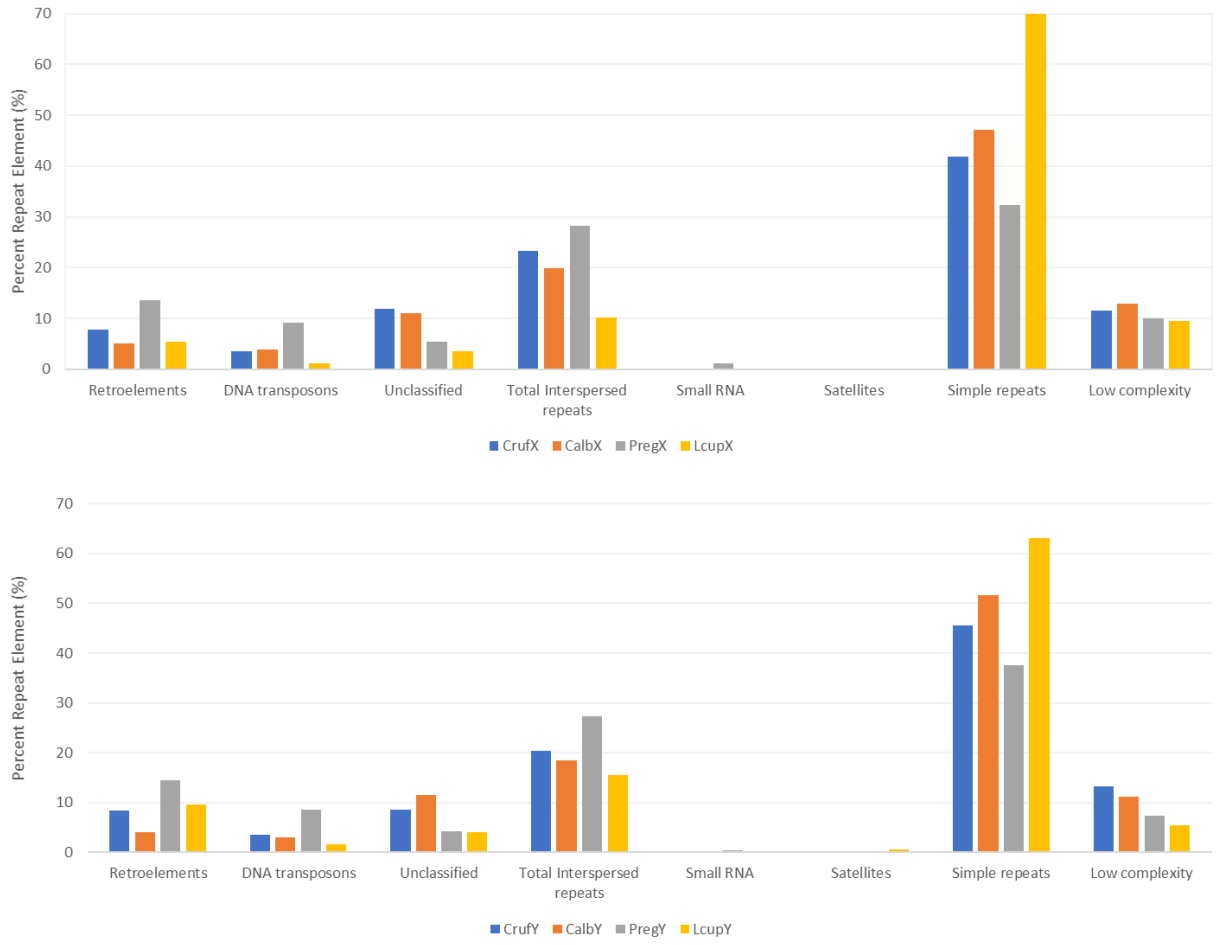


Figure 3.1. *Repetitive elements distribution in putative chromosome X and Y of all four blow fly species: Cruf = C. rufifacies, Calb = C. albiceps, Preg = P. regina, Lcup = L. cuprina*

3.4.2 Ancestral Dipteran Sex Chromosome – Muller Element F

Drosophila's fourth chromosome (also known as the dot chromosome or Muller element F) segregates as an X chromosome in higher Diptera (Brachycera) and has been identified as the ancestral sex chromosome in several fly lineages [1, 229]. Comparative mapping studies using differential sequence coverage approach between male and female flies showed that fly species with differentiated sex chromosomes exhibit an X-linked Muller element F due to the presence of an overrepresentation of female reads as compared to male [1]. Genes within Muller element F which were X-linked obtained mapped half as many male reads mapped onto them as female reads due to the male species having only one copy of the X chromosome as compared to two copies in females. In order to determine if Muller element F is widely X-linked in Calliphoridae, whole genome sequencing data was used to calculate the relative coverage of male and female reads in *P. regina, L. cuprina, C. rufifacies* and *C. albiceps.*

Approximately 67% and 55% of the contig sequences which segregated as belonging to the ancestral sex chromosome Muller element F were X-linked in *P. regina* and *L. cuprina* respectively. This was expected as both display differentiated sex chromosomes (Figure 3.2, Table A4). Less than 1% of the contig sequences segregating as belonging to the autosomal Muller elements A, B, C, D and E were determined to be X-linked, affirming the absence of other Muller elements segregating as the X chromosome. This also confirms the remaining five Muller elements (A, B, C, D and E) have autosomal properties with the same copy number of read coverage in both sexes. Despite the uncharacteristic size prediction of putative X and Y sex sequences via the CQ approach, a similar pattern was observed in *L. cuprina.* Approximately 57% of contig sequences identified as containing gene orthologs of *D. melanogaster's* Muller element F gene sequences were detected to be X-linked, a higher percentage than what was observed in the other autosomal Muller elements (Figure 3.2, Table A4).

Undifferentiated sex chromosomes are difficult and challenging to identify due to limited sequence divergence between the sex chromosome pairs [1, 114]. Additionally, karyotype data show that they appear morphologically undifferentiable (homomorphic) [59, 101, 249]**.** As expected, *C. rufifacies* and *C. albiceps* only had ~2% and >~10%, respectively, of the contig sequences characterized as belonging to Muller element F as X-linked. This signifies the presence of undifferentiated sex chromosomes where read coverage is not enough to characterize sex specific sequences. These same autosomal properties were also observed in all the other five

Muller elements (Figure 3.2, Table A8). Since *C. rufifacies* and *C. albiceps* belong in a family where the reproductive system of a majority of the blow fly species in it are amphogenic and display heteromorphic sex chromosomes [59, 117, 119, 237], this may suggest that monogeny and homomorphic sex chromosomes are derived rather than a representative of an ancestral state. Accordingly, this may suggest that the homomorphic nature of sex chromosomes in *C. albiceps* and *C. rufifacies* are young and in the early stages of evolution, however, additional studies will be needed to confirm this conclusion.

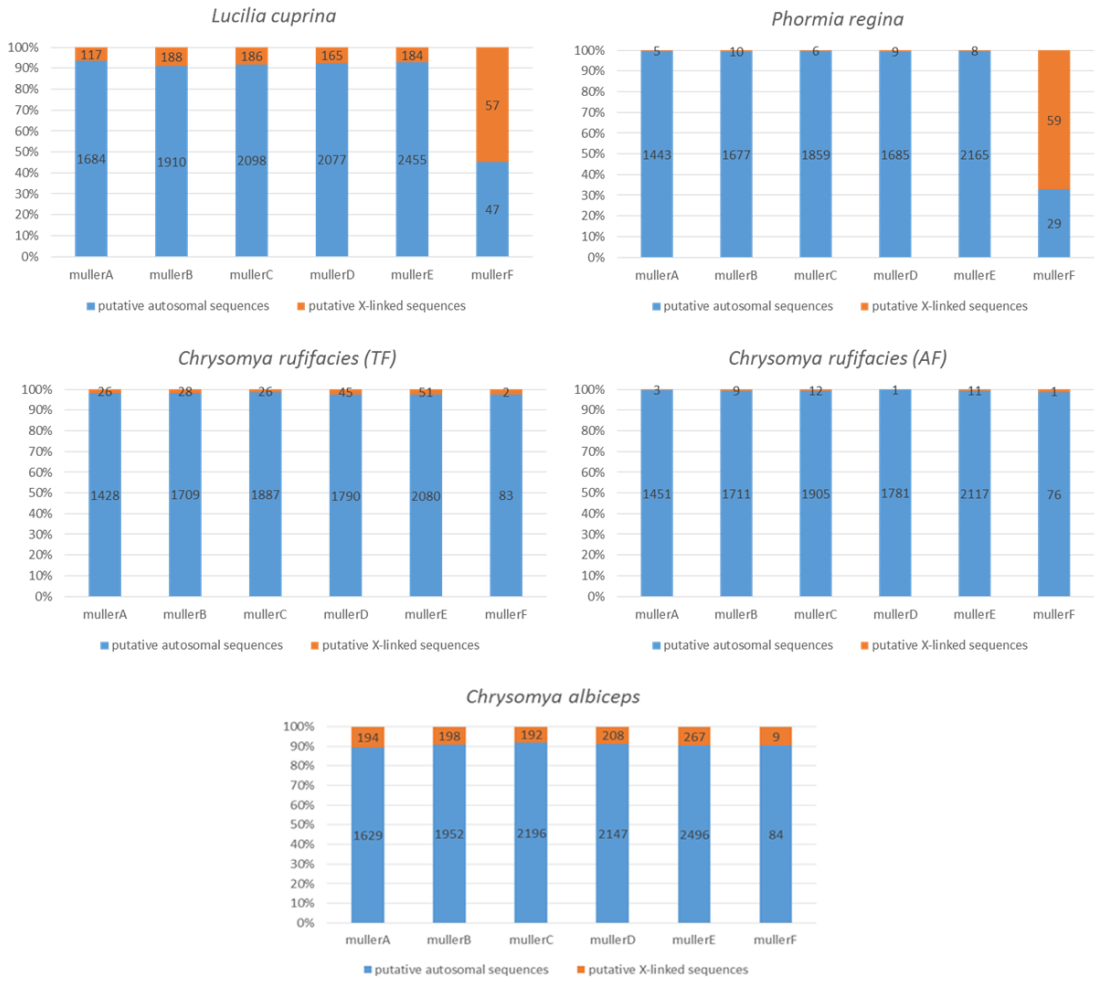


Figure 3.2. Muller element F, which is the ancestral sex chromosome in Dipteran species, is X-linked in *Phormia regina* and *Lucilia cuprina* – a characteristic expected in differentiated sex chromosomes. However, in *C. rufifacies* and *C. albiceps* (they have homomorphic sex chromosomes), Muller element F is not X-linked, suggesting undifferentiated sex chromosomes. The orange color represents the proportion of putative X-linked sequences while the blue color represents putative autosomal sequences. The number inside the bar graphs indicate the physical number of contig sequences representative of each section. Data shown is generated from female flies (AF = arrhenogenic female, TF = thelygenic female).

Of all the X-linked contig sequences that segregated as *D. melanogaster's* Muller element F, none of them was were found to be X-linked in all four blow fly species. However, one gene (the ortholog of the *Drosophila* gene legless (*lgs)),* was X-linked in three of the four blow flies which are *P. regina*, *L. cuprina* and *C. rufifacies* (AF *only*). Additionally, 5 other gene orthologs were identified as X-linked in *P. regina, L. cuprina* and *C. albiceps* (Table 3.4). Between *L. cuprina* and *P. regina,* a total of 34 contig sequences were found to be X-linked in both species (heteromorphic sex chromosomes) but not in *C. rufifacies* nor *C. albiceps* (homomorphic sex chromosomes). Table 3.4 provides a summary of the comparative analysis of these set of genes These sequences will be useful in providing insight into the evolutionary divergence of sex chromosome evolution in blow flies. A full list detailing the full Muller element F gene names can be found in Table A9.

Table 3.4. The table indicates a list of *D. melanogaster's* Muller element F genes whose orthologs in the blow flies were found to be X-linked. orthologous gene sequences of *Drosophila* Muller identified in contig sequences which segregates as Muller element F and were determined to be X-linked in each of the respective species. (Cruf = *C. rufifacies,* Calb = *C. albiceps,* Preg = *P. regina,* Lcup = *L. cuprina, AF = arrhenogenic female, TF = thelygenic female)*

Names	*total*	*D. melanogaster Muller element F gene*
Cruf AF, Lcup, Preg	1	*lgs*
Calb, Lcup, Preg	5	*CG33521, dpr7, Cals, pan, Arl4*
Lcup, Preg	34	*Actbeta, gw, Crk, unc-13, mav, CG1909, Kif3C, ey, ci, dati, zfh2, Thd1, CG33978, PlexB 4E-T, onecut, yellow-h, CG31999, MED26, CG31998, CG32006, Eph, CG31997, Ekar, Gyf, RhoGAP102A, bt, Sox102F, Ank, Asator, fd102C, CG2316, PlexA, Slip1*
Cruf TF, Preg	1	*CG11155*
Calb, Preg	1	*CG11076*
Calb, Lcup	1	*ND-49*
Preg	14	*bip2, PMCA, myo, CG1674, Gat, Zyx, Cadps, eIF4G, toy, sv, pho, Pur-alpha, PIP4K, CG11360*
Lcup	4	*Fuss, Syt7, CG32850, apolpp*
Cruf TF	1	*JYalpha*
Calb	1	*Zip102B*

3.4.2.1 Repetitive Element Analysis of Muller Elements in Blow Flies

In the genus *Drosophila*, Muller element F (chromosome 4 in *Drosophila melanogaster*) is unique and distinct from the other autosomal chromosomes because it contains a high proportion of repetitive elements (specifically transposable elements) than its euchromatic counterparts, but also a similar gene density as the other chromosomes [109, 110, 235]. Additionally, it undergoes

late replication and low rates of recombination, which are properties of heterochromatic DNA. Heterochromatic DNA, with a generally higher transposon density than euchromatic DNA [46], which explains the presence of a high density of transposable elements in *Drosophila's* Muller F's repeatome. An analysis of repetitive sequences in the Muller elements A – F of *P. regina, L. cuprina, C. rufifacies* and *C. albiceps* revealed that between ~5% - 7% of each Muller element is occupied by repetitive elements (Table A10). The repetitive landscape shows that in all the four species, Muller element F generally has a higher repeat density in total interspersed repeats, and retroelements (specifically LINEs and LTRs, –(Figure 3.3, Table A10) as compared to the other autosomal Muller elements. This is consistent with what is observed in Muller F (chromosome 4) of *D. melanogaster* [109, 110, 235] whose F element contains remnants of transposable elements [109]. Additionally, retrotransposons have been identified as the culprit behind the substantial increase in size of Muller element F in *D. ananassae* as compared to *D. melanogaster* [235], which most of it is composed of heterochromatic DNA. Transposons play a role in targeting heterochromatin formation [258]. Karyotype studies have shown that the X chromosome of *L. cuprina* is typically the largest chromosome in its karyotype and appears to be heterochromatic with few genes [236]. Consistent with this characteristic, its Muller element F has a higher density in retroelements and total interspersed repeats (Figure 3.3), as compared to the other three blow fly species, which may explain its significantly larger size.

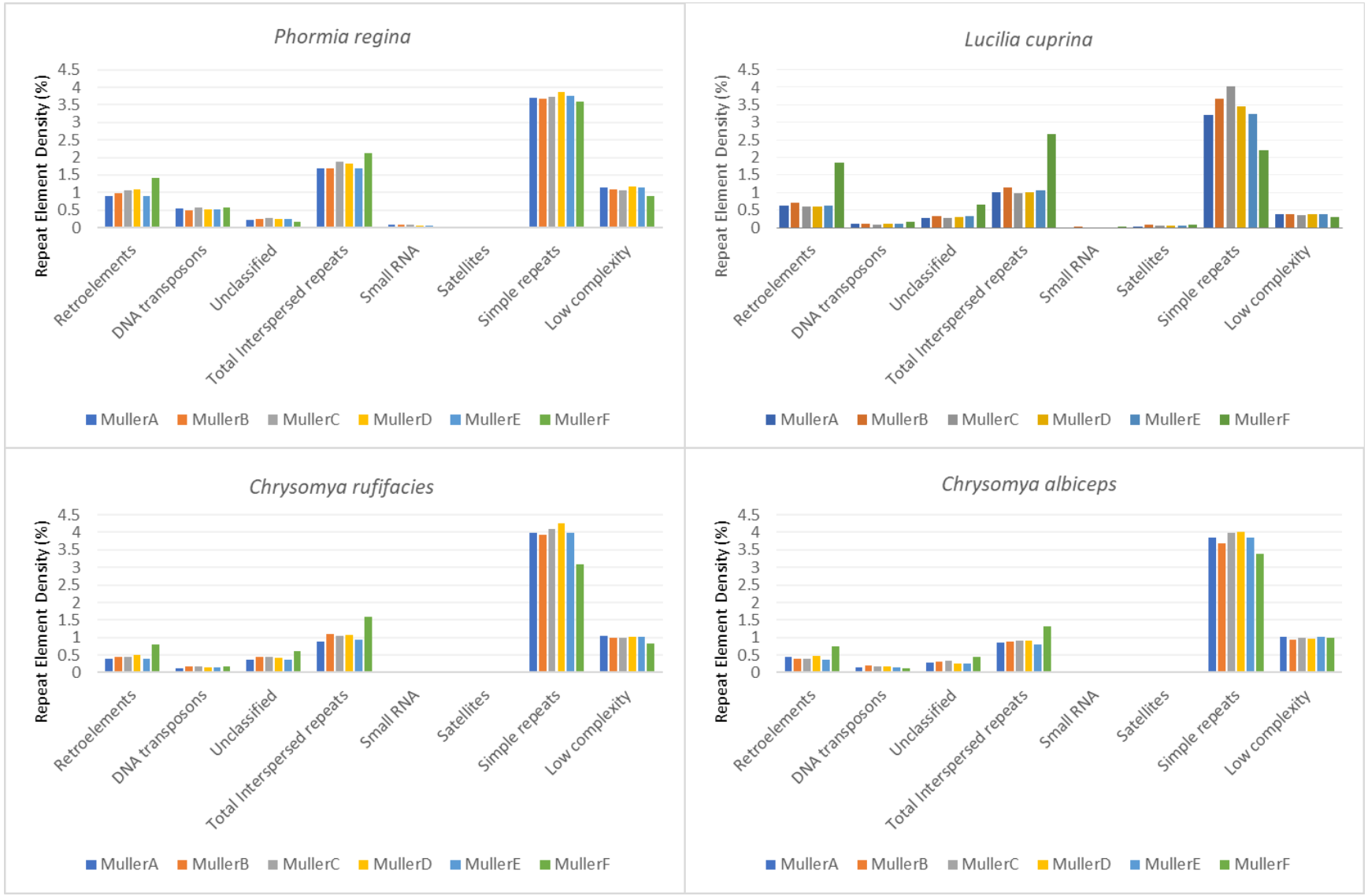


Figure 3.3. The repetitive element landscape of Muller elements in each of the four blow fly species is shown. Within each species, Muller element F generally shows a higher repeat density in retroelements (specifically LINEs and LTRs – Table A6) and total interspersed repeats.

3.5 Conclusion

This study utilized a comparative genomic approach in order to identify and characterize the sex chromosomes of four blow fly species: *Phormia regina, Chrysomya rufifacies* and *Chrysomya albiceps* all which belong to the Calliphoridae subfamily Chrysominae, and *Lucilia cuprina* which belongs to the Lucilinae subfamily. Results obtained from the genomic characterization of the sex chromosomes presents evidence that *L. cuprina* and *P. regina* do exhibit differentiated sex chromosomes. The chromosome quotient (CQ) approach, which exploits the difference in sex chromosome ploidy between males and females using next-generation sequencing data, successfully characterized heterogametic sex chromosomes in *P. regina* indicated by a smaller sized putative Y sequences as compared to the X. Additionally, annotation of the putative sex chromosomes in both *P. regina* and *L. cuprina* provided evidence of a dosage compensation mechanism due to the presence of the gene *no blokes* (*nbl*), which has been found to be important for X chromosome dosage compensation [112]. Expectedly, only a few sequences (~1.6 Mb) between the male and female genome of *C. rufifacies* was characterized as undifferentiated which is consistent with homogametic sex chromosomes.

The karyotype of the most recent common ancestor of higher Diptera (Brachycera) consists of five gene rich autosomes (named Muller elements A – E in *Drosophila*) and a smaller heterochromatic dot chromosome (Muller element F in *Drosophila*) which segregates as the sex chromosome in several other Dipteran species and has been classified as sex-linked [1, 107, 229]. A homology-based approach provided evidence that the sex chromosome of blow flies is homologous to the Dipteran ancestral sex chromosome Muller element F. Differential male and female read coverage analysis on the characterized Muller elements (A – F) showed a significant excess of X candidate sequences mainly on Muller element F in *P. regina* and *L. cuprina,* classifying Muller F as sex-linked in these two species and affirming the presence of differentiated sex chromosomes. However, *C. rufifacies* and *C. albiceps* showed no evidence of having differentiated sex chromosome as none of the Muller elements, including the ancestral sex chromosome Muller F, were classified as sex-linked. This demonstrates that *C. rufifacies* and *C. albiceps* have lost their ancestral sex chromosome and Muller element F is not the sex chromosome. It also suggests the presence of a young sex chromosome pair which is at an early stage of evolution.

Instead of a whole chromosome, a small region or a single sex-determining locus could be differentiated between the sexes contributing to the male and female phenotype. This means that the X/Y chromosome pairs would still undergo recombination, have similar gene content, and reads from the neo Y and X would still map to each other. With a young and possibly small sex-determining region, it is not expected that a male/female differential read coverage analysis would be successful in detecting differentiated sex chromosomes in these species. These results reveal that heteromorphic sex chromosomes are not a terminal end point in sex chromosome evolution. The absence of differentiated sex chromosomes within *C. rufifacies* and *C.albiceps* is unclear. However, the data obtained herein, open an avenue for more research to assess the unique evolutionary forces exposed to sex chromosomes which permit rapid sex chromosome turnover.

Sex chromosomes in numerous non – model organisms are being identified and characterized by new bioinformatic methods making genetic data from diverse sex chromosome systems easily available for analysis in an effort to gain a principal understanding of how evolutionarily labile sex chromosomes are. Blow flies hold promise for understanding the nature and evolutionary dynamics of sex chromosomes due to the presence of such a diverse sex chromosome system. The data obtained from this study will therefore reinforce the understanding of sex chromosome evolution and further refine its evolutionary history not only within the Calliphoridae family but in other Dipteran species as well.

CHAPTER 4. ASSESSING THE GENETIC DIVERGENCE OF SEX CHROMOSOMES IN FOUR BLOW FLY SPECIES

4.1 Abstract

Heteromorphic sex chromosomes are usually thought to represent an evolutionary end point in sex chromosome evolution. However, the presence and preservation of homomorphic sex chromosomes within species in lineages dominated by differentiated sex chromosomes largely remains elusive. The blow fly species *Chrysomya rufifacies* and its sister species *Chrysomya albiceps* are such an example as they possess homomorphic sex chromosomes yet many other blow fly species within the same family exhibit heteromorphic sex chromosomes. The evolutionary forces which contribute to the diversification of sex chromosomes within blow flies is relatively understudied. This study therefore utilized the ratio of synonymous (dS) and nonsynonymous (dN) substitution rates within protein coding sequences of genes located in the ancestral dipteran sex chromosome (*Drosophila melanogaster's* Muller element F) to provide preliminary insight into the evolutionary forces contributing to the divergence between homomorphic and heteromorphic sex chromosomes in blow flies. Muller F genes are usually X-linked in blow flies with differentiated heteromorphic sex chromosomes but in contrast possess autosomal properties in blow flies with undifferentiated homomorphic sex chromosomes. Our analysis demonstrate that Muller F genes within *C. rufifacies* and *C. albiceps* experienced weak selective pressures (weak negative selection) leading to a higher substitution rate which may contribute to the loss of sex chromosome properties within Muller F genes in *C. rufifacies* and *C. abliceps*; as compared to *Lucilia cuprina* and *Phormia regina* which possess differentiated sex chromosomes. Additionally, phylogenetic analysis provides some evidence that homomorphic sex chromosomes branched out of the most recent ancestral state and are therefore derived. The results obtained in this study will henceforth reinforce the understanding of sex chromosome evolution as well as offer new insight into the selective forces involved in the divergence of sex chromosomes within the Calliphoridae family.

4.2 Introduction

Genomes in males and females of most eukaryotic species are largely identical, with the only difference arising from their sex chromosomes [164, 259]. Sex chromosomes hence play a central role in fostering the evolution and development of sexual dimorphism, speciation and adaptation [1]. In many diploid eukaryotic species, the two well-known sex chromosome systems constitute of male heterogamety where the male genotype is denoted as XY and the female as XX; and female heterogamety where the female's genotype is ZW and the male's is ZZ. The accepted evolutionary theory of sex chromosomes states that sex chromosomes evolved from a pair of ordinary homologous autosomes [12]. One of the pairs (the newly formed Y/W chromosome) then acquired a sex-determining function, which could be a male-determining region initiating differentiation process between the pair [9, 11]. After the acquisition of the sex determining region (SDR), sexually antagonistic alleles beneficial to the sex-limited sex and therefore linked to the SDR begin accumulating selecting for the suppression of recombination between the now newly formed (nascent) pair of sex chromosomes [77, 126]. The non-recombining pair (Y/W chromosome) resultantly accumulates deleterious mutations becoming highly heterochromatic and leading to an accumulation of repetitive elements and a profound level of gene loss [11, 19, 85, 226]. Sequence divergence between the sex chromosomes is therefore catalyzed by suppression of recombination. In contrast, its homolog the X/Z chromosome remain functional and might evolve dosage compensation mechanism [67, 70, 76]. The end product therefore becomes a pair of highly differentiated heterogametic chromosomes which display a difference in size and gene content and a severely degenerated Y/Z chromosome [15]. Extreme cases lead to the continual loss of genes on the Y/W chromosome leading to an ultimate loss of the Y/W gene chromosome as is observed in the *Ellobius* genus of the mole voles [16]**.**

Studies have shown that sex chromosomes are evolutionarily labile and vary considerably across many eukaryotic species. These variations have been observed even among closely related species or populations of the same species [260]**.** Not all eukaryotic species exhibit differentiated heterogametic sex chromosomes. Exceptions to the old adage exists where some organisms have homomorphic sex chromosomes. Ratite birds such as emus [23, 24], mosquitos in the family Culicidae [26], most cold blooded vertebrates such as amphibians [137] and fishes [261], and many dioecious species of flowering plants [262] are examples of organisms which possess largely undifferentiated sex chromosomes

Sex chromosomes experience different evolutionary forces compared to the rest of the genome. Its evolution has partly been attributed to the presence of sexually antagonistic alleles. It has been hypothesized that sex chromosome systems which experience more sexual conflict will consequently undergo a rapid expansion of the non-recombining region increasing the rate of differentiation and degeneration of the sex-limited chromosome (Y chromosome). In the formation of a heterogametic sex chromosomes which is prevalent in many organisms, a male beneficial mutation which occurs close to a male sex determining region (or locus) has a high probability of spreading and being passed on to future generations even if it is highly detrimental to females. Linkage of the male beneficial mutation to the male determining locus increases the chances of it being transmitted to male offspring. The canonical model of sex chromosome evolution postulates that reduced recombination expands from the sex determining locus outwards [9]. Non-recombining populations accumulate deleterious mutations at a high rate via the evolutionary stochastic process Muller's ratchet [21]. Both Muller's ratchet and genetic hitchhiking are used to explain the degeneration of the Y chromosome [21]. They promote the fixation of deleterious mutations which are linked with beneficial genes which subsequently leads to reduced levels of nucleotide diversity [263] and signatures of reduced adaptation at the DNA and protein sequence levels [264].

Homomorphic sex chromosomes typically appear morphologically identical representing the assumption that they are undifferentiated. This assumption leads to the idea that these species have a small non-recombining region which has not spread beyond the sex determining region (or locus) allowing room for recombination to take place between the X and Y sex chromosomes [1]. The reason behind why some sex chromosomes remain largely undifferentiated is not well understood and remains elusive. However, a few hypotheses have been put forth to address this [13]. One of the accepted sex chromosome evolutionary theory proposes that sex chromosomes evolved from two homomorphic pairs of chromosomes. This consequently suggests that homomorphic sex chromosomes are by default young sex chromosomes in the early stages of evolving and have not yet fully degenerated [9]. However, some species have been shown to possess evolutionarily old homomorphic sex chromosomes, for example, frogs, emus and snakes [23, 25, 265] indicating that the level of differentiation should not solely be used as a predictor of evolutionary age.

Turnover of sex chromosomes (a switch in the chromosome pair recruited to determine sex) occur frequently in cold blooded vertebrates such as fish and frogs [265] for which most of them possess homomorphic sex chromosomes. Frequent sex chromosome turnovers have been hypothesized to have contributed to an increased rate of homomorphy in these organisms. In organisms which have undergone a turnover, sex chromosomes are presumptively relatively young, and the pair have not had time to substantially differentiate. Sex chromosome turnovers can occur due to the translocation or transposition of an existing male determining gene or sex-locus to an autosome [100, 266, 267]. For example, the fly *Megasalia scalaris,* whose karyotype consists of 3 chromosome pairs, is unique in that any of the 3 chromosomes can act as the sex chromosome pair in different strains [100, 267]. A sex determining factor, *M,* has been observed to move from one chromosome to any of the other two chromosomes either by translocation or transposition [100]. The now nascent sex chromosome pair shows no sign of morphological differentiation and is therefore homomorphic with a low degree of molecular genetic differentiation between the neo X and Y chromosomes [267]. Turnovers can also occur if a new gene on an autosome acquires a mutation which transforms it into a sex determining gene [268]. Substitution rates between gametologs of X and Y sex chromosomes, coupled with phylogenetic data can be used to identify the age of sex chromosomes and even trace their origin [269].

Most Dipteran flies typically have an XY sex chromosome system, and a mostly conserved karyotype [1, 229] of six chromosomal arms (2n = 12); five pairs of autosomes and a heteromorphic XX/XY sex chromosome pair . The gene contents of the chromosomes are designated as Muller elements A-F based on *Drosophila melanogaster* organization [229, 233]. One of the chromosomal arms, Muller element F, was identified as an ancestral X chromosome in many Dipteran species and considered sex-linked for over 200 million years [1, 107, 229]. Genes located within Muller element F are therefore expected to be X-linked. A whole genome analysis study within 37 species of flies identified fly species within Diptera which had genes characterized as belonging in Muller element F that were X-linked. This confirmed a level of conserved synteny of the gene elements in Muller F in Dipteran species [1]. In *Drosophila melanogaster,* Muller element F is known as the dot chromosome or chromosome 4 [107-109] however, it reverted back to an autosome but still features characteristics similar to an X chromosome [107]. Muller element A thereby became the X chromosome in *D. melanogaster* [107]. It is important to note that not all gene elements belonging to the ancestral X chromosome – Muller element F – in Dipteran species

are X-linked. For instance, in the blow fly *Lucilia cuprina,* some of the homologous genes characterized as Muller element F in *Drosophila* are X-linked within *L. cuprina* [112, 113], however, some other homologous genes lacked any X-linked properties and instead showed autosomal properties (same copy number of genomic reads in both male and female DNA) suggesting that these genes may have reverted back to autosomes [112]. Additionally, homologous Muller element F genes within insects exhibiting undifferentiated sex chromosomes possess autosomal properties as is observed in the some strains within the common house fly *Musca domestica* [114] and some other muscid flies such as the stable fly (*Stomoxys calcitrans)* and the horn fly (*Haematobia irritans)* [270].

Insects within the Dipteran family Calliphoridae are excellent models to study sex chromosome evolution as they possess varied sex chromosome systems in phylogenetically closely related species [59, 143]. They have a male heterogametic sex chromosome system (male = XY and female = XX), including heteromorphic and homomorphic sex chromosomes [59, 120]. Due to a rapid diversification caused by geographical changes and evolutionary processes within Calliphoridae [245], a divergence of sex chromosomes occurred within closely related species. For example, most of the blow fly species within the subfamily Chrysomyinae are amphogenic and possess heteromorphic sex chromosomes [59, 120, 271]. These also includes most species within the *Chrysomya* genus such as *Chrysomya megacephala*, *Chrysomya putoria*, *Chrysomya marginalis* and *Chrysomya saffranea* [59, 119]. However, *Chrysomya rufifacies* and its sister species *Chrysomya albiceps*, both members of this subfamily, are quite unique – they are monogenic (sex of offspring determined by maternal genotype) [123, 124], have two types of females, and they exhibit homomorphic sex chromosomes [59, 102, 122].

A variety of statistical tests to quantify selection pressures acting on protein coding sequences which may contribute to sequence evolution have been initiated by the advent of sequencing technologies [272]**.** The dynamics of DNA sequence evolution can be explained by calculating synonymous (dS) and nonsynonymous (dN) substitution rates [273]. One of the most widely used statistical test is the dN/dS ratio which quantifies selection pressures by comparing the substitution rates at synonymous sites (dS) which are presumed to be functionally silent and evolutionarily neutral; to the substitution rates on non-synonymous sites (dN) which are presumed to be non-silent sites and possibly are under strong selective pressures [272, 273]. Variable substitution rates may be indicative of adaptive evolution or relaxed selective constraints among

lineages [274]**.** Therefore, determining the selective forces responsible for evolution is fundamental as it enhances understanding what makes a species unique and also gives some insight into how a species may adapt into new environments [274]. Most metazoan species experience the 'faster-X effect', which is the rapid evolution of protein-coding genes on the X chromosome as compared to the autosomal chromosomes [275] . In these instances, the dN/dS ratio of protein-coding genes is predicted to be elevated on the X chromosome as compared to autosomes leading to a faster evolution rate on the X [276, 277], which may be attributed to the smaller population size of the X in a genome as compared to autosomal chromosomes [278]. This proposes that X-linked genes would evolve faster in protein sequences than on autosomes due to fixation of recessive beneficial mutations in the hemizygous state [275], or from fixation of recessive, mildly deleterious mutations via genetic drift as has been observed in some birds and aphids [277, 279]. However, some studies have shown that not all metazoan species experience the 'faster-X effect'. A comparison of protein sequence divergence (dN/dS) on the autosome vs X chromosome in the beetles (*Tribolium*) indicated an absence of a faster X-effect [280].

Not much is known about the evolutionary forces contributing to the diversification of the sex chromosomes within blow flies. This study analyzed the sex chromosome evolution of four closely related blow fly species using a comparative genomic approach. Two of the blow fly species possess differentiated sex chromosomes (*Phormia regina* and *Lucilia cuprina*), while the other two have undifferentiated sex chromosomes (*Chrysomya rufifacies* and *Chrysomya albiceps*). Substitution rate analysis (dN/dS ratio) of protein coding sequences homologous to the ancestral sex chromosome (*D. melanogaster's* Muller element F) was performed in order to examine the selective pressures exerted on sex chromosome systems with differentiated and undifferentiated sex chromosomes.

4.3 Methods

4.3.1 Substitution Rate Analysis

Three orthologous sequences from the ancestral dipteran sex chromosome Muller element F and two autosomal genes from Muller elements D and E from *Drosophila melanogaster* were extracted from four blow fly species (*Phormia regina, Lucilia cuprina, Chrysomya rufifacies, Chrysomya albiceps)* by BLASTn and tBLASTn with an E-value cutoff ≤ 1E-10. These genes

were *Calsyntenin I, Sox102F, Gawky* (all three in Muller F); *Spenito* (Muller D) and *Spartin* (Muller E). Annotation of the orthologs was performed using reference coding sequences from *Drosophila melanogaster* (FlyBase [281]), *Musca domestica* GCA_000371365.1 and *Lucilia cuprina* GCA_001187945.1 (both downloaded from NCBI). Multiple sequence alignment of the annotated protein coding sequences was performed on each set using default settings via the MUSCLE alignment plugin within MEGA-X (Molecular Evolutionary Genetics Analysis program, v10.1.7 [282]). Estimation of nonsynonymous (dN) and synonymous (dS) substitution ratios (dN/dS) per substitution type were based on the set of codon-aligned nucleotide sequences by doing a pairwise comparison of the annotated protein coding sequences for each gene in each species (both male and female). This was achieved using the program SNAP (Synonymous Non-synonymous Analysis Program, v2.1.1). The significance of the dN/dS ratios observed was calculated using the Student's T-test.

4.3.2 Phylogenetic Reconstruction

Phylogenetic analysis of the annotated protein coding gene sequences among the four blow fly species was reconstructed using Maximum Likelihood (ML) in MEGA-X v10.1.7 [282] following default settings. The phylogenetic reconstruction employed 500 rapid bootstrap replicates. The best-fit substitution model for the likelihood analysis was identified in MEGA and used for each gene as follows: *calsyntenin I and gawky* = TN93 + G, *sox102F* = HKY + G, *spartin* = T92 + G, and *spenito* = GTR + G*;* where TN93 = Tamura-Nei model, HKY = Hasegawa-Kishino-Yano model, T92 = Tamura 3-parameter model, and GTR = General Time Reversible model. Three dipteran species – *Drosophila melanogaster* (fruit fly), *Bactrocera dorsalis* (oriental fruit fly) and *Musca domestica* (the common house fly) were used as outgroups in the reconstruction of the phylogeny trees. The outgroup sequences for each of the protein coding genes were downloaded from NCBI. The respective accession numbers are as follows: Bdor_*nito*_XM_019991296, Mdom_*nito*_XM_011293877, Dmel_*nito*_FBpp0308267, Mdom_*spartin*_XM_005178237, Bdor_*spartin*_XM_011208384, Dmel_*spartin*_FBpp0300977, Dmel_*sox*_NM_001272145, Mdom_sox_XM_020039069, Mdom_*gw*_XM_011295768, Bdor_*gw*_XM_029550231, Dmel_*gw*_NM_001014691, Bdor_*cals1*_XM_011200878, and Dmel_cals1_AJ289018. (*nito = spenito, gw = gawky, cals1 = calsyntenin 1*. Mdom = *M, domestica*, Dmel = *D. melanogaster*, Bdor = *Bactrocera dorsalis.*

4.4 Results and Discussion

4.4.1 Substitution Rate Analysis

Selection pressure among protein coding gene sequences is typically inferred by estimating the evolutionary rate ratio. This widely used method is performed by calculating the ratio between nonsynonymous substitutions (dN) and synonymous substitutions (dS) [283]. Differential male and female read coverage analysis on the characterized Muller element F (as described in Chapter 3) affirmed the presence of differentiated sex chromosomes in *P. regina* and *L. cuprina*, and largely undifferentiated sex chromosomes in *C. rufifacies* and *C. albiceps*. In order to gain some insight into the evolutionary forces acting on the ancestral X chromosome, three orthologous gene sequences of *D. melanogaster's* Muller element F, in the four blow fly species, were arbitrarily chosen to analyze the substitution rates within the undifferentiated and differentiated sex chromosomes (Table 4.1). Additionally, two orthologous gene sequences from *D. melanogaster's* autosomal Muller elements D and E were also arbitrarily chosen for comparative purposes (Table 4.1). Muller element F genes are expected to experience different selective pressures in a hemizygous sex system as is observed in differentiated sex chromosomes with a degenerated Y chromosome, as compared to a non-hemizygous system in a species with undifferentiated sex chromosomes [284]. From the differential male and female read coverage analysis in Chapter 3, the genes *gawky* and *Sox102F* were X-linked in both *P. regina* and *L.cuprina* while the gene *Calsyntenin-1* was X-linked in *L. cuprina, P. regina* and *C. albiceps.* Both genes *Spartin* and *Spenito* were autosomal in all the blow flies (not X-linked).

Table 4.1. Average substitution rate ratios (dN/dS) and their significance values, calculated via a pairwise comparison between orthologs belonging to blow fly species with homomorphic (*C. albiceps* and *C. rufifacies*) and heteromorphic sex chromosomes (*P. regina* and *L. cuprina*).

			Mean dN/dS ratio		*p-value*
		Gene	Homomorphic (Undifferentiated sex chromosomes)	Heteromorphic (Differentiated sex chromosomes)	
Dipteran ancestral sex chromosome	Muller F	*Calsyntenin-1*	0.6970	0.4262	*0.003841
		Sox102F	0.6485	0.5752	*3.24E-12
		Gawky	0.7151	0.6803	0.82081
Autosomal chromosomes	Muller E	*Spartin*	0.7796	0.6862	0.51218
	Muller D	*Spenito*	0.4657	0.4536	0.84068

**denote significant p-values*

The ratio of non-synonymous to synonymous substitutions (dN/dS) is a useful measure of denoting the natural selection acting on protein-coding genes. A dN/dS ratio < 1 indicates the presence of more synonymous substitutions suggesting the presence of an evolutionary pressure to conserve the ancestral state by removing deleterious alleles (purifying, negative selection) [272, 273]. Moreover, a dN/dS ratio > 1 suggests an increase in non-synonymous substitutions, signifying an upsurge in evolutionary pressure to escape from the ancestral state (diversifying, positive selection) due to the evolution of new advantageous genetic variants which may serve a novel function [272, 273]. If the dN/dS ratio =1, it implies the presence of an equal number of synonymous and non-synonymous substitutions from the ancestral state to the modern versions of the gene sequences indicating neutral selection [272, 273]. The substitution rate values obtained from the pairwise comparison of the protein coding sequences of *cals* and *sox102F* gene orthologs in *C. rufifacies* and *C. albiceps* (undifferentiated sex chromosomes) showed a relatively higher dN/dS ratio (Table 4.1, Figure 4.1): average dN/dS values of 0.6970 and 0.6485, respectively**)** as compared to their counterparts in the differentiated sex chromosomes (Table 4.1, Figure 4.1) dN/dS values of 0.4262 and 0.5752 respectively). The differences in the substitution rates were statistically significant with *p*-values of 0.003841 and 3.24E-12 in *cals* and *sox102F*, respectively. The gene *gawky* also showed a similar pattern in the distribution of the substitution ratios (Table 4.1, Figure 4.1) with an average dN/dS ratio of 0.7151 in *C. rufifacies* and *C. albiceps* as compared to *P. regina* and *L. cuprina* (average dN/dS ratio of 0.6803). However, the results were not statistically significant *(p*-value 0.82081). This could be due to a skewed data point introduced by the pairwise comparison between the male and female *L. cuprina.* Exclusion of this data point produces a significant difference between the two groups (*p*-value 8.80747E-05). A common feature among these three genes is that they are X-linked in *P. regina* and *L. cuprina* (differentiated sex chromosomes), but are not X-linked in *C. rufifacies* and *C.albiceps.* A higher dN/dS values which is less than 0 but closer to 1 suggests that these genes are under a weak negative selection in blow fly species with undifferentiated sex chromosomes as compared to the ones with differentiated sex chromosomes. Since this observation occurs within protein coding gene sequences belonging to the ancestral sex chromosome, it provides some insight as to the almost absence of X-linked genes within the Muller F element in *C. rufifacies* and *C. albiceps*. Weakening or removal of selection which was important in the maintenance of a trait, has been observed to occur due to 'relaxed selection' in natural populations [274]. This suggests *C. rufifacies* and *C.*

albiceps may have lost the differentiated sex chromosome trait within these genes and consequently reverted to a homomorphic sex chromosome system.

The plant species *Silene latifolia* (white campion) has recently evolved sex chromosomes (young sex chromosomes) and as such many of its genes still retain functional X and Y-linked gametologs [284, 285]. A comparative study was performed between *S. latifolia* and *Silene vulgaris*, a closely related species to *S. latifolia*, which has differentiated sex chromosomes and thus hemizygous X-linked genes on the male. Genes in the non-hemizygous X (*S. latifolia)* showed a significant shift to a higher dN/dS ratio while the X-linked genes on the hemizygous X (*S. vulgaris)* which has no Y-linked gametolog showed a significantly reduced dN/dS ratio [284]. This suggested that substitution rates in X-linked genes are affected by haploid selection in hemizygous males [284]. In a heteromorphic system, where the X chromosome is hemizygous in males, partially or completely recessive alleles on the X chromosome are more exposed to selection as compared to autosomes implying that slightly deleterious and recessive alleles are more efficiently purged on the X chromosome than on an autosome [5, 286]**.** In a system with undifferentiated sex chromosomes, the assumption is that there is continued recombination between the homologous sex chromosome. Therefore, recessive mutations would be shielded from selection as would happen in autosomes To compare the result with autosomal genes, the protein coding sequences of two autosomal genes from Muller elements D and E (*spenito* and *spartin*, respectively) were arbitrarily chosen. As shown in Table 4.1 and Figure 4.1 a pairwise comparison of the distribution of dN/dS values between *C. rufifacies* and *C.albiceps* (undifferentiated sex chromosomes) and *P. regina* and *L. cuprina* (differentiated sex chromosomes) revealed no significant difference (*p*-value *spartin* = 0.51218, *spenito* = 0.84068). This suggests that substitution rate divergence within autosomal genes is similar in both blow fly species with homomorphic and heteromorphic sex chromosomes, and highlights on the differences in how selection works on sex chromosomes (Muller F) and autosomes in blow flies. The results obtained were not sufficient to conclude the presence or absence of a faster-X effect in these blow fly species. However, it will be of interest for a future study to compare a faster-X effect in both a differentiated and undifferentiated sex chromosome system of closely related species as these four blow fly species.

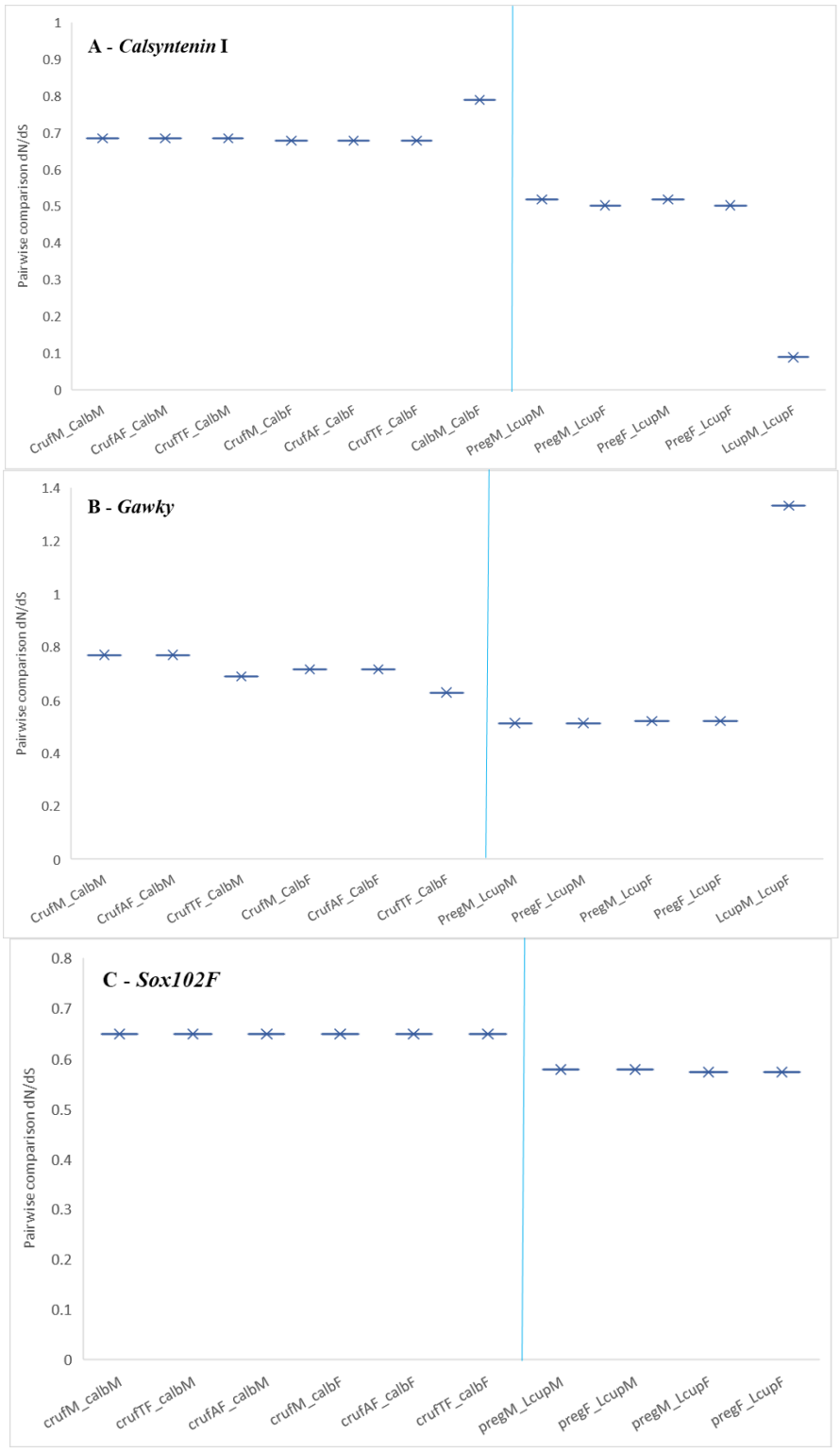
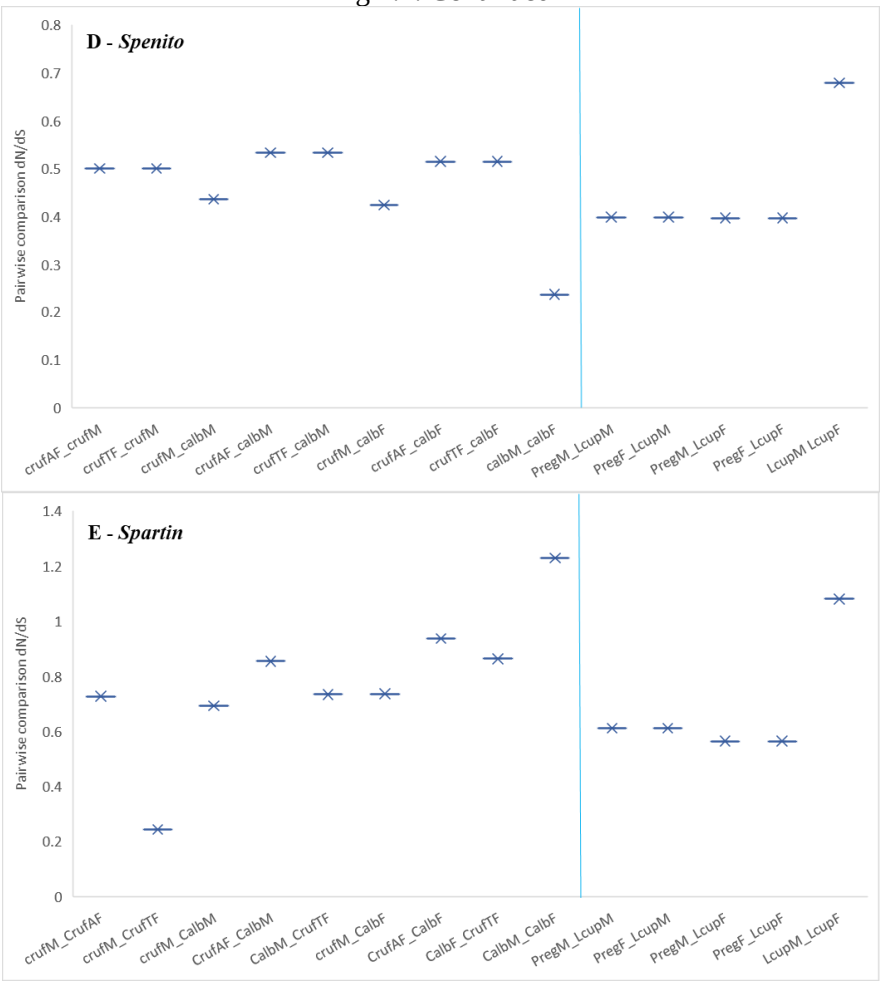


Figure 4.1. Pairwise comparison of the synonymous (dS) and non-sysnonymous (dN) ratio (dN/dS) in gene orthologs of *D. melanogaster's* Muller element F gene *calsyntenin* I (cals), *gawky* and *sox102F* between the male and female of the four blow fly species; and two autosomal genes *spartin* and *spenito* (nito). *Cruf = *C. rufifacies*, Preg = *P. regina*, Calb = *C. albiceps*, Lcup = *L. cuprina*.

Fig 4.1. Continued



4.4.2 Phylogenetic Analysis

A species (taxonomic) phylogenetic tree reflects actual evolutionary pathways and relationships among species while a gene tree typically represents evolutionary history of a set of orthologous genes included in a specific study [287]. Gene trees can be used to estimated species divergence time and ancestral population sizes however they do not necessarily follow the same topology as a species tree due to genetic polymorphisms, horizontal gene transfer, gene duplication and even gene loss [287-289]. In order to infer evolutionary relationships and identify evolutionary patterns present, the protein coding sequences of all the five genes were used for phylogenetic analysis using the maximum likelihood (ML) model. Outgroup dipteran species used in the analysis included the common housefly *Musca domestica*, the fruit fly *Drosophila melanogaster*, and the oriental fruit fly *Bactrocera dorsalis.* Within Calliphoridae, a rapid radiation of the

subfamilies Chrysomyinae and Luciliinae occurred ~17.74 and ~16.32 million years ago respectively, branching them from a common ancestor into the two groups [115]. Therefore in a phylogenetic tree based on taxonomy, *P.regina* groups together with both *C. rufifacies* and *C albiceps* as they belong to the same subfamily Chrysomyinae, while *L. cuprina* groups separately as it belongs to a different subfamily – Luciliinae [115, 290]. The phylogenetic trees of the analyzed Muller F genes – *cals, sox102F* and *gw* (Figure 4.2 A, B, C) show that these genes are closely related in *L. cuprina* and *P. regina* despite belonging in different subfamilies, while *C. albiceps* and *C, rufifacies* branched off and are grouped together. The data generated from the trees are in contrast to what is expected from a taxonomic phylogenetic tree. The distance scale bar located on each of the phylogenetic graphs represent the number of genetic differences (substitutions per site) between the sequences and how much change is reflected in the lengths of the horizontal branches. This means that overall, in these Muller F genes (Figure 4.2 A, B, C) similar selective pressures occurred on the *L. cuprina* and *P. regina* conserving Muller F as the sex chromosome while *C. albiceps* and *C. rufifacies* underwent an evolutionary event subsequently losing it as the sex chromosome. A majority of the blow fly species within the Chrysomyinae subfamily display differentiated sex chromosomes [59, 117, 119, 237] denoting that sex chromosomes in all blow flies before the derivation of the undifferentiated sex chromosomes had Muller F as the sex chromosome in the ancestral state. Therefore, undifferentiated sex chromosomes in *C. rufifacies* and *C. albiceps* are not a representative of the most recent ancestral state. The phylogenetic analyses of these three genes support this hypothesis. *Chrysomya rufifacies* and *C. albiceps* seem to have undergone genetic mutations different from *L.cuprina* and *P. regina,* since the split from a most recent common ancestor, as they are seen to branch out further from the ancestral state, and are also separately grouped.

In comparison, the phylogenetic trees of the autosomal genes *spenito* (Figure 4.2 D) has similar topology as a taxonomic phylogenetic tree which indicates that *L. cuprina* is distantly related to *P. regina, C. albiceps* and *C. rufifacies* as it has been grouped more as an outgroup as is expected taxonomically [115]. For an autosomal gene, this phylogenetic relationship would be supported by the fact that *L. cuprina* belongs to a different subfamily (Luciliinae) while *C. rufifacies, C. albiceps* and *P. regina* all belong to the same subfamily Chrysomyinae. This also suggests that the gene *spenito* undergoes similar selection pressure as in the taxonomic tree in these blow fly species. In contrast, the phylogenetic tree of the autosomal gene *spartin* (Figure 4.2 E)

displays similar branching pattern to the Muller F genes which is in contrast to the taxonomic grouping (*L, cuprina* and *P. regina* being closely related).Sequence analysis of the dN/dS ratio in *spartin* did not however provide a statistical significance on the differences in the substitution rates between the differentiated and undifferentiated sex chromosome systems. Therefore, selection observed in this gene could be for different biological functions which are not sex related. *Chrysomya rufifacies* and *albiceps* contain unique features which are not common in other related blow fly species. For example, the larvae of both *C. rufifacies* and *C. albiceps* are referred to as hairy maggot in response to the presence of spines in their larvae. In contrast, other Calliphorids have a smooth larval morphology. Additionally, the larvae of *C. rufifacies* and *C. albiceps* have been observed to be predators as they feed on other dipteran larvae as alternative food sources [291], and are also cannibalistic in nature. Both of these traits are not common in other Calliphorid larvae [237]**.**

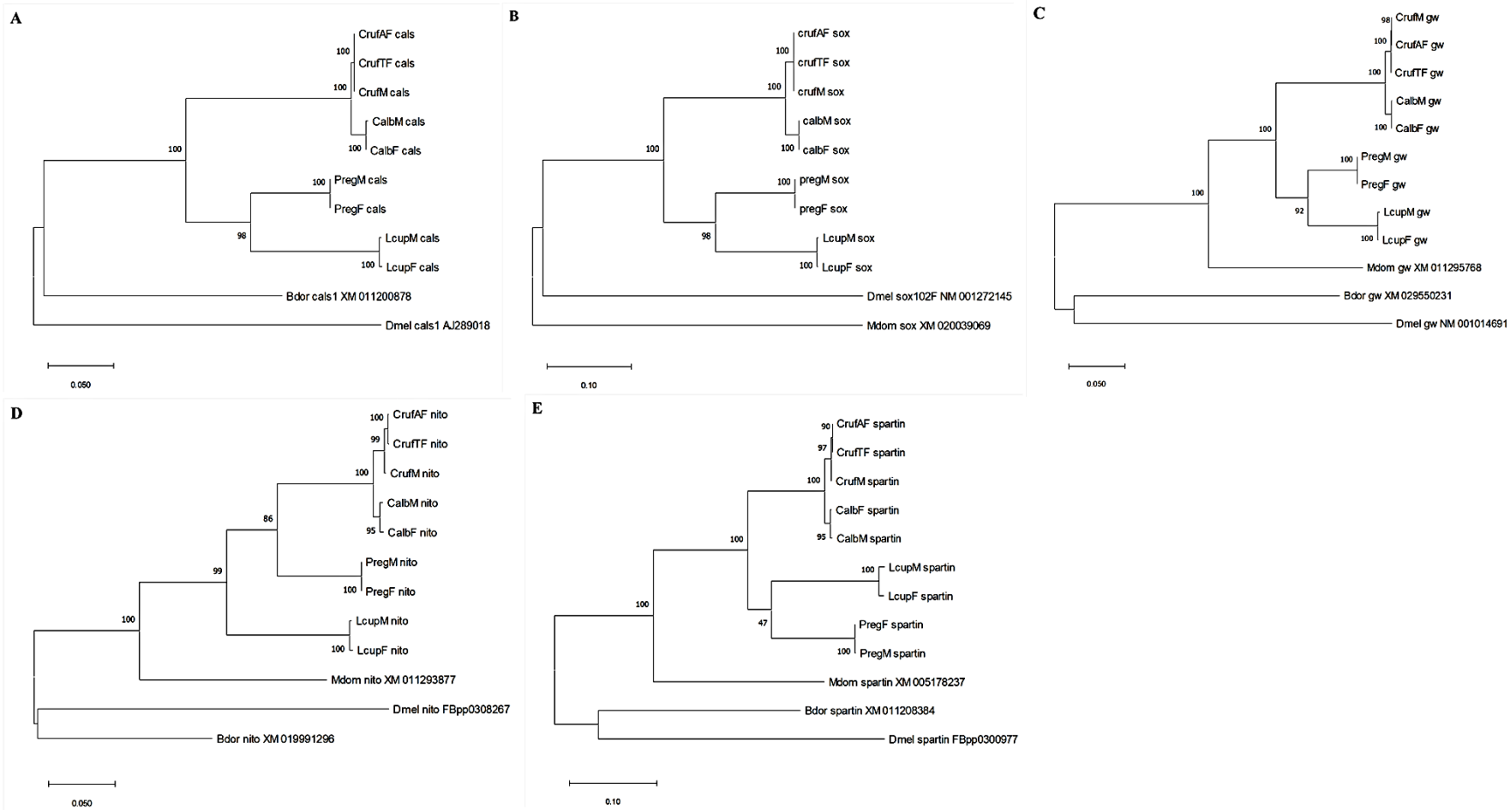


Figure 4.2. Phylogeny-based analyses of substitution rates in the genes cals (A), sox (B), gw (C), nito (D) and spartin (E), calculated by maximum likelihood analysis. A majority of the branches received between 86-100% of bootstrap support, except one branch in the spartin (E) phylogenetic tree which received poor support of 47%. Abbreviations: Preg = P. regina, Lcup = L. cuprina, Cruf = C. rufifacies – AF = arrhenogenic female – TF = thelygenic female, Calb = C. albiceps, Mdom = M. domestica, Bdor = B. dorsalis, Dmel = D. melanogaster. The 'M' and 'F' after every abbreviation = Male and Female respectively.

4.5 Conclusion

Changes in the genetic makeup of an organism can either shield genes or alleles from selection causing them to retain the original function of the genes or expose them to selective pressures which would relax or weaken selection. The results obtained in this study reveal that genes located on the ancestral sex chromosomes (Muller F) experience different selective pressures in blow fly species with differentiated sex chromosomes (where these genes are X-linked) as compared to those with largely undifferentiated sex chromosomes (where these genes are not X-linked). An elevated ratio of substitution rates (dN/dS) which was observed within protein coding sequences of genes in Muller F in *C. rufifacies* and *C. albiceps* is a characteristic reminiscent of adaptive evolution signifying the presence of a weaker negative selective pressure. A weak purifying selection within these genes, allowed genetic changes to occur in the Muller F genes which may have contributed to the rapid diversification observed within the sex chromosomes in this lineage, losing their sex chromosome trait, and departing from the most recent ancestral sex chromosome state (differentiated sex chromosome). Any sex chromosome characteristic that would have been present in the Muller F genes may have moved from this ancestral sex chromosome to any of the other autosomal chromosomes. Additionally, results obtained from the phylogenetic analysis of the Muller F genes supports the hypothesis that *C. rufifacies* and *C. albiceps* undifferentiated sex chromosomes are derived and branched out from the most recent sex chromosome ancestral state introducing a genetic difference in Muller F genes.

CHAPTER 5. CONCLUSION

The extraordinary diversity of sex chromosome evolution within eukaryotes remains a fundamental question and a major field of research. The diversity of sex chromosomes represented within eukaryotes is mind-blowing and research on sex chromosome evolution has increased and is an active field [1, 23, 60, 224, 262, 270, 292-294]. Understanding the biology and evolution behind diverse sex chromosomes is therefore important because in addition to sex determination, sex chromosomes, play a critical role in the survival and evolution of sexually reproducing organisms [5, 38]. Additionally, they are involved in the process of speciation by reducing fitness of hybrids between species [5], and they can also be used to genetically modify specific sexes in the development of viable tools for pest control [295]. Many sex chromosome systems in eukaryotic species have proven that theoretical predictions do not always match to empirical evidence. For example, highly differentiated sex chromosomes and a degenerate, gene poor Y chromosome does not necessarily represent an evolutionary dead-end [1, 25, 28, 270]. Numerous eukaryotes have been under investigation in an effort to understand this unique phenomenon [1, 25, 28, 270]. Insects within the order Diptera are one such group, as they are ubiquitous in nature and display a large variation in sex chromosomes [1].

Blow flies (Diptera: Calliphoridae) are a perfect model system to study sex chromosome evolution as they present an array of sex chromosome systems even within phylogenetically closely related species [59, 119]. The presence of both a heteromorphic sex chromosome system (degenerated and differentiated Y and a homozygous XX female [59], and a homomorphic sex chromosome system – almost morphologically identical X and Y [101, 144]) provided an opportunity to investigate sex chromosome evolution and the respective evolutionary forces contributing to their divergence. The evolution of sex chromosomes and the selective pressures involved in their divergence is relatively understudied within blow fly species. In this dissertation, I used whole-genome analysis to investigate sex chromosome structure and evolution within four blow fly species; *Lucilia cuprina* and *Phormia regina* which exhibit heteromorphic sex chromosomes, and *Chrysomya rufifacies* and *Chrysomya albiceps* which are sister species and exhibit homomorphic sex chromosomes. *Lucilia cuprina* and *Phormia regina* are amphogenic, as is the case in most blow fly species [59], where female and male offspring are produced in the same ratio. However, *Chrysomya rufifacies* and *C. albiceps* are unique – in addition to the absence

of obvious sex chromosomes, females are monogenic such that sex of the offspring is determined by the mother's genotype [101, 122, 144]. They have two types of females – one which produces exclusively male offspring (arrhenogenic females) and another which produces exclusively female offspring (thelygenic females) [101, 122]. Mating studies hypothesize that thelygenic females are heterozygous for a dominant female-determiner (*F/f*) while both arrhenogenic females and males are homozygous (*f/f*) at this same locus [121, 123]. Monogeny in these species could have evolved in order to avoid inbreeding and consequently increase the overall fitness of individuals. This would ensure that at worst any mating that would occur will only be of half siblings. In addition, to a homomorphic sex chromosome system and a unique sex determination system, the larvae of *C. rufifacies* is referred to as hairy maggot in response to the presence of spines in their larvae. This morphological departure from the normal 'smooth' larvae observed in other Calliphorids is purported to have occurred in order to reduce cannibalism and protect them from eating each other as they have been found to be cannibalistic in nature, and also avoid being preyed upon by other species making them powerful ecological competitors for carrion resource [237]**.**

Genomic sequences and the assembled genomes of the blow fly *Chrysomya rufifacies,* specifically the male, thelygenic female (female producing females), and arrhenogenic females (male producing females) are presented firsthand. Additionally, the male and female genomes of *L. cuprina* and *C. albiceps* used in this study were also locally assembled but are currently unpublished. *Phormia regina* male and female genomes were from a previous publication [148] . These individual genomes have provided access to Calliphorid genomic data which will facilitate the investigation of diverse sex determining mechanisms and sex chromosome evolution within Calliphoridae. A comparative genomic analysis of the sex chromosome structures of the blow fly species *P. regina, L. cuprina, C. rufifacies* and *C. albiceps* was executed. Muller element analysis confirmed that *Drosophila melanogaster's* Muller element F (dot chromosome), which is the ancestral dipteran sex chromosome, has remained the sex chromosome in some blow flies. Differential male and female read coverage analysis on genomic sequences representing Muller element F affirmed the presence of differentiated (heteromorphic) sex chromosomes in *L. cuprina* and *P. regina*; and undifferentiated (homomorphic) sex chromosomes in *C. rufifacies* and *C. albiceps.* This data supports previous evidence from cytological karyotypic analysis displaying highly differentiated sex chromosomes (an X and Y chromosome which differ in shape and size) in *P. regina* and *L. cuprina;* and a sex chromosome pair which appears morphologically

indistinguishable in *C. rufifacies* and *C. albiceps* [59, 101, 144]. Additionally, the specific Muller F genes which were established to be either X-linked (2:1 female to male read coverage) or autosomal (1:1 female to male read coverage) in all the four blow fly species were identified. Evidence of a dosage compensation system was observed in *P. regina* and *L. cuprina* as the gene *no blokes* (*nbl*) which is required for normal levels of gene expression of most X-linked genes in *L. cuprina* [76] was annotated in the characterized putative sex chromosomes in both species. In *C. rufifacies* and *C. albiceps* only a few genomic sequences aligned to the *nbl* gene. Since *C. rufifacies* and *C. albiceps* have homomorphic sex chromosomes, it is probable that dosage compensation does not function the same way as it does in a species with heteromorphic sex chromosomes.

Finally, a close examination of genes hypothesized to be a part of the ancestral dipteran sex chromosome (Muller F) in the four blow fly species was performed in order to gain some insight into how sex chromosomes diverged causing some species to maintain a differentiated sex chromosome system and others to lose the 'normal' sex chromosome and instead develop undifferentiated sex chromosomes. Substitution ratio (dN/dS) analysis results revealed that different selection pressures were exerted on the same ancestral sex chromosome genes in homomorphic and heteromorphic sex chromosome system. A statistically significant lower substitution ratio (relatively stronger negative selection) was observed in X-linked Muller F protein coding sequences in the heteromorphic sex chromosome, indicating constraints such that emergence of deleterious alleles within the coding sequences of these genes are disfavored and purged. In contrast, a moderately higher dN/dS ratio (weaker negative selection) was observed in these same genes, which in contrast, are autosomal in a homomorphic sex chromosome system. This suggests partial relaxation of purifying selection on these genes in *C. rufifacies* and *C. albiceps*, rendering them susceptible to favorable changes (mutations) in the protein coding sequences. The absence of a differentiated sex chromosome system in cooperation with the presence of genetic changes within Muller F genes in *C. rufifacies* and *C. albiceps* provide clues indicating a probable loss of a 'normal' sex chromosome in these two species. Absence of a heteromorphic sex chromosome system proposes a dominant sex determining locus could be present within an autosomal chromosome. Consequently, this would transform the autosomes into a neo-sex chromosome which is in its early stages of evolution and thus not easily detected using differential male/female read coverage analysis.

One avenue of future work is to perform an analysis of single nucleotide polymorphisms (SNPs) in males versus females in order to gain some insight into identifying the Muller elements (if any) which may have recently been transformed into the neo-sex chromosomes within *C. rufifacies* and *C. albiceps.* The expectation is that recently formed X and Y sex chromosomes still share considerable homology, but due to an early X-Y differentiation, they should have an increased SNP density – indicative of young evolving sex chromosomes. Moreover, additional substitution ratio analysis on the remaining annotated Muller F genes will be needed to obtain a complete picture on the selection forces exerted on their protein coding genes. This would provide additional data to gain a robust understanding in the evolution of sex chromosomes within blowflies and the role of natural selection in this process.

Sex chromosomes in numerous non – model organisms are being identified and characterized by new bioinformatic methods. This effort makes it possible to address some unanswered questions in sex chromosome evolution, as well as getting a principal understanding of how evolutionarily labile sex chromosomes are. Blow flies hold promise for understanding the nature and evolutionary dynamics of sex chromosomes. Although the understanding of sex chromosome evolution within blow flies is still in its infancy, this dissertation provides a foundational set of genomic tools to reinforce the study of sex chromosome evolution within Calliphoridae and other species with diverse sex chromosome systems.

REFERENCES

1. Vicoso, B. and D. Bachtrog, *Numerous transitions of sex chromosomes in Diptera.* PLoS Biol, 2015. **13**(4): p. e1002078.

2. Wilson Sayres, M.A., *Genetic Diversity on the Sex Chromosomes.* Genome Biol Evol, 2018. **10**(4): p. 1064-1078.

3. Arbeitman, M.N., et al., *The genetics of sex: exploring differences.* Genetics, 2014. **197**(2): p. 527-9.

4. Bachtrog, D., et al., *Are all sex chromosomes created equal?* Trends Genet, 2011. **27**(9): p. 350-7.

5. Johnson, N.A. and J. Lachance, *The genetics of sex chromosomes: evolution and implications for hybrid incompatibility.* Ann N Y Acad Sci, 2012. **1256**: p. E1-22.

6. Zhou, Q., et al., *Complex evolutionary trajectories of sex chromosomes across bird taxa.* Science, 2014. **346**(6215): p. 1246338.

7. Sahara, K., A. Yoshido, and W. Traut, *Sex chromosome evolution in moths and butterflies.* Chromosome Res, 2012. **20**(1): p. 83-94.

8. Charlesworth, D. and B. Charlesworth, *Sex chromosomes: evolution of the weird and wonderful.* Curr Biol, 2005. **15**(4): p. R129-31.

9. Charlesworth, D., B. Charlesworth, and G. Marais, *Steps in the evolution of heteromorphic sex chromosomes.* Heredity (Edinb), 2005. **95**(2): p. 118-28.

10. Traut, W., K. Sahara, and F. Marec, *Sex chromosomes and sex determination in Lepidoptera.* Sex Dev, 2007. **1**(6): p. 332-46.

11. Charlesworth, B., *The evolution of sex chromosomes.* Science, 1991. **251**(4997): p. 1030-3.

12. Bull, J.J., *Evolution of Sex Determining Mechanisms*. 1983, Menlo Park California: Benjamin Cummings Publishing Company.

13. Wright, A.E., et al., *How to make a sex chromosome.* Nature Communications, 2016. **7**.

14. Bachtrog, D., et al., *Sex determination: why so many ways of doing it?* PLoS Biol, 2014. **12**(7): p. e1001899.

15. Bachtrog, D., *A dynamic view of sex chromosome evolution.* Curr Opin Genet Dev, 2006. **16**(6): p. 578-85.

16. Matveevsky, S., et al., *Chromosomal Evolution in Mole Voles Ellobius (Cricetidae, Rodentia): Bizarre Sex Chromosomes, Variable Autosomes and Meiosis.* Genes (Basel), 2017. **8**(11).

17. Palacios-Gimenez, O.M., et al., *Tracking the evolution of sex chromosome systems in Melanoplinae grasshoppers through chromosomal mapping of repetitive DNA sequences.* BMC Evol Biol, 2013. **13**: p. 167.

18. Carvalho, A.B., *Origin and evolution of the Drosophila Y chromosome.* Curr Opin Genet Dev, 2002. **12**(6): p. 664-8.

19. Bachtrog, D., *Sex chromosome evolution: molecular aspects of Y-chromosome degeneration in Drosophila.* Genome Res, 2005. **15**(10): p. 1393-401.

20. Bachtrog, D., et al., *Genomic degradation of a young Y chromosome in Drosophila miranda.* Genome Biol, 2008. **9**(2): p. R30.

21. Kaiser, V.B. and B. Charlesworth, *Muller's ratchet and the degeneration of the Drosophila miranda neo-Y chromosome.* Genetics, 2010. **185**(1): p. 339-48.

22. Nozawa, M., et al., *Accelerated pseudogenization on the neo-X chromosome in Drosophila miranda.* Nat Commun, 2016. **7**: p. 13659.

23. Vicoso, B., V.B. Kaiser, and D. Bachtrog, *Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution.* Proc Natl Acad Sci U S A, 2013. **110**(16): p. 6453-8.

24. Yazdi, H.P. and H. Ellegren, *Old but not (so) degenerated--slow evolution of largely homomorphic sex chromosomes in ratites.* Mol Biol Evol, 2014. **31**(6): p. 1444-53.

25. Vicoso, B., et al., *Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation.* PLoS Biol, 2013. **11**(8): p. e1001643.

26. Hall, A.B., et al., *Insights into the preservation of the homomorphic sex-determining chromosome of Aedes aegypti from the discovery of a male-biased gene tightly linked to the M-locus.* Genome Biol Evol, 2014. **6**(1): p. 179-91.

27. Becak, W. and M.L. Becak, *Cytotaxonomy and chromosomal evolution in Serpentes.* Cytogenetics, 1969. **8**(4): p. 247-62.

28. Fontaine, A., et al., *Extensive Genetic Differentiation between Homomorphic Sex Chromosomes in the Mosquito Vector, Aedes aegypti.* Genome Biol Evol, 2017. **9**(9): p. 2322-2335.

29. Toups, M.A. and M.W. Hahn, *Retrogenes reveal the direction of sex-chromosome evolution in mosquitoes.* Genetics, 2010. **186**(2): p. 763-6.

30. Ogawa, A., K. Murata, and S. Mizuno, *The location of Z- and W-linked marker genes and sequence on the homomorphic sex chromosomes of the ostrich and the emu.* Proc Natl Acad Sci U S A, 1998. **95**(8): p. 4415-8.

31. Shetty, S., D.K. Griffin, and J.A. Graves, *Comparative painting reveals strong chromosome homology over 80 million years of bird evolution.* Chromosome Res, 1999. **7**(4): p. 289-95.

32. Ellegren, H. and J. Parsch, *The evolution of sex-biased genes and sex-biased gene expression.* Nat Rev Genet, 2007. **8**(9): p. 689-98.

33. Connallon, T. and L.L. Knowles, *Intergenomic conflict revealed by patterns of sex-biased gene expression.* Trends Genet, 2005. **21**(9): p. 495-9.

34. Pala, I., et al., *Evidence of a neo-sex chromosome in birds.* Heredity (Edinb), 2012. **108**(3): p. 264-72.

35. Carvalho, A.B. and A.G. Clark, *Y chromosome of D. pseudoobscura is not homologous to the ancestral Drosophila Y.* Science, 2005. **307**(5706): p. 108-10.

36. Charlesworth, B., *The evolution of chromosomal sex determination.* Novartis Found Symp, 2002. **244**: p. 207-19; discussion 220-4, 253-7.

37. Blackmon, H., L. Ross, and D. Bachtrog, *Sex Determination, Sex Chromosomes, and Karyotype Evolution in Insects.* J Hered, 2017. **108**(1): p. 78-93.

38. Abbott, J.K., A.K. Norden, and B. Hansson, *Sex chromosome evolution: historical insights and future perspectives.* Proc Biol Sci, 2017. **284**(1854).

39. Orr, H.A. and Y. Kim, *An adaptive hypothesis for the evolution of the Y chromosome.* Genetics, 1998. **150**(4): p. 1693-8.

40. Bachtrog, D., *Accumulation of Spock and Worf, two novel non-LTR retrotransposons, on the neo-Y chromosome of Drosophila miranda.* Mol Biol Evol, 2003. **20**(2): p. 173-81.

41. Nanda, I., et al., *Early stages of sex chromosome differentiation in fish as analysed by simple repetitive DNA sequences.* Chromosoma, 1992. **101**(5-6): p. 301-10.

42. Hobza, R., et al., *An accumulation of tandem DNA repeats on the Y chromosome in Silene latifolia during early stages of sex chromosome evolution.* Chromosoma, 2006. **115**(5): p. 376-82.

43. Matsubara, K., et al., *Sex chromosome evolution in snakes inferred from divergence patterns of two gametologous genes and chromosome distribution of sex chromosome-linked repetitive sequences.* Zoological Lett, 2016. **2**(1): p. 19.

44. Hobza, R., et al., *Impact of repetitive DNA on sex chromosome evolution in plants.* Chromosome Res, 2015. **23**(3): p. 561-70.

45. Kraemer, C. and E.R. Schmidt, *The sex determining region of Chironomus thummi is associated with highly repetitive DNA and transposable elements.* Chromosoma, 1993. **102**(8): p. 553-62.

46. Smith, C.D., et al., *The Release 5.1 annotation of Drosophila melanogaster heterochromatin.* Science, 2007. **316**(5831): p. 1586-91.

47. Steinemann, M. and S. Steinemann, *Enigma of Y chromosome degeneration: neo-Y and neo-X chromosomes of Drosophila miranda a model for sex chromosome evolution.* Genetica, 1998. **102-103**(1-6): p. 409-20.

48. Steinemann, M. and S. Steinemann, *Degenerating Y chromosome of Drosophila miranda: a trap for retrotransposons.* Proc Natl Acad Sci U S A, 1992. **89**(16): p. 7591-5.

49. Cermak, T., et al., *Survey of repetitive sequences in Silene latifolia with respect to their distribution on sex chromosomes.* Chromosome Res, 2008. **16**(7): p. 961-76.

50. Kejnovsky, E., et al., *The role of repetitive DNA in structure and evolution of sex chromosomes in plants.* Heredity (Edinb), 2009. **102**(6): p. 533-41.

51. Okada, S., et al., *The Y chromosome in the liverwort Marchantia polymorpha has accumulated unique repeat sequences harboring a male-specific gene.* Proc Natl Acad Sci U S A, 2001. **98**(16): p. 9454-9.

52. Liu, Z., et al., *A primitive Y chromosome in papaya marks incipient sex chromosome evolution.* Nature, 2004. **427**(6972): p. 348-52.

53. Kondo, M., et al., *Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka.* Genome Res, 2006. **16**(7): p. 815-26.

54. Peichel, C.L., et al., *The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome.* Curr Biol, 2004. **14**(16): p. 1416-24.

55. Steinemann, S. and M. Steinemann, *Y chromosomes: born to be destroyed.* Bioessays, 2005. **27**(10): p. 1076-83.

56. Chuong, E.B., N.C. Elde, and C. Feschotte, *Regulatory activities of transposable elements: from conflicts to benefits.* Nat Rev Genet, 2017. **18**(2): p. 71-86.

57. Munoz-Lopez, M. and J.L. Garcia-Perez, *DNA transposons: nature and applications in genomics.* Curr Genomics, 2010. **11**(2): p. 115-28.

58. Ellegren, H., *Sex-chromosome evolution: recent progress and the influence of male and female heterogamety.* Nat Rev Genet, 2011. **12**(3): p. 157-66.

59. Ullerich, F.H. and M. Schottke, *Karyotypes, constitutive heterochromatin, and genomic DNA values in the blowfly genera Chrysomya, Lucilia, and Protophormia (Diptera: Calliphoridae).* Genome, 2006. **49**(6): p. 584-97.

60. Otto, S.P., et al., *About PAR: the distinct evolutionary dynamics of the pseudoautosomal region.* Trends Genet, 2011. **27**(9): p. 358-67.

61. Graves, J.A., M.J. Wakefield, and R. Toder, *The origin and evolution of the pseudoautosomal regions of human sex chromosomes.* Hum Mol Genet, 1998. **7**(13): p. 1991-6.

62. Lahn, B.T., N.M. Pearson, and K. Jegalian, *The human Y chromosome, in the light of evolution.* Nat Rev Genet, 2001. **2**(3): p. 207-16.

63. Raudsepp, T. and B.P. Chowdhary, *The horse pseudoautosomal region (PAR): characterization and comparison with the human, chimp and mouse PARs.* Cytogenet Genome Res, 2008. **121**(2): p. 102-9.

64. Mangs, A.H. and B.J. Morris, *The human pseudoautosomal region (PAR): Origin, function and future.* Current Genomics, 2007. **8**(2): p. 129-136.

65. Burgoyne, P.S., M.J. Sutcliffe, and S.K. Mahadevaiah, *The role of unpaired sex chromosomes in spermatogenic failure.* Andrologia, 1992. **24**(1): p. 17-20.

66. Mohandas, T.K., et al., *Role of the pseudoautosomal region in sex-chromosome pairing during male meiosis: meiotic studies in a man with a deletion of distal Xp.* Am J Hum Genet, 1992. **51**(3): p. 526-33.

67. Steinemann, M., S. Steinemann, and B.M. Turner, *Evolution of dosage compensation.* Chromosome Res, 1996. **4**(3): p. 185-90.

68. Ercan, S., *Mechanisms of x chromosome dosage compensation.* J Genomics, 2015. **3**: p. 1-19.

69. Cloutier, J.M. and J.M. Turner, *Meiotic sex chromosome inactivation.* Curr Biol, 2010. **20**(22): p. R962-3.

70. Charlesworth, B., *Sex chromosomes: evolving dosage compensation.* Curr Biol, 1998. **8**(25): p. R931-3.

71. Lee, H., et al., *Dosage-Dependent Expression Variation Suppressed on the Drosophila Male X Chromosome.* G3 (Bethesda), 2018. **8**(2): p. 587-598.

72. Franke, A., et al., *Evidence that MSL-mediated dosage compensation in Drosophila begins at blastoderm.* Development, 1996. **122**(9): p. 2751-60.

73. Marin, I., et al., *The dosage compensation system of Drosophila is co-opted by newly evolved X chromosomes.* Nature, 1996. **383**(6596): p. 160-3.

74. Larsson, J., et al., *Painting of fourth in genus Drosophila suggests autosome-specific gene regulation.* Proc Natl Acad Sci U S A, 2004. **101**(26): p. 9728-33.

75. Larsson, J., et al., *Painting of fourth, a chromosome-specific protein in Drosophila.* Proc Natl Acad Sci U S A, 2001. **98**(11): p. 6273-8.

76. Brockdorff, N. and B.M. Turner, *Dosage compensation in mammals.* Cold Spring Harb Perspect Biol, 2015. **7**(3): p. a019406.

77. Bachtrog, D., *Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration.* Nat Rev Genet, 2013. **14**(2): p. 113-24.

78. Charlesworth, D., *Young sex chromosomes in plants and animals.* New Phytol, 2019. **224**(3): p. 1095-1107.

79. Gabriel, W., M. Lynch, and R. Burger, *Muller's Ratchet and Mutational Meltdowns.* Evolution, 1993. **47**(6): p. 1744-1757.

80. Kondrashov, A.S., *Muller's ratchet under epistatic selection.* Genetics, 1994. **136**(4): p. 1469-73.

81. Howe, D.K. and D.R. Denver, *Muller's Ratchet and compensatory mutation in Caenorhabditis briggsae mitochondrial genome evolution.* BMC Evol Biol, 2008. **8**: p. 62.

82. Gordo, I. and B. Charlesworth, *The degeneration of asexual haploid populations and the speed of Muller's ratchet.* Genetics, 2000. **154**(3): p. 1379-87.

83. Gordo, I. and B. Charlesworth, *The speed of Muller's ratchet with background selection, and the degeneration of Y chromosomes.* Genet Res, 2001. **78**(2): p. 149-61.

84. Stephan, W., L. Chao, and J.G. Smale, *The advance of Muller's ratchet in a haploid asexual population: approximate solutions based on diffusion theory.* Genet Res, 1993. **61**(3): p. 225-31.

85. Charlesworth, B., *The role of background selection in shaping patterns of molecular evolution and variation: evidence from variability on the Drosophila X chromosome.* Genetics, 2012. **191**(1): p. 233-46.

86. Kaiser, V.B. and B. Charlesworth, *The effects of deleterious mutations on evolution in non-recombining genomes.* Trends Genet, 2009. **25**(1): p. 9-12.

87. Barton, N.H., *Genetic hitchhiking.* Philos Trans R Soc Lond B Biol Sci, 2000. **355**(1403): p. 1553-62.

88. Rice, W.R., *Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome.* Genetics, 1987. **116**(1): p. 161-7.

89. Cranston, P.S. and P.J. Gullan, *The Insects: An outline of Entomology*. 4 ed. 2010: Wiley-Blackwell.

90. Kaiser, V.B. and D. Bachtrog, *Evolution of sex chromosomes in insects.* Annu Rev Genet, 2010. **44**: p. 91-112.

91. Sanchez, L., *Sex-determining mechanisms in insects.* Int J Dev Biol, 2008. **52**(7): p. 837-56.

92. Scott, J.G., et al., *Genome of the house fly, Musca domestica L., a global vector of diseases with adaptations to a septic environment.* Genome Biol, 2014. **15**(10): p. 466.

93. Salz, H.K. and J.W. Erickson, *Sex determination in Drosophila: The view from the top.* Fly (Austin), 2010. **4**(1): p. 60-70.

94. Saccone, G., A. Pane, and L.C. Polito, *Sex determination in flies, fruitflies and butterflies.* Genetica, 2002. **116**(1): p. 15-23.

95. Hewitt, G.M. and B. John, *Inter-population sex chromosome polymorphism in the grasshopper Podisma pedestris. II. Population parameters.* Chromosoma, 1972. **37**(1): p. 23-42.

96. John, B. and G.M. Hewitt, *Inter-population sex chromosome polymorphism in the grasshopper Podisma pedestris. I. Fundamental facts.* Chromosoma, 1970. **31**(3): p. 291-308.

97. Meisel, R.P., P.J. Delclos, and J.R. Wexler, *The X chromosome of the German cockroach, Blattella germanica, is homologous to a fly X chromosome despite 400 million years divergence.* bioRxiv, 2019: p. 279737.

98. Shearman, D.C., *The evolution of sex determination systems in dipteran insects other than Drosophila.* Genetica, 2002. **116**(1): p. 25-43.

99. Chirino, M.G., et al., *Comparative study of mitotic chromosomes in two blowflies, Luciliasericata and L.cluvia (Diptera, Calliphoridae), by C- and G-like banding patterns and rRNA loci, and implications for karyotype evolution.* Comp Cytogenet, 2015. **9**(1): p. 103-18.

100. Traut, W., *Sex determination in the fly Megaselia scalaris, a model system for primary steps of sex chromosome evolution.* Genetics, 1994. **136**(3): p. 1097-104.

101. Ullerich, F.H., *Monogene Fortpflanzung bei der Fliege Chrysomyia albiceps.* Zeitschrift für Naturforschung, 1958. **13b**: p. 473-474.

102. Ullerich, F.H., *Analysis of Sex Determination in the Monogenic Blowfly Chrysomya rufifacies by Pole Cell Transplantation.* Molecular & General Genetics, 1984. **193**: p. 479-487.

103. Pane, A., et al., *The transformer gene in Ceratitis capitata provides a genetic basis for selecting and remembering the sexual fate.* Development, 2002. **129**(15): p. 3715-25.

104. Concha, C. and M.J. Scott, *Sexual development in Lucilia cuprina (Diptera, Calliphoridae) is controlled by the transformer gene.* Genetics, 2009. **182**(3): p. 785-98.

105. Meisel, R.P., et al., *Is Multifactorial Sex Determination in the House Fly, Musca domestica (L.), Stable Over Time?* J Hered, 2016. **107**(7): p. 615-625.

106. Hamm, R.L., R.P. Meisel, and J.G. Scott, *The evolving puzzle of autosomal versus Y-linked male determination in Musca domestica.* G3 (Bethesda), 2014. **5**(3): p. 371-84.

107. Vicoso, B. and D. Bachtrog, *Reversal of an ancient sex chromosome to an autosome in Drosophila.* Nature, 2013. **499**(7458): p. 332-5.

108. Schaeffer, S.W., *Muller "Elements" in Drosophila: How the Search for the Genetic Basis for Speciation Led to the Birth of Comparative Genomics.* Genetics, 2018. **210**(1): p. 3-13.

109. Leung, W., et al., *Drosophila muller f elements maintain a distinct set of genomic properties over 40 million years of evolution.* G3 (Bethesda), 2015. **5**(5): p. 719-40.

110. Riddle, N.C. and S.C.R. Elgin, *The Drosophila Dot Chromosome: Where Genes Flourish Amidst Repeats.* Genetics, 2018. **210**(3): p. 757-772.

111. Leung, W., et al., *Evolution of a distinct genomic domain in Drosophila: comparative analysis of the dot chromosome in Drosophila melanogaster and Drosophila virilis.* Genetics, 2010. **185**(4): p. 1519-34.

112. Davis, R.J., et al., *no blokes Is Essential for Male Viability and X Chromosome Gene Expression in the Australian Sheep Blowfly.* Curr Biol, 2018.

113. Linger, R.J., E.J. Belikoff, and M.J. Scott, *Dosage Compensation of X-Linked Muller Element F Genes but Not X-Linked Transgenes in the Australian Sheep Blowfly.* PLoS One, 2015. **10**(10): p. e0141544.

114. Meisel, R.P., C.A. Gonzales, and H. Luu, *The house fly Y Chromosome is young and minimally differentiated from its ancient X Chromosome partner.* Genome Res, 2017. **27**(8): p. 1417-1426.

115. Junqueira, A.C.M., et al., *Large-scale mitogenomics enables insights into Schizophora (Diptera) radiation and population diversity.* Scientific Reports, 2016. **6**.

116. Bedo, D.G. and G.G. Foster, *Cytogenetic mapping of the male-determining region of Lucilia cuprina (Diptera: Calliphoridae) Chromosoma.* Chromosoma, 1985. **92**: p. 344-350.

117. Batista, M.R., et al., *Photographic map of the polytene chromosomes of Cochliomyia hominivorax.* Med Vet Entomol, 2009. **23 Suppl 1**: p. 92-7.

118. Childress, D., *Polytene chromosomes and linkage group-chromosome correlations in the Australian sheep blowfly Lucilia cuprina (Diptera: Calliphoridae).* Chromosoma, 1969. **26**(2): p. 208-14.

119. Parise-Maltempi, P.P. and R.M. Avancini, *C-banding and FISH in chromosomes of the blow flies Chrysomya megacephala and Chrysomya putoria (Diptera, Calliphoridae).* Mem Inst Oswaldo Cruz, 2001. **96**(3): p. 371-7.

120. Boyes, J., *Somatic chromosomes of higher diptera: V. Interspecific and intraspecific variation in the Calliphoridae.* Canadian Journal of Zoology, 1961. **39**(4): p. 549-570.

121. Scott, M.J., M.L. Pimsler, and A.M. Tarone, *Sex determination mechanisms in the Calliphoridae (blow flies).* Sex Dev, 2014. **8**(1-3): p. 29-37.

122. Ullerich, F.H., *Die genetische Grundlage der Monogenie bei der Schmeißfliege Chrysomya rufifacies (Calliphoridae, Diptera).* Molecular and General Genetics, 1973. **125**(2): p. 157-172.

123. Ullerich, F.H., *[Identification of the genetic sex chromosomes in the monogenic blowfly Chrysomya rufifacies (Calliphoridae, Diptera)].* Chromosoma, 1975. **50**(4): p. 393-419.

124. Ullerich, F.H., *Production of male and female offspring in the strictly monogenic fly Chrysomya rufifacies after ovary transplantation.* Naturwissenschaften, 1977. **64**(5): p. 277-8.

125. Barton, N.H. and B. Charlesworth, *Why sex and recombination?* Science, 1998. **281**(5385): p. 1986-90.

126. Bergero, R. and D. Charlesworth, *The evolution of restricted recombination in sex chromosomes.* Trends Ecol Evol, 2009. **24**(2): p. 94-102.

127. Kubat, Z., et al., *Microsatellite accumulation on the Y chromosome in Silene latifolia.* Genome, 2008. **51**(5): p. 350-6.

128. Charlesworth, B., P. Sniegowski, and W. Stephan, *The evolutionary dynamics of repetitive DNA in eukaryotes.* Nature, 1994. **371**(6494): p. 215-20.

129. Schutt, C. and R. Nothiger, *Structure, function and evolution of sex-determining systems in Dipteran insects.* Development, 2000. **127**(4): p. 667-77.

130. Hall, A.B., et al., *SEX DETERMINATION. A male-determining factor in the mosquito Aedes aegypti.* Science, 2015. **348**(6240): p. 1268-70.

131. Sanchez, L., *Sex-determining mechanisms in insects.* International Journal of Developmental Biology, 2008. **52**: p. 837-856.

132. Nigro, R.G., M.C.C. Campos, and A.L.P. Perondini, *Temperature and the progeny sex-ratio in Sciara ocellaris (Diptera, Sciaridae).* Genetics and Molecular Biology, 2007. **30**(1): p. 152-158.

133. Conte, M.A., et al., *A high quality assembly of the Nile Tilapia (Oreochromis niloticus) genome reveals the structure of two sex determination regions.* BMC Genomics, 2017. **18**(1): p. 341.

134. Zhang, J., et al., *Genomics of sex determination.* Curr Opin Plant Biol, 2014. **18**: p. 110-6.

135. Zhou, Q., et al., *Deciphering neo-sex and B chromosome evolution by the draft genome of Drosophila albomicans.* BMC Genomics, 2012. **13**: p. 109.

136. Sarre, S.D., T. Ezaz, and A. Georges, *Transitions between sex-determining systems in reptiles and amphibians.* Annu Rev Genomics Hum Genet, 2011. **12**: p. 391-406.

137. Eggert, C., *Sex determination: the amphibian models.* Reprod Nutr Dev, 2004. **44**(6): p. 539-49.

138. Miura, I., *Sex Determination and Sex Chromosomes in Amphibia.* Sex Dev, 2017. **11**(5-6): p. 298-306.

139. Schmid, M. and C. Steinlein, *Sex chromosomes, sex-linked genes, and sex determination in the vertebrate class amphibia.* EXS, 2001(91): p. 143-76.

140. Luckenbach, J.A., et al., *Sex determination in flatfishes: Mechanisms and environmental influences.* Semin Cell Dev Biol, 2009. **20**(3): p. 256-63.

141. Bedo, D.G., *Differential sex chromosome replication and dosage compensation in polytene trichogen cells of Lucilia cuprina (Diptera: Calliphoridae).* Chromosoma, 1982. **87**(1): p. 21-32.

142. Picard, C.J., J.S. Johnston, and A.M. Tarone, *Genome sizes of forensically relevant Diptera.* Journal of Medical Entomology, 2012. **49**(1): p. 192-197.

143. Ullerich, F.H., *Geschlechtschromosomen und Geschlechtsbestimmung bei einigen Calliphorinen (Calliphoridae, Diptera). .* Chromosoma, 1963. **14**: p. 45-110.

144. Ullerich, F.H., *Identification of the genetic sex chromosomes in the monogenic blowfly Chrysomya rufifacies (Calliphoridae, Diptera).* Chromosoma, 1975. **50**(4): p. 393-419.

145. Roy, D.N. and L.B. Siddons, *On the life history and bionomics of Chrysomyia rufifacies Macq. (Order Diptera, Family Calliphoridae).* Parasitology, 1939. **31**(4): p. 442-447.

146. Kirchhoff, C. and V. Schroeren, *Monogenic reproduction allows comparison of protein patterns of female and male predetermined ovaries and embryos in Chrysomya rufifacies (Diptera, Calliphoridae).* Comp Biochem Phys B, 1986. **85**: p. 693-699.

147. Ullerich, F.H., *Analysis of the Predetermining Effect of a Sex Realizer by Ovary Transplantations in the Monogenic Fly Chrysomya rufifacies.* Wilhelm Roux's Archives of Developmental Biology, 1980. **188**: p. 37-43.

148. Andere, A.A., et al., *Genome sequence of Phormia regina Meigen (Diptera: Calliphoridae): implications for medical, veterinary and forensic research.* BMC Genomics, 2016. **17**(1): p. 842.

149. Xue, W., et al., *L_RNA_scaffolder: scaffolding genomes with transcripts.* BMC Genomics, 2013. **14**: p. 604.

150. Gurevich, A., et al., *QUAST: quality assessment tool for genome assemblies.* Bioinformatics, 2013. **29**(8): p. 1072-5.

151. Simao, F.A., et al., *BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.* Bioinformatics, 2015. **31**(19): p. 3210-2.

152. Cantarel, B.L., et al., *MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes.* Genome Res, 2008. **18**(1): p. 188-96.

153. Conesa, A., et al., *Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.* Bioinformatics, 2005. **21**(18): p. 3674-6.

154. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.

155. Wang, Y., et al., *OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species.* Nucleic Acids Res, 2015. **43**(W1): p. W78-84.

156. Hall, A.B., et al., *Six novel Y chromosome genes in Anopheles mosquitoes discovered by independently sequencing males and females.* BMC Genomics, 2013. **14**: p. 273.

157. Fu, L., et al., *CD-HIT: accelerated for clustering the next-generation sequencing data.* Bioinformatics, 2012. **28**(23): p. 3150-2.

158. Huang, Y., et al., *CD-HIT Suite: a web server for clustering and comparing biological sequences.* Bioinformatics, 2010. **26**(5): p. 680-2.

159. Camacho, C., et al., *BLAST+: architecture and applications.* BMC Bioinformatics, 2009. **10**: p. 421.

160. Supek, F., et al., *REVIGO summarizes and visualizes long lists of gene ontology terms.* PLoS One, 2011. **6**(7): p. e21800.

161. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biol, 2010. **11**(10): p. R106.

162. Smit, A.F.A., R. Hubley, and P. Green, *RepeatMasker*. 2013-2015.

163. Sze, S.H., et al., *A scalable and memory-efficient algorithm for de novo transcriptome assembly of non-model organisms.* Bmc Genomics, 2017. **18**.

164. Picard, C.J., J.S. Johnston, and A.M. Tarone, *Genome sizes of forensically relevant Diptera.* J Med Entomol, 2012. **49**(1): p. 192-7.

165. Adams, M.D., et al., *The genome sequence of Drosophila melanogaster.* Science, 2000. **287**(5461): p. 2185-95.

166. Anstead, C.A., et al., *Lucilia cuprina genome unlocks parasitic fly biology to underpin future interventions.* Nature Communications, 2015. **6**.

167. Arabidopsis Genome, I., *Analysis of the genome sequence of the flowering plant Arabidopsis thaliana.* Nature, 2000. **408**(6814): p. 796-815.

168. Bennett, M.D., et al., *Comparisons with Caenorhabditis (approximately 100 Mb) and Drosophila (approximately 175 Mb) using flow cytometry show genome size in Arabidopsis to be approximately 157 Mb and thus approximately 25% larger than the Arabidopsis genome initiative estimate of approximately 125 Mb.* Ann Bot, 2003. **91**(5): p. 547-57.

169. Anstead, C.A., et al., *A blow to the fly - Lucilia cuprina draft genome and transcriptome to support advances in biology and biotechnology.* Biotechnol Adv, 2016. **34**(5): p. 605-620.

170. Blanc, G., et al., *Extensive duplication and reshuffling in the Arabidopsis genome.* Plant Cell, 2000. **12**(7): p. 1093-101.

171. Blanc, G., K. Hokamp, and K.H. Wolfe, *A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome.* Genome Res, 2003. **13**(2): p. 137-44.

172. Schoenfelder, K.P. and D.T. Fox, *The expanding implications of polyploidy.* J Cell Biol, 2015. **209**(4): p. 485-91.

173. Hare, E.E. and J.S. Johnston, *Genome Size Determination Using Flow Cytometry of Propidium Iodide-Stained Nuclei.* , in *Molecular Methods for Evolutionary Genetics. Methods in Molecular Biology (Methods and Protocols)*, R.M. Orgogozo V., Editor. 2012, Humana Press. p. 3-12.

174. Attrill, H., et al., *FlyBase: establishing a Gene Group resource for Drosophila melanogaster.* Nucleic Acids Res, 2016. **44**(D1): p. D786-92.

175. Bownes, M., M. Dempster, and M. Blair, *The regulation of yolk protein gene expression in Drosophila melanogaster.* Ciba Found Symp, 1983. **98**: p. 63-79.

176. Negre, B. and P. Simpson, *Diversity of transposable elements and repeats in a 600 kb region of the fly Calliphora vicina.* Mob DNA, 2013. **4**(1): p. 13.

177. Wrischnik, L.A., et al., *Recruitment of the proneural gene scute to the Drosophila sex-determination pathway.* Genetics, 2003. **165**(4): p. 2007-27.

178. Martinez, A. and M. Bownes, *The specificity of yolk protein uptake in cyclorrhaphan diptera is conserved through evolution.* J Mol Evol, 1992. **35**(5): p. 444-53.

179. Barnett, T., et al., *The isolation and characterization of Drosophila yolk protein genes.* Cell, 1980. **21**(3): p. 729-38.

180. Scott, M.J., et al., *Organisation and expression of a cluster of yolk protein genes in the Australian sheep blowfly, Lucilia cuprina.* Genetica, 2011. **139**(1): p. 63-70.

181. Tarone, A.M., et al., *Genetic variation in the Yolk protein expression network of Drosophila melanogaster: sex-biased negative correlations with longevity.* Heredity, 2012. **109**(4): p. 226-234.

182. Tarone, A.M., Y.M. Nasser, and S.V. Nuzhdin, *Genetic variation for expression of the sex determination pathway genes in Drosophila melanogaster.* Genetics Research, 2005. **86**(1): p. 31-40.

183. Burtis, K.C. and B.S. Baker, *Drosophila Doublesex Gene Controls Somatic Sexual-Differentiation by Producing Alternatively Spliced Messenger-Rnas Encoding Related Sex-Specific Polypeptides.* Cell, 1989. **56**(6): p. 997-1010.

184. Majewska, M.M., et al., *Yolk proteins in the male reproductive system of the fruit fly Drosophila melanogaster: spatial and temporal patterns of expression.* Insect Biochem Mol Biol, 2014. **47**: p. 23-35.

185. Burtis, K.C., et al., *The doublesex proteins of Drosophila melanogaster bind directly to a sex-specific yolk protein gene enhancer.* EMBO J, 1991. **10**(9): p. 2577-82.

186. Concha, C., et al., *Organization and expression of the Australian sheep blowfly (Lucilia cuprina) hsp23, hsp24, hsp70 and hsp83 genes.* Insect Mol Biol, 2012. **21**(2): p. 169-80.

187. Marshall, O.J. and V.R. Harley, *Identification of an interaction between SOX9 and HSP70.* FEBS Lett, 2001. **496**(2-3): p. 75-80.

188. He, Y., et al., *Identification of a testis-enriched heat shock protein and fourteen members of Hsp70 family in the swamp eel.* PLoS One, 2013. **8**(6): p. e65269.

189. Chebbi, M.A., et al., *The Genome of Armadillidium vulgare (Crustacea, Isopoda) Provides Insights into Sex Chromosome Evolution in the Context of Cytoplasmic Sex Determination.* Molecular Biology and Evolution, 2019. **36**(4): p. 727-741.

190. Morran, L.T., et al., *Running with the Red Queen: host-parasite coevolution selects for biparental sex.* Science, 2011. **333**(6039): p. 216-8.

191. Sturtevant, A.H. and E. Novitski, *The Homologies of the Chromosome Elements in the Genus Drosophila.* Genetics, 1941. **26**(5): p. 517-41.

192. Boyes, J.W. and H.M. Slatis, *Somatic chromosomes of higher Diptera. IV. A biometrical study of the chromosomes of Hylemya.* Chromosoma, 1954. **6**(6-7): p. 479-88.

193. Jurka, J., et al., *Repetitive sequences in complex genomes: structure and evolution.* Annu Rev Genomics Hum Genet, 2007. **8**: p. 241-59.

194. Schmidt, A.L. and L.M. Anderson, *Repetitive DNA elements as mediators of genomic change in response to environmental cues.* Biol Rev Camb Philos Soc, 2006. **81**(4): p. 531-43.

195. Olafson, P.U., S. Aksoy, and e. al., *Functional genomics of the stable fly, Stomoxys calcitrans, reveals mechanisms underlying reproduction, host interactions, and novel targets for pest control.* BioRxiv, 2019.

196. Martin, A., et al., *A transposon-induced epigenetic change leads to sex determination in melon.* Nature, 2009. **461**(7267): p. 1135-8.

197. Sahebi, M., et al., *Contribution of transposable elements in the plant's genome.* Gene, 2018. **665**: p. 155-166.

198. Anstead, C.A., et al., *Lucilia cuprina genome unlocks parasitic fly biology to underpin future interventions.* Nat Commun, 2015. **6**: p. 7344.

199. Cockburn, A.F. and S.E. Mitchell, *Repetitive DNA Interspersion Patterns in Diptera* Arch. Insect Biochem. Physiol., 1989. **10**(2): p. 105-113.

200. Li, S., et al., *The genomic and functional landscapes of developmental plasticity in the American cockroach.* Nat Commun, 2018. **9**(1): p. 1008.

201. Gempe, T. and M. Beye, *Function and evolution of sex determination mechanisms, genes and pathways in insects.* Bioessays, 2011. **33**(1): p. 52-60.

202. Green, R.F., G. Gordh, and B.A. Hawkins, *Precise sex ratios in highly inbred parasitic wasps.* The American Naturalist, 1982. **120**(5): p. 653-665.

203. Tabadkani, S.M., M. Khansefid, and A. Ashouri, *Monogeny, a neglected mechanism of inbreeding avoidance in small populations of gall midges.* Entomologia Experimentalis Et Applicata, 2011. **140**(1): p. 77-84.

204. Benatti, T.R., et al., *A neo-sex chromosome that drives postzygotic sex determination in the hessian fly (Mayetiola destructor).* Genetics, 2010. **184**(3): p. 769-77.

205. Dubendorfer, A. and M. Hediger, *The female-determining gene F of the housefly, Musca domestica, acts maternally to regulate its own zygotic activity.* Genetics, 1998. **150**(1): p. 221-226.

206. Tomberlin, J.K., et al., *A Review of Bacterial Interactions With Blow Flies (Diptera: Calliphoridae) of Medical, Veterinary, and Forensic Importance.* Annals of the Entomological Society of America, 2017. **110**(1): p. 19-36.

207. Stevens, J.R., *The evolution of myiasis in blowflies (Calliphoridae).* Int J Parasitol, 2003. **33**(10): p. 1105-13.

208. Byrd, J.H. and J.L. Castner, eds. *Forensic Entomology: The utility of arthropods in legal investigations.* 2nd Edition ed. 2010, CRC Press: Boca Raton, FL. 681.

209. Byrd, J.H. and J.K. Tomberlin, *Forensic Entomology: The Utility of Arthropods in Legal Investigations, Third Edition*. 2020, Boca Raton, FL: CRC Press, Taylor & Francis Groiup.

210. Tarone, A.M. *Genomic tools to reduce error in PMI estimates derived from Entomological Evidence*. 2016 [cited 2018 April 18]; Available from: https://www.ncjrs.gov/pdffiles1/nij/grants/250086.pdf.

211. Payne, J.A., *A summer carrion study of the baby pig Sus scrofa Linnaeus.* Ecology, 1965. **46**(5): p. 592-602.

212. Benbow, M.E., J.K. Tomberlin, and A.M. Tarone, *Carrion ecology, evolution, and their applications.* 2015, Boca Raton, FL: CRC Press. 577 pages.

213. Benbow, M.E., et al., *Necrobiome framework for bridging decomposition ecology of autotrophically and heterotrophically derived organic matter.* Ecological Monographs, 2019. **89**(1): p. e01331.

214. Scott, M.J., et al., *Genetic control of screwworm using transgenic male-only strains.* Transgenic Research, 2018. **27**(5): p. 474-474.

215. Heinrich, J.C. and M.J. Scott, *A repressible female-specific lethal genetic system for making transgenic insect strains suitable for a sterile-release program.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(15): p. 8229-8232.

216. Clausen, S. and F.-H. Ullerich, *Sequence homology between a polytene band in the genetic sex chromosomes of Chrysomya rufifacies and the daughterless gene of Drosophila melanogaster.* Naturwissenschaften, 1990. **77**(3): p. 137-138.

217. Cabrera, C.V. and M.C. Alonso, *Transcriptional Activation by Heterodimers of the Achaete Scute and Daughterless Gene-Products of Drosophila.* Embo Journal, 1991. **10**(10): p. 2965-2973.

218. Deshpande, G., J. Stukey, and P. Schedl, *Scute (Sis-B) Function in Drosophila Sex Determination.* Molecular and Cellular Biology, 1995. **15**(8): p. 4430-4440.

219. Tanaka-Matakatsu, M., et al., *Daughterless homodimer synergizes with Eyeless to induce Atonal expression and retinal neuron differentiation.* Developmental Biology, 2014. **392**(2): p. 256-265.

220. Hu, Y., et al. *FlyBi (BDGP/CCSB/DRSC) Preliminary fly interactome data.* 2017 [cited 2020 7-April]; Available from: http://flybi.hms.harvard.edu/results.php.

221. Blake, J.A. and M.R. Ziman, *Pax genes: regulators of lineage specification and progenitor cell maintenance.* Development, 2014. **141**(4): p. 737-51.

222. Otto, S.P., *The evolutionary enigma of sex.* Am Nat, 2009. **174 Suppl 1**: p. S1-S14.

223. Mank, J.E., *Sex chromosomes and the evolution of sexual dimorphism: lessons from the genome.* Am Nat, 2009. **173**(2): p. 141-50.

224. Gammerdinger, W.J., et al., *Novel Sex Chromosomes in 3 Cichlid Fishes from Lake Tanganyika.* J Hered, 2018. **109**(5): p. 489-500.

225. Gammerdinger, W.J. and T.D. Kocher, *Unusual Diversity of Sex Chromosomes in African Cichlid Fishes.* Genes (Basel), 2018. **9**(10).

226. Charlesworth, B. and D. Charlesworth, *The degeneration of Y chromosomes.* Philos Trans R Soc Lond B Biol Sci, 2000. **355**(1403): p. 1563-72.

227. van Doorn, G.S., *Intralocus sexual conflict.* Ann N Y Acad Sci, 2009. **1168**: p. 52-71.

228. Sinclair, A.H., et al., *A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif.* Nature, 1990. **346**(6281): p. 240-4.

229. Muller, H.J., *Bearing of the Drosophila work on systematics.* The new systematics, 1940: p. 185-268.

230. Foster, G., et al., *Autosomal genetic maps of the Australian Sheep Blowfly, Lucilia cuprina dorsalis R.-D. (Diptera: Calliphoridae), and possible correlations with the linkage maps of Musca domestica L. and Drosophila melanogaster (Mg.).* Genetical Research, 1981. **37**(1): p. 55-69.

231. Sved, J.A., et al., *Extraordinary conservation of entire chromosomes in insects over long evolutionary periods.* Evolution, 2016. **70**(1): p. 229-34.

232. Dubendorfer, A., et al., *Musca domestica, a window on the evolution of sex-determining mechanisms in insects.* Int J Dev Biol, 2002. **46**(1): p. 75-9.

233. White, M.J., *Cytological evidence on the phylogeny and classification of the Diptera.* Evolution, 1949. **3**(3): p. 252-61.

234. Landeen, E.L. and D.C. Presgraves, *Evolution: from autosomes to sex chromosomes--and back.* Curr Biol, 2013. **23**(18): p. R848-50.

235. Leung, W., et al., *Retrotransposons Are the Major Contributors to the Expansion of the Drosophila ananassae Muller F Element.* G3-Genes Genomes Genetics, 2017. **7**(8): p. 2439-2460.

236. Bedo, D.G., *C, Q and H-banding in the analysis of Y chromosome rearrangements in Lucilia cuprina (Wiedemann) (Diptera: Calliphoridae).* Chromosoma, 1980. **77**(3): p. 299-308.

237. Singh, B., H. Kurahashi, and J.D. Wells, *Molecular phylogeny of the blowfly genus Chrysomya.* Med Vet Entomol, 2011. **25**(2): p. 126-34.

238. Whitworth, T., *Keys to the genera and species of blow flies (Diptera : Calliphoridae) of America North of Mexico.* Proceedings of the Entomological Society of Washington 2006. **108**: p. 689-725.

239. Williams, K.A., C.S. Richards, and M.H. Villet, *Predicting the geographic distribution of Lucilia sericata and Lucilia cuprina (Diptera: Calliphoridae) in South Africa.* African Invertebrates, 2014. **55**(1): p. 157-170.

240. Wardhaugh, K.G., et al., *Estimating the incidence of fly myiases in Australian sheep flocks: development of a weather-driven regression model.* Med Vet Entomol, 2007. **21**(2): p. 153-67.

241. Byrd, J.H. and J.L. Castner, *Forensic Entomology: The Utility of Arthropods in Legal Investigations*. 2001: CRC Press.

242. Williams, K. and M.H. Villet, *Ancient and modern hybridization between Lucilia sericata and L. cuprina (Diptera: Calliphoridae).* European Journal of Entomology, 2013. **110**(2): p. 187-196.

243. Byrd, J.H. and J.L. Castner, *Forensic Entomology: The utility of arthropods in legal investigations*. 2 ed. 2010, Boca Raton, FL: CRC Press.

244. Tellam, R.L. and V.M. Bowles, *Control of blowfly strike in sheep: current strategies and future prospects.* Int J Parasitol, 1997. **27**(3): p. 261-73.

245. Junqueira, A.C., et al., *Large-scale mitogenomics enables insights into Schizophora (Diptera) radiation and population diversity.* Sci Rep, 2016. **6**: p. 21762.

246. Palmer, D.H., et al., *How to identify sex chromosomes and their turnover.* Molecular Ecology, 2019. **28**(21): p. 4709-4724.

247. Moghadam, H.K., et al., *W chromosome expression responds to female-specific selection.* Proc Natl Acad Sci U S A, 2012. **109**(21): p. 8207-11.

248. Jeffries, D.L., et al., *A rapid rate of sex-chromosome turnover and non-random transitions in true frogs.* Nat Commun, 2018. **9**(1): p. 4088.

249. Wells, J.D. and F.A. Sperling, *Molecular phylogeny of Chrysomya albiceps and C. rufifacies (Diptera: Calliphoridae).* J Med Entomol, 1999. **36**(3): p. 222-6.

250. Koblmuller, S., et al., *Cichlid evolution: lessons in diversification.* Int J Evol Biol, 2011. **2011**: p. 847969.

251. Takahashi, T. and S. Koblmuller, *The adaptive radiation of cichlid fish in lake tanganyika: a morphological perspective.* Int J Evol Biol, 2011. **2011**: p. 620754.

252. Brawand, D., et al., *The genomic substrate for adaptive radiation in African cichlid fish.* Nature, 2014. **513**(7518): p. 375-381.

253. Li, F., et al., *Insect genomes: progress and challenges.* Insect Mol Biol, 2019. **28**(6): p. 739-758.

254. Izumi, S., et al., *Yolk Proteins from Insect Eggs - Structure, Biosynthesis and Programmed Degradation during Embryogenesis.* Journal of Insect Physiology, 1994. **40**(9): p. 735-746.

255. Ding, S., et al., *Large-scale analysis reveals that the genome features of simple sequence repeats are generally conserved at the family level in insects.* BMC Genomics, 2017. **18**(1): p. 848.

256. Bourque, G., et al., *Ten things you should know about transposable elements.* Genome Biology, 2018. **19**.

257. Platt, R.N., 2nd, L. Blanco-Berdugo, and D.A. Ray, *Accurate Transposable Element Annotation Is Vital When Analyzing New Genome Assemblies.* Genome Biol Evol, 2016. **8**(2): p. 403-10.

258. Grewal, S.I. and S.C. Elgin, *Transcription and RNA interference in the formation of heterochromatin.* Nature, 2007. **447**(7143): p. 399-406.

259. Hjelmen, C.E., et al., *Genome Size Evolution within and between the Sexes.* J Hered, 2019. **110**(2): p. 219-228.

260. Tree of Sex, C., *Tree of Sex: a database of sexual systems.* Sci Data, 2014. **1**: p. 140015.

261. Devlin, R.H. and Y. Nagahama, *Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences.* Aquaculture, 2002. **208**(3-4): p. 191-364.

262. Charlesworth, D., *Plant sex chromosome evolution.* J Exp Bot, 2013. **64**(2): p. 405-20.

263. Bachtrog, D. and B. Charlesworth, *Reduced adaptation of a non-recombining neo-Y chromosome.* Nature, 2002. **416**(6878): p. 323-6.

264. Bartolome, C. and B. Charlesworth, *Evolution of amino-acid sequences and codon usage on the Drosophila miranda neo-sex chromosomes.* Genetics, 2006. **174**(4): p. 2033-44.

265. Stock, M., et al., *Ever-young sex chromosomes in European tree frogs.* PLoS Biol, 2011. **9**(5): p. e1001062.

266. Woram, R.A., et al., *Comparative genome analysis of the primary sex-determining locus in salmonid fishes.* Genome Res, 2003. **13**(2): p. 272-80.

267. Willhoeft, U. and W. Traut, *Molecular differentiation of the homomorphic sex chromosomes in Megaselia scalaris (Diptera) detected by random DNA probes.* Chromosoma, 1990. **99**(4): p. 237-42.

268. Roberts, R.B., J.R. Ser, and T.D. Kocher, *Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes.* Science, 2009. **326**(5955): p. 998-1001.

269. Cortez, D., et al., *Origins and functional evolution of Y chromosomes across mammals.* Nature, 2014. **508**(7497): p. 488-93.

270. Meisel, R.P., et al., *Sex Chromosome Evolution in Muscid Flies.* G3 (Bethesda), 2020.

271. Bedo, D.G., *Polytene chromosomes of the Old World screwworm fly (Chrysomya bezziana) and its evolutionary relationships with Lucilia cuprina and Cochliomyia hominivorax (Diptera: Calliphoridae).* Genome, 1992. **35**(2): p. 294-303.

272. Kryazhimskiy, S. and J.B. Plotkin, *The population genetics of dN/dS.* PLoS Genet, 2008. **4**(12): p. e1000304.

273. Yang, Z. and R. Nielsen, *Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models.* Mol Biol Evol, 2000. **17**(1): p. 32-43.

274. Lahti, D.C., et al., *Relaxed selection in the wild.* Trends Ecol Evol, 2009. **24**(9): p. 487-96.

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

276. Mank, J.E., E. Axelsson, and H. Ellegren, *Fast-X on the Z: rapid evolution of sex-linked genes in birds.* Genome Res, 2007. **17**(5): p. 618-24.

277. Mank, J.E., K. Nam, and H. Ellegren, *Faster-Z evolution is predominantly due to genetic drift.* Mol Biol Evol, 2010. **27**(3): p. 661-70.

278. Presgraves, D.C., *Sex chromosomes and speciation in Drosophila.* Trends Genet, 2008. **24**(7): p. 336-43.

279. Parsch, J. and H. Ellegren, *The evolutionary causes and consequences of sex-biased gene expression.* Nat Rev Genet, 2013. **14**(2): p. 83-7.

280. Whittle, C.A., A. Kulkarni, and C.G. Extavour, *Absence of a Faster-X Effect in Beetles (Tribolium, Coleoptera).* G3 (Bethesda), 2020. **10**(3): p. 1125-1136.

281. Thurmond, J., et al., *FlyBase 2.0: the next generation.* Nucleic Acids Res, 2019. **47**(D1): p. D759-D765.

282. Kumar, S., et al., *MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.* Mol Biol Evol, 2018. **35**(6): p. 1547-1549.

283. Zhang, Z. and J. Yu, *Evaluation of six methods for estimating synonymous and nonsynonymous substitution rates.* Genomics Proteomics Bioinformatics, 2006. **4**(3): p. 173-81.

284. Krasovec, M., B. Nevado, and D.A. Filatov, *A Comparison of Selective Pressures in Plant X-Linked and Autosomal Genes.* Genes (Basel), 2018. **9**(5).

285. Krasovec, M., et al., *The Mutation Rate and the Age of the Sex Chromosomes in Silene latifolia.* Curr Biol, 2018. **28**(11): p. 1832-1838 e4.

286. Kirkpatrick, M. and D.W. Hall, *Male-biased mutation, sex linkage, and the rate of adaptive evolution.* Evolution, 2004. **58**(2): p. 437-40.

287. Pamilo, P. and M. Nei, *Relationships between gene trees and species trees.* Mol Biol Evol, 1988. **5**(5): p. 568-83.

288. Edwards, S.V. and P. Beerli, *Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies.* Evolution, 2000. **54**(6): p. 1839-54.

289. Hey, J. and R. Nielsen, *Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis.* Genetics, 2004. **167**(2): p. 747-60.

290. Singh, B. and J.D. Wells, *Chrysomyinae (Diptera: Calliphoridae) is monophyletic: A molecular systematic analysis.* Systematic Entomology 2011. **36**(3): p. 415-420.

291. Baumgartner, D.L., *Review of Chrysomya rufifacies (Diptera: Calliphoridae).* J Med Entomol, 1993. **30**(2): p. 338-52.

292. Charlesworth, B., *The evolution of chromosomal sex determination and dosage compensation.* Curr Biol, 1996. **6**(2): p. 149-62.

293. Charlesworth, B., *The evolutionary biology of sex.* Curr Biol, 2006. **16**(17): p. R693-5.

294. Gammerdinger, W.J., et al., *Comparative analysis of a sex chromosome from the blackchin tilapia, Sarotherodon melanotheron.* BMC Genomics, 2016. **17**(1): p. 808.

295. Prowse, T.A., et al., *A Y-chromosome shredding gene drive for controlling pest vertebrate populations.* Elife, 2019. **8**.

APPENDIX A. SUPPLEMENTAL TABLES

Table A1. BUSCO Completeness report of the male, thelygenic female and arrhenogenic females of *C. rufifacies*. Complete and single-copy means the gene ortholog has been found once as expected and is within 2 standard-deviations in length of the BUSCO meeting the score criteria for the alignment. Complete and duplicated means there's more than one copy found that meets length and score criteria. Fragmented means part of the gene was found but not the full length and missing means the BUSCO group was not found.

Thelygenic Female	**Eukaryota**	**Diptera**	**Arthropoda**	**Insecta**
Total BUSCO groups	303	2799	1066	1658
Complete BUSCOs (C)	278 (91.8%)	2385 (85.2%)	989 (92.8%)	1508 (91.0%)
Complete and single-copy BUSCOs (S)	276 (91.1%)	2362 (84.4%)	977 (91.7%)	1485 (89.6%)
Complete and duplicated BUSCOs (D)	2 (0.7%)	23 (0.8%)	12 (1.1%)	23 (1.4%)
Fragmented BUSCOs (F)	6 (2.0%)	130 (4.6%)	15 (1.4%)	31 (1.9%)
Missing BUSCOs (M)	19 (6.2%)	284 (10.2%)	62 (5.8%)	119 (7.1%)

Male	**Eukaryota**	**Diptera**	**Arthropoda**	**Insecta**
Total BUSCO groups	303	2799	1066	1658
Complete BUSCOs (C)	276 (91.1%)	2400 (85.8%)	994 (93.3%)	1518 (91.5%)
Complete and single-copy BUSCOs (S)	273 (90.1%)	2384 (85.2%)	986 (92.5%)	1504 (90.7%)
Complete and duplicated BUSCOs (D)	3 (1.0%)	16 (0.6%)	8 (0.8%)	14 (0.8%)
Fragmented BUSCOs (F)	7 (2.3%)	145 (5.2%)	12 (1.1%)	24 (1.4%)
Missing BUSCOs (M)	20 (6.6%)	254 (9.0%)	60 (5.6%)	116 (7.1%)

Arrhenogenic Female	**Eukaryota**	**Diptera**	**Arthropoda**	**Insecta**
Total BUSCO groups	303	2799	1066	1658
Complete BUSCOs (C)	272 (89.8%)	2381 (85.0%)	993 (93.1%)	1508 (90.9%)
Complete and single-copy BUSCOs (S)	270 (89.1%)	2366 (84.5%)	981 (92.0%)	1494 (90.1%)
Complete and duplicated BUSCOs (D)	2 (0.7%)	15 (0.5%)	12 (1.1%)	14 (0.8%)
Fragmented BUSCOs (F)	6 (2.0%)	138 (4.9%)	17 (1.6%)	40 (2.4%)
Missing BUSCOs (M)	25 (8.2%)	280 (10.1%)	56 (5.3%)	110 (6.7%)

Table A2. Orthologous clusters of predicted protein sequences from the thelygenic female, arrhenogenic female and the male assemblies. The table shows the number of clusters and the total number of protein sequences within each cluster which is (i) unique to each group, (ii) shared between each pair, and (iii) shared among the three groups. Cluster classification was done according to sequence analysis data, protein similarity comparisons, and phylogenetic relationships (AF = arrhenogenic female, TF = thelygenic female, M = male).

	Unique Clusters	***Total #protein sequences composing the cluster***
Thelygenic Female	17	35
Arrhenogenic Female	30	66
Male	20	44
AF – TF	644	1,313 (659 TF, 654 AF)
AF – M	732	1,490 (745 AF, 745 M)
TF – M	774	1,567 (783 M, 784 TF)
AF – TF – M	10,354	31,812 (10,602 AF, 10,630 TF,10,580 M)

Table A3. Orthologous contig sequences in *C. rufifacies* of the chromosomal gene contents (Muller elements) from D. melanogaster's Muller elements. Male vs female read coverage ratios was used to determine which set of orthologous Muller elements had a 2X sequence coverage (X-linked). Due to the presence of two types of female, read coverage ratios on the male was performed twice (AF-M and TF_M).

	AF		***TF***		***AF-M***		***TF-M***	
Muller elements (*D. melanogaster *chromosome)	No. of contigs with tBLASTx hits	No. of X-linked contigs (%)	No. of contigs with tBLASTx hits	No. of X-linked contigs	No. of contigs with tBLASTx hits	No. of X-linked contigs	No. of contigs with tBLASTx hits	No. of X-linked contigs
A (Chr X)	1,454	3 (0.21%)	1,454	26 (1.79%)	1,464	3 (0.20%)	1,464	16 (1.09%)
B (Chr 2L)	1,720	9 (0.62%)	1,737	28 (1.61%)	1,706	6 (0.35%)	1,706	32 (1.88%)
C (Chr 2R)	1,917	12 (0.63%)	1,913	26 (2.40%)	1,896	8 (0.42%)	1,896	29 (1.53%)
D (Chr 3L)	1,782	1 (0.10%)	1,835	45 (2.45%)	1,831	10 (0.55%)	1,831	31 (1.69%)
E (Chr 3R)	2,128	11 (0.51%)	2,131	51 (2.39%)	2,100	7 (0.33%)	2,100	32 (1.52%)
F (Chr 4)	77	1 (1.30%)	85	3 (3.53%)	84	0 (0%)	84	1 (1.19%)

Table A4. A summary of the percent of the genome composed of repetitive sequences in the thelygenic female, arrhenogenic female and the male *Chrysomya rufifacies*, their copy number and the number of bases of each. Repetitive elements were identified using homology to known Diptera repetitive elements

		Thelygenic Female			**Arrhenogenic Female**			**Male**		
		Copy number	**Bases (bp)**	**% Genome**	**Copy number**	**Bases (bp)**	**% Genome**	**Copy number**	**Bases (bp)**	**% Genome**
Class 1 Retrotransposons										
LTRs	Copia	441	104,256	0.04	463	106,151	0.04	456	103,567	0.04
	Gypsy	3,958	471,065	0.17	3,816	463,860	0.16	3,532	440,356	0.15
	Pao	1,130	176,457	0.06	1,131	187,231	0.06	1,119	186,236	0.07
LINEs	CR1	254	22,554	0.00	264	23,764	0.01	260	23,089	0.01
	Dong-R4	10	706	0.00	10	721	0.00	8	553	0.00
	I	2,743	319,008	0.11	2,875	337,734	0.11	2,752	318,547	0.11
	L1	84	4,898	0.00	11	5,971	0.00	88	4,900	0.00
	L2	309	26,244	0.01	318	27,745	0.01	301	26,002	0.01
	Penelope	4	260	0.00	7	356	0.00	3	148	0.00
	R1	1,273	190,621	0.07	1,350	202,744	0.07	1,343	198,440	0.07
	R2	8	1,847	0.00	8	1,869	0.00	8	1,924	0.00
	RTE	829	105,303	0.04	950	121,165	0.04	887	109,982	0.04
SINEs	tRNA	97	6,646	0.00	111	7,381	0.00	179	11,402	0.00
Class II DNA Transposons										
Cut and Paste Transposons	DNA	213	99,351	0.04	265	123,658	0.04	237	108,259	0.04
	CMC	911	58,029	0.02	1,020	65,497	0.02	976	61,004	0.02
	hAT	2,767	2,771	0.00	3,387	322,255	0.11	3,059	287,617	0.10
	Kolobok	5	258	0.00	5	255	0.00	8	463	0.00
	Maverick	173	48,891	0.02	180	47,465	0.02	191	53,806	0.02
	Merlin	0	0	0	1	59	0.00	0	0	0.00
	MULE	67	3,842	0.00	62	3,560	0.00	66	3,979	0.00
	P	245	13,844	0.01	226	12,963	0.00	236	13,848	0.01
	PIF	137	6,431	0.00	164	7,692	0.00	140	6,661	0.00
	PiggyBac	21	4,161	0.00	24	4,588	0.00	23	3,970	0.00
	Sola	30	1,629	0.00	35	2,078	0.00	38	2,114	0.00
	TcMar	607	92,035	0.03	662	101,190	0.03	586	92,098	0.03
	Zator	5	364	0.00	5	353	0.00	6	416	0.00
Rolling Circle	Helitron	17,100	1,790,559	0.64	21,232	2,269,371	0.77	18,757	1,955,084	0.68
rRNA		61	9,567	0.00	74	10,276	0.00	64	11,651	0.00

Table A4. Continued

Satellites	46	11,112	0.00	58	15,714	0.01	58	14,388	0.01
Simple repeats	234,230	11,974,585	4.29	253,970	12,687,148	4.3	268,134	12,844,756	4.45
Low complexity	53,832	2,766,628	0.99	55,834	2,843,978	0.96	55,983	2,821,282	0.98
Unknown	1,788	151,518	0.05	2,131	186,861	0.06	1,987	166,619	0.06
Total Repeat Content	**323,378**	**18,465,440**	**6.61**	**350,649**	**20,191,653**	**6.84**	**361,485**	**19,873,161**	**6.89**

Table A5. A comparison of enriched Gene ontology (GO) terms within biological processes (BP), cellular component (CC) and molecular function (MF), respectively shared among the putative X sequences in the four blow fly species.

Biological Processes (ChrX)	***Total***	***Elements***
CalbX CrufX LcupX PregX	11	reproduction
		developmental process
		cell communication
		cellular process
		signaling
		biological regulation
		cellular component organization or biogenesis
		response to stimulus
		protein folding
		metabolism
		localization
CalbX LcupX PregX	4	multicellular organismal process
		reproductive process
		multi-organism process
		methylation
CrufX LcupX	2	anatomical structure development
		DNA metabolism
LcupX PregX	1	biological adhesion
CrufX	5	cell cycle
		nitrogen compound transport
		macromolecular complex assembly
		response to stress
		homeostatic process
CalbX	5	ribonucleoprotein complex subunit organization
		divalent metal ion transport
		multicellular organism development
		nucleobase-containing small molecule metabolism

Table A5. Continued

		response to toxic substance
LcupX	5	regulation of catalytic activity
		ion transmembrane transport
		chromosome organization
		cellular response to DNA damage stimulus
		microtubule-based process
PregX	21	anatomical structure formation involved in morphogenesis
		organic hydroxy compound metabolism
		superoxide metabolism
		response to endogenous stimulus
		regulation of protein catabolism
		reactive oxygen species metabolism
		mitochondrial transmembrane transport
		antibiotic metabolism
		cell division
		protein refolding
		system process
		locomotion
		glycosyl compound metabolism
		autophagy
		organic hydroxy compound biosynthesis
		growth
		immune system process
		cell wall organization or biogenesis
		adhesion of symbiont to host
		cell proliferation
		prenylation
Cellular Components (ChrX)	**Total**	**Elements**
CalbX CrufX LcupX PregX	4	organelle
		membrane-enclosed lumen
		extracellular region
		membrane
CalbX LcupX PregX	12	extracellular space
		organelle envelope
		cell junction
		cell projection
		envelope
		cell
		anchoring junction
		supramolecular complex

Table A5. Continued

		supramolecular fiber
		synapse
		macromolecular complex
		cell surface
CrufX LcupX	1	intracellular
CalbX LcupX	1	postsynapse
CalbX PregX	5	apical part of cell
		thylakoid
		cell periphery
		coated membrane
		endomembrane system
CrufX	1	endoplasmic reticulum
CalbX	1	proteasome complex
LcupX	1	membrane protein complex
PregX	2	presynapse
		tethering complex
Molecular Functions (ChrX)	**Total**	**Elements**
CalbX CrufX LcupX PregX	6	binding
		transporter activity
		translation regulator activity
		molecular function regulator
		catalytic activity
		structural molecule activity
CalbX CrufX LcupX	1	enzyme regulator activity
CalbX LcupX PregX	17	structural constituent of ribosome
		transcription factor activity, sequence-specific DNA binding
		lipid binding
		metal cluster binding
		amide binding
		molecular transducer activity
		ribonucleoprotein complex binding
		antioxidant activity
		carbohydrate binding
		cofactor binding
		drug binding
		electron carrier activity
		transcription cofactor activity
		peptide binding
		binding, bridging
		sulfur compound binding
		iron-sulfur cluster binding

Table A5. Continued

CrufX LcupX	1	translation regulator activity, nucleic acid binding
CalbX LcupX	4	chitin binding
		ion transmembrane transporter activity
		coenzyme binding
		iron ion binding
CrufX	3	cytoskeletal protein binding
		transmembrane transporter activity
		nucleoside-triphosphatase activity
CalbX	4	vitamin binding
		enzyme binding
		transferase activity, transferring acyl groups other than amino-acyl groups
		signaling receptor activity
LcupX	3	transmembrane signaling receptor activity
		hydrolase activity, acting on ester bonds
		protein dimerization activity
PregX	10	ATPase regulator activity
		sulfur compound transmembrane transporter activity
		peroxiredoxin activity
		nucleoside-triphosphate diphosphatase activity
		macromolecular complex binding
		ubiquitin binding
		phosphoenolpyruvate carboxykinase activity
		DNA topoisomerase activity
		manganese ion binding
		sequence-specific double-stranded DNA binding

Table A6. *A comparison of enriched Gene ontology (GO) terms within biological processes (BP), cellular component (CC) and molecular function (MF), respectively shared among the putative Y sequences in the four blow fly species.*

Molecular Functions	*Total*	*Elements*
CalbY CrufY LcupY PregY	19	structural constituent of ribosome
		binding
		lipid binding
		molecular function regulator
		amide binding
		molecular transducer activity
		carbohydrate binding
		catalytic activity
		cofactor binding
		drug binding
		sulfur compound binding
		iron-sulfur cluster binding
		structural molecule activity
		transcription factor activity, sequence-specific DNA binding
		transporter activity
		metal cluster binding
		antioxidant activity
		electron carrier activity
CalbY CrufY LcupY	1	ribonucleoprotein complex binding
CalbY LcupY PregY	1	binding, bridging
CrufY LcupY	5	vitamin binding
		ion transmembrane transporter activity
		transmembrane signaling receptor activity
		coenzyme binding
		enzyme regulator activity
CalbY LcupY	2	transcription cofactor activity
		peptide binding
CalbY PregY	3	sulfur compound transmembrane transporter activity
		macromolecular complex binding
		DNA topoisomerase activity
CrufY	4	protein dimerization activity
		nucleoside-triphosphatase activity
		translation regulator activity, nucleic acid binding
		iron ion binding
CalbY	7	ATPase regulator activity
		peroxiredoxin activity

Table A6. Continued

		nucleoside-triphosphate diphosphatase activity
		ubiquitin binding
		phosphoenolpyruvate carboxykinase activity
		manganese ion binding
		sequence-specific double-stranded DNA binding
LcupY	3	enzyme binding
		transferase activity, transferring acyl groups other than amino-acyl groups
		magnesium ion binding
PregY	20	chromatin binding
		GTP cyclohydrolase I activity
		oxygen binding
		kinase activator activity
		sigma factor activity
		deaminase activity
		recombinase activity
		polysaccharide binding
		histone binding
		glycogen debranching enzyme activity
		telomeric DNA binding
		3-hydroxyacyl-CoA dehydrogenase activity
		chitin binding
		3-phosphoshikimate 1-carboxyvinyltransferase activity
		neurotransmitter binding
		CTP synthase activity
		3-dehydroquinate synthase activity
		peptidoglycan muralytic activity
		potassium ion binding
		modified amino acid binding
Cellular Components	**Total**	**Elements**
CalbY CrufY LcupY PregY	14	envelope
		membrane-enclosed lumen
		membrane
		cell periphery
		extracellular space
		cell projection
		cell
		endomembrane system
		supramolecular complex
		organelle
		supramolecular fiber

Table A6. Continued

		extracellular region
		macromolecular complex
		cell surface
CalbY CrufY LcupY	5	cell junction
		anchoring junction
		thylakoid
		organelle envelope
		synapse
CalbY LcupY PregY	1	coated membrane
CrufY PregY	3	nucleoid
		periplasmic space
		cell division site
CalbY LcupY	1	presynapse
CalbY PregY	1	tethering complex
CrufY	3	membrane protein complex
		intermediate filament cytoskeleton
		cell septum
CalbY	1	apical part of cell
LcupY	1	proteasome complex
PregY	7	site of polarized growth
		spindle pole body
		cell tip
		cellular bud neck
		TOR complex
		cellular bud
		cytoplasmic vesicle

Table A7. Repetitive element landscape of the isolated putative sex chromosomal sequences in *P. regina, L. cuprina, C. rufifacies* and *C. albiceps*

	Percent sequence of the characterized putative sex chromosome composed of repetitive elements							
Repeat elements	**CrufX**	**CrufY**	**CalbX**	**CalbY**	**PregX**	**PregY**	**LcupX**	**LcupY**
Retroelements	0.63	0.29	0.39	0.34	1.38	0.92	0.51	1.31
DNA transposons	0.28	0.12	0.3	0.26	0.94	0.54	0.11	0.23
Unclassified	0.96	0.3	0.85	0.99	0.55	0.27	0.33	0.56
Total Interspersed repeats	1.87	0.71	1.54	1.58	2.87	1.73	0.95	2.1
Small RNA	0	0	0	0.01	0.12	0.03	0.01	0
Satellites	0	0.01	0.01	0	0.01	0.01	0.01	0.07
Simple repeats	3.37	1.58	3.64	4.41	3.29	2.38	6.54	8.55
Low complexity	0.93	0.46	1	0.96	1.02	0.47	0.89	0.74
%masked	7.11	3.01	6.73	7.59	9.16	5.88	8.46	12.82

Table A8. The table below shows the number of contig sequences containing orthologous genes of *Drosophila melanogaster's* Muller elements. Muller element F, which is the ancestral sex chromosome in Dipteran species, is X-linked in *Phormia regina* and *Lucilia cuprina* – a characteristic expected in differentiated sex chromosomes. However, in *C. rufifacies* and *C. albiceps*, the blow fly species which exhibit homomorphic sex chromosomes, Muller element F is not X-linked, suggesting undifferentiated sex chromosomes. Data shown is generated from female flies (AF = arrhenogenic female, TF = thelygenic female).

	Phormia regina		***Lucilia cuprina***		***Chrysomya rufifacies AF***		***Chrysomya rufifaces TF***		***Chrysomya albiceps***	
Muller Elements (D. melanogaster chromosome)	**No. of contigs with tblastN hits**	**No. of X-linked contigs**	**No. of contigs with tblastN hits**	**No. of X-linked contigs**	**No. of contigs with tblastN hits**	**No. of X-linked contigs**	**No. of contigs with tblastN hits**	**No. of X-linked contigs**	**No. of contigs with tblastN hits**	**No. of X-linked contigs**
A (Chr X)	1,448	5 (0.35%)	1,801	117 (6.49%)	1,454	3 (0.21%)	1,454	26 (1.79%)	1,823	194 (10.64%)
B (Chr 2L)	1,687	10 (0.59%)	2,098	188 (8.96%)	1,720	9 (0.62%)	1,737	28 (1.61%)	2,150	198 (9.21%)
C (Chr 2R)	1,865	6 (0.32%)	2,284	186 (8.14%)	1,917	12 (0.63%)	1,913	26 (2.40%)	2,388	192 (8.04%)
D (Chr 3L)	1,694	9 (0.53%)	2,242	165 (7.35%)	1,782	1 (0.10%)	1,835	45 (2.45%)	2,355	208 (8.83%)
E (Chr 3R)	2,173	8 (0.37%)	2,639	184 (6.97%)	2,128	11 (0.51%)	2,131	51 (2.39%)	2,763	267 (9.66%)
F (Chr 4)	88	59 (67.05%)	104	57 (54.81%)	77	1 (1.30%)	85	2 (2.35%)	93	9 (9.67%)

Table A9. A list of the X-linked orthologous Muller element F gene sequences in all four blow fly species

Phormia regina	***Lucilia cuprina***	***Chrysomya rufifacies (TF)***	***Chrysomya rufifacies (AF)***	***Chrysomya albiceps***
Arl4 (ADP ribosylation factor-like 4)	*Arl4* (ADP ribosylation factor-like 4)	*CG11155*	*lgs* (legless)	*Arl4* (ADP ribosylation factor-like 4)
Cals (Calsyntenin-1)	*Cals* (Calsyntenin-1)	JYalpha		*Cals* (Calsyntenin-1)
CG11076	*CG33521*			*CG11076*
CG33521	*dpr7* (defective proboscis extension response 7)			*CG33521*
dpr7 (defective proboscis extension response 7)	*ND-49* (NADH dehydrogenase (ubiquinone) 49 kDa subunit)			*dpr7* (defective proboscis extension response 7)
pan (pangolin)	*pan* (pangolin)			*ND-49* (NADH dehydrogenase (ubiquinone) 49 kDa subunit)
CG11155	*lgs* (legless)			*pan* (pangolin)
lgs (legless)	*4E-T* (eIF4E-Transporter)			*Zip102B* (Zinc/iron regulated transporter-related protein 102B)
4E-T (eIF4E-Transporter)	*Actbeta* (Activin-β)			
Actbeta (Activin-β)	*Ank* (Ankyrin)			
Ank (Ankyrin)	*apolpp* (apolipophorin)			
Asator	*Asator*			
bip2	*bt* (bent)			
bt (bent)	*CG1909*			
Cadps (Calcium-dependent secretion activator)	*CG2316* (ATP binding cassette subfamily D)			
CG11360	*CG31997*			
CG1674	*CG31998*			
CG1909	*CG31999*			
CG2316 (ATP binding cassette subfamily D)	*CG32006*			
CG31997	CG32850			
CG31998	*CG33978*			
CG31999	*ci* (cubitus interruptus)			
CG32006	Crk (Crk oncogene)			
CG33978	*dati* (datilografo)			

Table A9. Continued

ci (cubitus interruptus)	*Ekar* (Eye-enriched kainate receptor)			
Crk (Crk oncogene)	*Eph* (Eph receptor tyrosine kinase)			
dati (datilografo)	*ey* (eyeless)			
eIF4G (eukaryotic translation initiation factor 4G1)	*fd102C* (forkhead domain 102C)			
Ekar (Eye-enriched kainate receptor)	*fuss* (fussel)			
Eph (Eph receptor tyrosine kinase)	*gw* (gawky)			
ey (eyeless)	*Gyf* (Gigyf)			
fd102C (forkhead domain 102C)	*Kif3C* (Kinesin family member 3C)			
Gat (GABA transporter)	*mav* (maverick)			
gw (gawky)	*MED26* (Mediator complex subunit 26)			
Gyf (Gigyf)	*onecut*			
Kif3C (Kinesin family member 3C)	*PlexA* (Plexin A)			
mav (maverick)	*PlexB* (Plexin B)			
MED26 (Mediator complex subunit 26)	*RhoGAP102A* (Rho GTPase activating protein at 102A)			
myo (myoglianin)	*Slip1* (SLo interacting protein 1)			
onecut	*Sox102F*			
pho	*Syt7* (Synaptotagmin 7)			
PIP4K (Phosphatidylinositol 5-phosphate 4-kinase)	*Thd1*			
PlexA (Plexin A)	*unc-13*			
PlexB (Plexin B)	*yellow-h*			
PMCA (plasma membrane calcium ATPase)	*zfh2* (Zn finger homeodomain 2)			
Pur-alpha (Purine-rich binding protein-α)				
RhoGAP102A (Rho GTPase activating protein at 102A)				

Table A9. Continued

Slip1 (Slo interacting protein 1)				
Sox102F				
sv (shaven)				
Thd1				
toy (twin of eyeless)				
unc-13				
yellow-h				
zfh2 (Zn finger homeodomain 2)				
Zyx				

Table A10. Repetitive element landscape of Muller elements in *P. regina, L. cuprina, C. rufifacies* and *C. albiceps*

PregM	***MullerA***	***MullerB***	***MullerC***	***MullerD***	***MullerE***	***MullerF***
Retroelements	0.9	0.98	1.06	1.09	0.91	1.41
DNA transposons	0.55	0.48	0.57	0.51	0.53	0.56
Unclassified	0.23	0.24	0.26	0.24	0.25	0.16
Total Interspersed repeats	1.68	1.7	1.89	1.83	1.69	2.13
Small RNA	0.08	0.07	0.07	0.06	0.05	0
Satellites	0	0	0.01	0	0	0
Simple repeats	3.69	3.68	3.73	3.86	3.76	3.59
Low complexity	1.15	1.09	1.05	1.17	1.13	0.91
Retroelements	*MullerA*	*MullerB*	*MullerC*	*MullerD*	*MullerE*	*MullerF*
SINES	0	0	0	0	0	0
LINES	0.62	0.67	0.68	0.75	0.59	0.89
LTR	0.28	0.31	0.38	0.33	0.32	0.52
Bases masked	6.58	6.54	6.74	6.92	6.63	6.63
Rolling circles	0	0	0	0	0	0
LcupM	**MullerA**	**MullerB**	**MullerC**	**MullerD**	**MullerE**	**MullerF**
Retroelements	0.63	0.7	0.61	0.6	0.62	1.86
DNA transposons	0.1	0.12	0.09	0.11	0.11	0.17
Unclassified	0.27	0.32	0.28	0.29	0.32	0.65
Total Interspersed repeats	1	1.14	0.98	1	1.05	2.68
Small RNA	0	0.01	0	0	0	0.01
Satellites	0.02	0.07	0.06	0.05	0.05	0.07
Simple repeats	3.2	3.68	4.02	3.45	3.24	2.21
Low complexity	0.37	0.39	0.36	0.37	0.39	0.31
Retroelements	*MullerA*	*MullerB*	*MullerC*	*MullerD*	*MullerE*	*MullerF*
SINES	0	0	0	0	0	0
LINES	0.3	0.33	0.29	0.29	0.3	0.76
LTR	0.33	0.37	0.32	0.3	0.32	1.1
Bases masked	4.6	5.29	5.42	4.87	4.72	5.29
Rolling circles	0	0	0	0	0	0
CrufM	**MullerA**	**MullerB**	**MullerC**	**MullerD**	**MullerE**	**MullerF**
Retroelements	0.4	0.46	0.44	0.49	0.4	0.8
DNA transposons	0.12	0.18	0.18	0.16	0.15	0.17
Unclassified	0.36	0.46	0.44	0.43	0.38	0.62
Total Interspersed repeats	0.88	1.1	1.06	1.08	0.93	1.59
Small RNA	0.01	0.01	0	0	0	0

Table A10. Continued

Satellites	0	0.01	0	0	0	0
Simple repeats	3.98	3.92	4.1	4.26	3.99	3.1
Low complexity	1.04	0.98	0.98	1.01	1.03	0.84
Retroelements	*MullerA*	*MullerB*	*MullerC*	*MullerD*	*MullerE*	*MullerF*
SINES	0	0	0	0	0	0
LINES	0.27	0.27	0.29	0.32	0.28	0.45
LTR	0.14	0.18	0.15	0.17	0.12	0.35
Bases masked	5.9	6.01	6.14	6.34	5.95	5.54
Rolling circles	0	0	0	0	0	0
CalbM	**MullerA**	**MullerB**	**MullerC**	**MullerD**	**MullerE**	**MullerF**
Retroelements	0.46	0.4	0.4	0.47	0.38	0.74
DNA transposons	0.14	0.19	0.18	0.18	0.15	0.13
Unclassified	0.28	0.3	0.34	0.27	0.26	0.46
Total Interspersed repeats	0.87	0.89	0.92	0.92	0.79	1.32
Small RNA	0.01	0.01	0.01	0	0	0
Satellites	0	0	0	0	0	0
Simple repeats	3.84	3.68	3.99	4.02	3.85	3.39
Low complexity	1.01	0.93	1	0.97	1.01	0.98
Retroelements	*MullerA*	*MullerB*	*MullerC*	*MullerD*	*MullerE*	*MullerF*
SINES	0	0	0	0	0	0
LINES	0.31	0.23	0.26	0.31	0.27	0.42
LTR	0.15	0.17	0.15	0.15	0.11	0.32
Bases masked	5.72	5.5	5.91	5.91	5.65	5.7
Rolling circles	0	0	0	0	0	0

APPENDIX B. SUPPLEMENTAL FIGURES

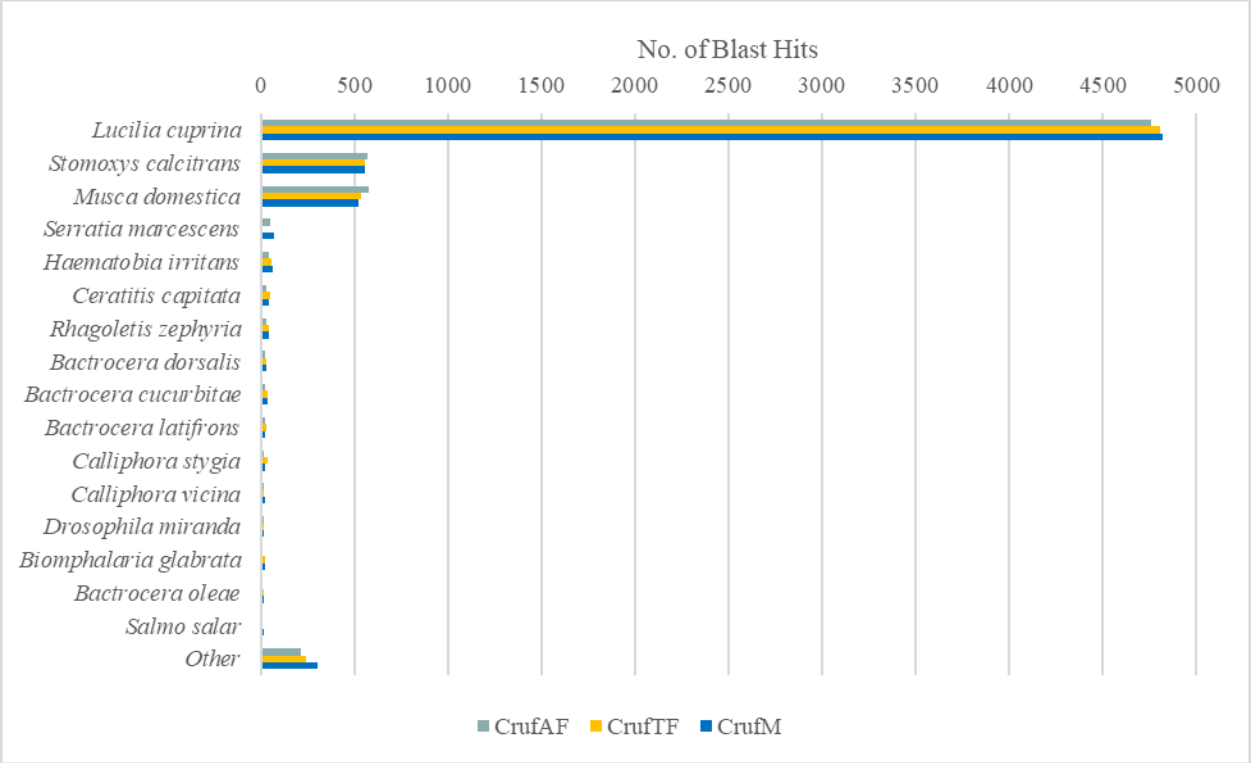


Figure B1. Top-hit species distribution from Blast2GO for arrhenogenic female, thelygenic female and the male. The top hit species is the blow fly *L. cuprina*. The species listed are those with >10 hits. Those with less than the threshold are summed and grouped in the 'Other' category