The House Fly Y Chromosome is Young and Undifferentiated from its Ancient X Chromosome Partner

Richard P. Meisel[1]*, Christopher A. Gonzales[1], and Hoang Luu[1]

1. Department of Biology and Biochemistry, University of Houston

*To whom correspondence should be addressed:
Richard Meisel, 3455 Cullen Blvd #342, Houston, TX 77204-5001, rpmeisel@uh.edu

Running title: House fly sex chromosomes

Keywords: sex chromosomes; neo-Y chromosome; *Musca domestica*

September 1, 2016

Abstract

Ancient or canonical sex chromosome pairs consist of a gene rich X (or Z) chromosome and a male- (or female-) limited Y (or W) chromosome that is gene poor. In contrast to highly differentiated sex chromosomes, nascent sex chromosome pairs are homomorphic or very similar in sequence content. Nascent sex chromosomes arise frequently over the course of evolution, as evidenced by differences in sex chromosomes between closely related species and sex chromosome polymorphisms within species. Sex chromosome turnover typically occurs when an existing sex chromosome becomes fused to an autosome or an autosome acquires a new sex-determining locus/allele. Previously documented sex chromosome transitions involve changes to both members of the sex chromosome pair (X and Y, or Z and W). The house fly has sex chromosomes that resembles the ancestral fly karyotype that originated 100 million years ago, and therefore house fly is expected to have differentiated X and Y chromosomes. We tested this hypothesis using whole genome sequencing and transcriptomic data, and we surprisingly discovered little evidence for X-Y differentiation in house fly. We propose that house fly has retained the ancient X chromosome, but the ancestral Y was replaced by an X chromosome carrying a male determining gene. In this evolutionary scenario, the house fly has an ancient X chromosome that is partnered with with a neo-Y chromosome. This example of sex chromosome recycling illustrates how one member of a sex chromosome pair can experience evolutionary turnover while the other member remains unaffected.

1. Introduction

In organisms where sex is determined by genetic factors, sex determining loci reside on sex chromosomes. Sex chromosome systems can be divided into two broad categories: 1) males are the heterogametic sex (XY); or 2) females are the heterogametic sex (ZW). In long established sex chromosomes—such as in birds, eutherian mammals, and *Drosophila*—the X and Y (or Z and W) chromosomes are typically highly differentiated (Charlesworth, 1996; Charlesworth et al., 2005). The X (or Z) chromosome usually resembles an autosome in size and gene density, although there are some differences in gene content between the X and autosomes (Ellegren, 2011; Meisel et al., 2012). In contrast, Y (or Z) chromosomes tend contain a small number of genes with male- (or female-) specific functions and are often enriched with repetitive DNA as a result of male- (or female-) specific selection pressures, a low recombination rate, and a reduced effective population size (Rice, 1996; Bachtrog, 2013). This X-Y (or Z-W) differentiation results in a heterogametic sex that is effectively haploid for most or all X (or Z) chromosome genes.

Highly divergent X-Y (or Z-W) pairs trace their ancestry to a pair of undifferentiated autosomes (Bull, 1983; Charlesworth, 1991). Many species harbor undifferentiated sex chromosomes because they are either of recent origin or non-canonical evolutionary trajectories have prevented X-Y (or Z-W) divergence (Stöck et al., 2011; Bachtrog, 2013; Vicoso et al., 2013; Yazdi and Ellegren, 2014). Recently derived sex chromosomes typically result from Robertsonian fusions between an existing sex chromosome and an autosome, or they can arise through a mutation that creates a new sex determining locus on an autosome (Bachtrog et al., 2014; Beukeboom and Perrin, 2014). In both cases, one of the formerly autosomal homologs evolves into an X (or Z) chromosome, and the other homolog evolves into a Y (or W) chromosome. In some cases, one or both of the ancestral sex chromosomes can revert back to an autosome when a new chromosome becomes sex-linked (Carvalho and Clark, 2005; Larracuente et al., 2010; Vicoso and Bachtrog, 2013). In all of the scenarios described above, the X and Y (or Z and W) chromosomes evolve in concert, with an evolutionary transition in one sex chromosome producing a corresponding change in its partner.

Sex chromosome evolution has been extensively studied in higher dipteran flies (Brachyc-

era), where sex chromosome transitions involving X-autosome fusions are common (Patterson and Stone, 1952; Schaeffer et al., 2008; Baker and Wilkinson, 2010; Vicoso and Bachtrog, 2015). The ancestral brachyceran karyotype consists of five large autosomal pairs (known as Muller elements A–E) and a small sex chromosome pair (element F is the X chromosome), and this genomic arrangement has been conserved for ∼100 million years in some lineages (Muller, 1940; Foster et al., 1981; Weller and Foster, 1993; Vicoso and Bachtrog, 2013; Sved et al., 2016). In species with the ancestral karyotype, females are XX and males are XY, with a male-determining locus (M factor) on the Y chromosome (Bopp et al., 2014; Hamm et al., 2015). Many sex chromosome transitions have occurred across Brachycera, including complete reversions from an X to an autosome and fusions of ancestral autosomes with the X chromosome (Schaeffer et al., 2008; Baker and Wilkinson, 2010; Vicoso and Bachtrog, 2013, 2015).

The house fly (*Musca domestica*) is a classic model system for studying sex determination because it harbors a vast array of natural and laboratory genetic variation (Dübendorfer et al., 2002). For example, the M factor in house flies has been mapped to the Y chromosome, each of the five autosomes, and even the X chromosome (Hamm et al., 2015). Cytological evidence suggests the house fly X and Y chromosomes are the ancient sex chromosome pair shared by the common ancestor of Brachycera (Boyes et al., 1964; Hamm et al., 2015). If the ancestral karyotype segregates in house fly populations, we expect that the Y chromosome is differentiated from its gametologous X chromosome (Vicoso and Bachtrog, 2013; Linger et al., 2015; Vicoso and Bachtrog, 2015). We tested this hypothesis using whole genome and transcriptome sequencing of house flies to examine sequence divergence between the X and Y chromosomes. Unexpectedly, we observed minimal differentiation in sequence and gene content between X and Y chromosomes in genomes that were previously thought to carry the ancestral karyotype. We propose that the ancestral Brachyceran Y chromosome has been lost from house fly populations, and that all existing Y chromosomes in natural populations arose from the recent translocation of the M factor onto an ancestral X chromosome. This represents, to the best of our knowledge, the first example of the "recycling" of a sex chromosome pair through the creation of a nascent Y from an ancient X chromosome (Graves, 2005).

2. Results

2.1. The house fly X and Y chromosomes do not have unique sequences

Our first goal was to identify house fly X chromosome sequences not found on the Y, which would be consistent with the hypothesis that house flies have an ancient, differentiated sex chromosome pair. Males of the house fly genomic reference strain (aabys) have been previously characterized as possessing the XY karyotype (Wagoner, 1967; Tomita and Wada, 1989; Scott et al., 2014). To identify X-linked genes and examine differentiation between X and Y chromosomes, we used the Illumina technology to sequence genomic DNA (gDNA) separately from male (XY) and female (XX) aabys flies (3 replicates of each sex), and we aligned the reads to the annotated genome. If house fly males have a Y chromosome that is fully differentiated from the X, we expect females to have twice the sequencing coverage ($\log_2 \frac{M}{F} = -1$) within genes on Muller element F (the ancestral X chromosome) as males (Vicoso and Bachtrog, 2013). We instead surprisingly find that the average sequencing coverage in males and females is almost identical ($\log_2 \frac{M}{F} = 0$) for genes on all six chromosomes (**Fig 1**).

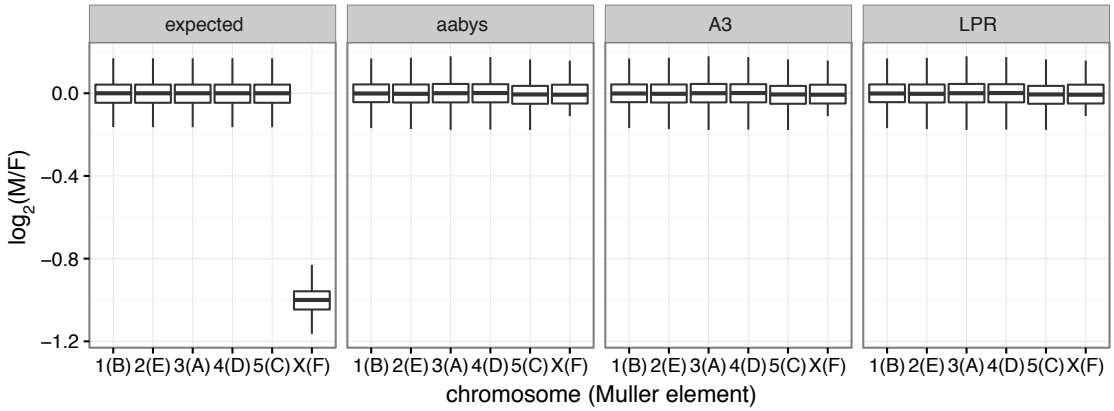

Figure 1: Expected sequencing coverage in males relative to females ($\log_2 \frac{M}{F}$) in an XY system with a degenerated Y chromosomes (left), and observed coverage in three house fly strains (aabys, A3, and LPR) for each house fly chromosome (Muller elements in parentheses).

To determine whether lack of X-Y differentiation is common to other XY strains of the house fly, we sought to identify X-linked genes in two additional strains previously reported to have XY males: A3 and LPR (Scott and Georghiou, 1985; Scott et al., 1996; Liu and

Yue, 2001). We sequenced gDNA from males and females of the A3 and LPR strains, and we aligned those reads to the reference genome. Consistent with the results from aabys, both the A3 and LPR strains had identical sequencing coverage within genes across all six chromosomes in males and females (**Fig 1**). Our results suggest that there are no genes found on the house fly X chromosome that are not present on the Y chromosome.

To ensure that our results are not an artifact of incomplete annotation of house fly X-linked genes, we calculated the male:female fold-difference in sequence mapping coverage ($\log_2 \frac{M}{F}$) across non-overlapping 1 kb intervals in the reference genome. The distribution of $\log_2 \frac{M}{F}$ across autosomes is expected to be centered at zero. If males have a single copy of the X chromosome, we should observe a second peak at $\log_2 \frac{M}{F} = -1$, indicating a 2-fold enrichment of X-linked sequences in females. We do indeed observe that the distribution of $\log_2 \frac{M}{F}$ is centered near zero for all three house fly strains in our analysis (**Fig 2**). However, we do not observe a second peak at $\log_2 \frac{M}{F} = -1$ in any of the distributions (**Fig 2**). This result provides further evidence that the house fly X chromosome does not contain sequences absent from the Y chromosome.

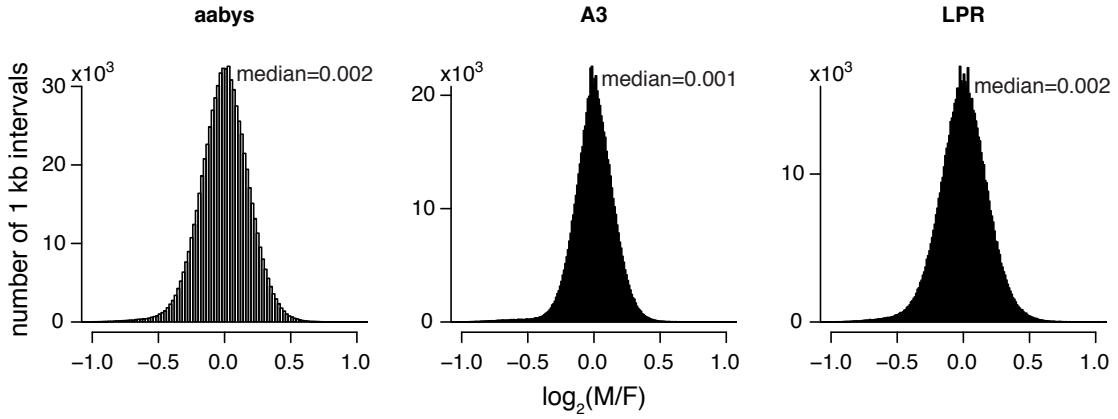

Figure 2: Histograms of $\log_2 \frac{M}{F}$ for 1 kb intervals across three strains.

We next sought to identify Y-linked sequences that are absent from the X chromosome (i.e., the reciprocal of the analyses described above). To this end, we first used the male sequencing reads from the aabys strain to assemble a genome that contains a Y chromosome. It was necessary to assemble a male genome because the genome project sequenced gDNA from female flies (Scott et al., 2014). Then we used a k-mer comparison approach to identify

male-specific sequences by searching for male genomic scaffolds that are not matched by female sequencing reads (Carvalho and Clark, 2013). Most of the scaffolds in the male genome assembly were (nearly) completely matched by female sequencing reads, and none of the male scaffolds were completely unmatched by female sequencing reads (**Fig 3**). In contrast, when this approach was used to identify Y-linked scaffolds in species with differentiated sex chromosomes (*Drosophila* and humans), a substantial number of Y-linked scaffolds were completely unmatched by female sequencing reads (Carvalho and Clark, 2013). Our results therefore suggest that there are very few, if any, Y-specific sequences in the house fly genome, other than the M factor which we failed to detect. We therefore hypothesize that the house fly "Y chromosome" is actually an X chromosome that carries an M-factor (X^M), and house fly males previously characterized as XY are better described as being XX^M.

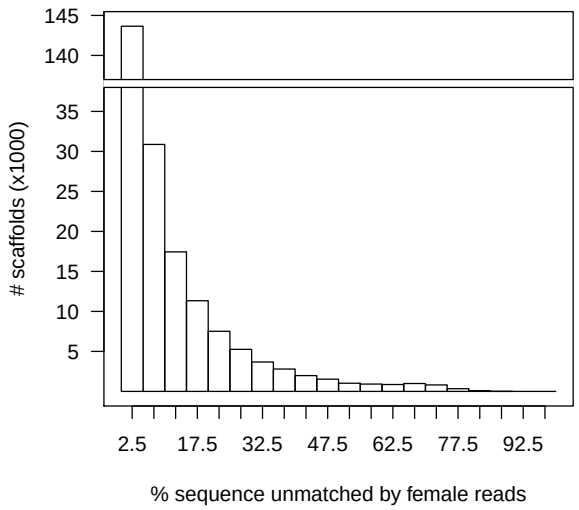

Figure 3: Histogram of female read mapping coverage to scaffolds assembled from a male genome.

2.2. Moderate differences in sequence abundance between house fly males and females

We next examined whether housefly X and Y chromosomes might exhibit differential representation of shared sequences, as might be expected from expansion or contraction of satellite repeats or other repetitive elements. We first used a principal components (PC) analysis to compare read mapping coverage of the male and female sequencing libraries across non-

overlapping 1 kb intervals in the reference (female) genome. The first PC (PC1) explains 81.5–91.1% of the variance in coverage across libraries in the three strains, and PC1 clearly separates the male and female sequencing libraries in all three strains (**Fig 4**). Therefore, house fly males and females, and by association X and Y chromosomes, exhibit systematic differences in the abundance of some sequences, even if neither sex chromosome contains unique sequences.

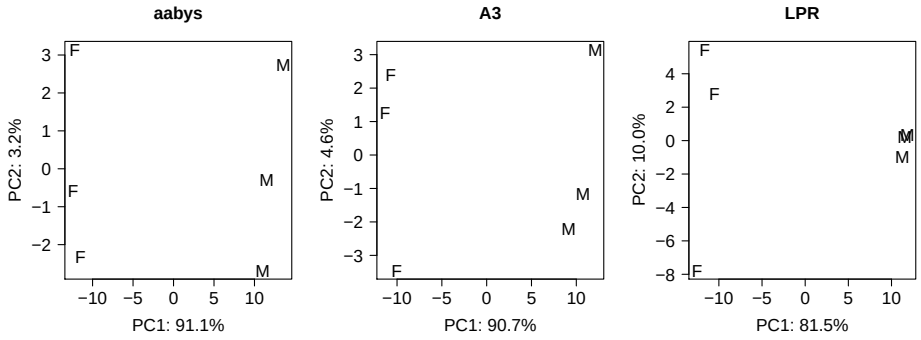

Figure 4: Plot of the first two principal components explaining differential sequencing coverage between female (F) and male (M) libraries.

We applied two different approaches to characterize sequences enriched on the X and Y chromosomes (i.e., differentially abundant in female and male genomes). First, we searched for 1 kb windows with significantly different coverage between males and females (false discovery rate corrected $P < 0.05$ and $|\log_2 \frac{M}{F}| > 1$). We identified 214 "sex-biased" windows: 63 are >2-fold enriched in females, and 151 are >2-fold enriched in males (Supplementary Data). The X and Y chromosomes of house fly are largely heterochromatic (Boyes et al., 1964; Hediger et al., 1998b), and it is possible that differences in the abundances of particular repetitive DNA sequences (e.g., transposable elements and other interspersed repeats) between the X and Y chromosomes are responsible for the differences in read coverage between females and males. Sequences from repetitive heterochromatic regions of the genome are less likely to be mapped to a genomic location (Smith et al., 2007), and we therefore expect sex-biased windows to be located on scaffolds that are not mapped to a house fly chromosome. Only 2/63 (3.2%) female-enriched windows are within a scaffold that we were able to map to a chromosome (neither was mapped to element F, the ancestral X chromosome). In addition, 59/151 (39.1%) male-enriched windows are within a scaffold that maps

to a Muller element (only one of those scaffolds maps to element F). In contrast, 65.7% of 1 kb windows that are not differentially covered between males and females are on scaffolds that we are able to map to Muller elements (2033/3096 windows with $P > 0.05$ and $|\log_2 \frac{M}{F}| < 0.01$). These unbiased windows are more likely to be mapped to a Muller element than the sex-biased windows ($P < 10^{-15}$ in Fisher's exact test), providing some evidence that differential coverage between males and females is driven by repeat content differences between the X and Y chromosomes.

We next tested for an enrichment of annotated repeats within the female- and male-biased 1 kb windows, and we found that all 63 of the female-biased windows and most of the male-biased windows (149/151) contain sequences masked as repetitive during the house fly genome annotation (Supplementary Data). However, 3071/3096 (>99%) of the 1 kb windows that are not differentially covered between males and females also contain repeat masked sequences; this fraction is not significantly different than the fraction of repeat masked sex-biased windows ($P = 1$ for female-biased and $P = 0.6$ for male-biased windows using Fisher's exact test). In addition, the proportion of sites within male-biased and female-biased windows that are repeat masked is less than that of unbiased windows, suggesting that the sex-biased windows are actually depauperate for annotated repeats (**Fig S1**). However, these analyses are limited because a large fraction (≥52%) of the house fly genome is composed of interspersed repeats that are poorly annotated (Scott et al., 2014). Future improvements to repeat annotation in the housefly genome may therefore shed light on the nature of repetitive sequences that differentiate the X and Y chromosomes.

As a second approach to identify candidate X- or Y-enriched sequences, we first determined the abundances of all possible 2–10mers in the male and female aabys sequencing reads. This approach will identify smaller sequence motifs that may differentiate the X and Y chromosomes than the analysis described above, and it does not require any *a priori* repeat annotations. The 100 most common k-mers are found at similar frequencies in both males and females (**Fig 5**), with the abundances highly correlated between sexes ($r = 0.999$). We considered a k-mer to be over-represented in one sex if the minimum abundance across the three replicate libraries for that sex is greater than the maximum in the other sex. Six k-mers are over-represented in males using this cutoff, but they are all less than 2-fold en-

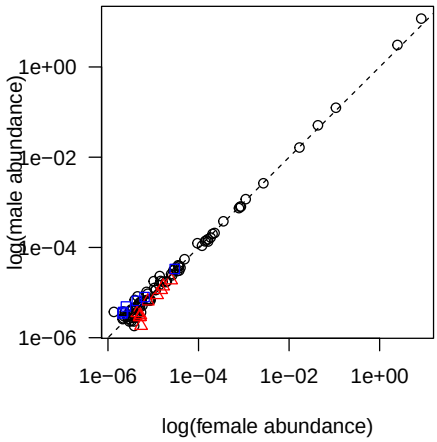

Figure 5: The abundances of the 100 most common k-mers in the male and female sequencing reads from the aabys strains are graphed. The mean across all three libraries for males and females is plotted for each k-mer. Triangles indicate k-mers where abundances in all three female libraries are greater than the three male libraries, and squares indicate k-mers that are more abundant in male libraries. The dashed line indicates equal representation in males and females.

riched in males (**Figs 5 & S2**). These results suggest that short sequence repeats do not predominantly differentiate the X and Y chromosomes.

2.3. Relative heterozygosity in males and female suggests that the house fly Y chromosome is very young

Our data suggest that, other than the unidentified M factor, the house fly Y chromosome is not highly differentiated from the X. We therefore hypothesize that the house fly Y chromosome is the result of a recent transition of an ancestral X chromosome into a neo-Y through the acquisition of an M factor. While recently derived neo-Y chromosomes may not differ in gene content from the gametologous X chromosome, modest sequence-level X-Y differentiation can result in elevated heterozygosity within sex-linked genes in males (Vicoso and Bachtrog, 2015). We tested for elevated sex-linked heterozygosity by first identifying polymorphic sites (SNPs) within genes in aabys males and females. We then calculated the proportion of heterozygous SNPs in males relative to females for genes on each chromosome (**Fig 6**A). Genes on the ancestral X chromosome (element F) have equivalent heterozy-

gosity in males and females ($P = 0.45$ in a Mann-Whitney test comparing male:female heterozygosity on element F with the other chromosomes), demonstrating that the house fly Y chromosome is so young that it has not yet accumulated modest sequence differences from the X chromosome.

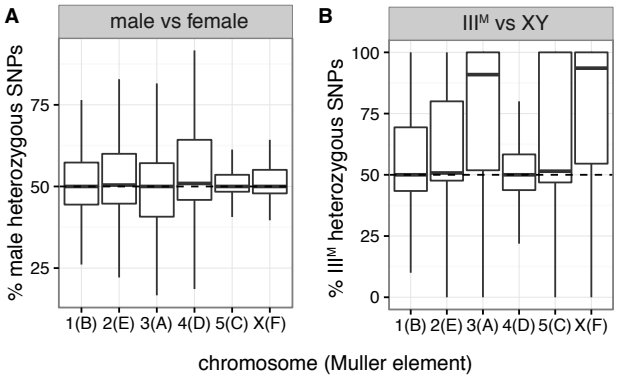

Figure 6: Elevated heterozygosity on the third chromosome in III^{M} males, but not on the X chromosome in XY males. Box plots show the distribution of the percent of heterozygous SNPs within genes on each chromosome in either (**A**) XY males relative to XX females or (**B**) III^{M} males relative to XY males. Values >50% indicate elevated heterozygosity in XY males or III^{M} males. The median across all autosomes is indicated by a dashed line.

Some house fly males carry the M factor on the third chromosome (III^{M}) and two copies of the X chromosome, neither of which has an M factor (Hamm et al., 2015). The III^{M} chromosome is therefore a recently derived neo-Y chromosome, and we expect that males heterozygous for III^{M} (hereafter III^{M} males) will have an excess of heterozygous SNPs on the third chromosome. To test this hypothesis, we used available RNA-Seq data (Meisel et al., 2015) to calculate the proportion of heterozygous SNPs in III^{M} males relative to males previously classified as XY (**Fig 6**B). As predicted, there is an excess of heterozygous SNPs on the third chromosome in III^{M} males relative to XY males ($P = 10^{-122}$ in a Mann-Whitney test comparing chromosome III with the other autosomes). Surprisingly, there is also elevated heterozygosity on the X chromosome in III^{M} males relative to XY males ($P = 10^{-4}$) even though III^{M} males have the XX genotype. These results further support our conclusion that the house fly Y chromosome is not differentiated from the X chromosome. In contrast, the III^{M} chromosome harbors evidence that it is partially differentiated from the non-M-bearing third chromosome, suggesting that the III^{M} chromosome has been a neo-Y

chromosome for more time than the "canonical" house fly Y.

3. Discussion

Cytological examination suggests that the house fly has the ancestral karyotype of higher dipterans, which includes a Y chromosome that is differentiated from the X chromosome (Boyes et al., 1964; Boyes and Van Brink, 1965; Vicoso and Bachtrog, 2013). However, we find almost no evidence for X-Y differentiation in the house fly genome: we do not find any sequences unique to the X or Y (**Figs 1, 2, & 3**); there is very little evidence for differential abundance of specific sequences on the X and Y (**Fig 5**, but see **Fig 4**); and there is not elevated heterozygosity within X chromosome genes in males (**Fig 6**). Curiously, *in situ* hybridizations of chromosomal dissections to mitotic chromosomes have detected Y-specific sequences in the house fly genome (Hediger et al., 1998b), but the sequences of these chromosomal segments are unknown. In contrast, we fail to detect Y-specific or highly Y-enriched sequences in the house fly genome (**Figs 3 & 5**), which suggests that the male-specific region of the Y chromosome (including the M factor) is small relative to the rest of the chromosome and/or difficult to assemble using short sequencing reads. We therefore hypothesize that the house fly Y chromosome is actually an ancestral brachyceran X chromosome that very recently acquired an M factor. In our model, after the X chromosome acquired an M factor, the ancestral Y chromosome was lost from house fly populations (**Fig 7**). Our results suggest that the X-to-Y conversion happened after the creation of the III^M chromosome because, unlike XY males, III^M males have elevated heterozygosity on their neo-sex chromosome (**Figs 6 & 7**).

There are four additional lines of evidence to support our hypothesis that the house fly Y chromosome is recently derived from the ancestral X chromosome. First, house fly X and Y chromosomes are largely monomorphic in cytological examinations and can only be distinguished through careful examination of their mitotic morphology (Denholm et al., 1983; Cakir and Kence, 1996). Our results suggest that the morphological differences between the X and Y chromosomes could result from the differential abundance of particular sequences between X and Y (**Fig 4**) rather than extensive sequence differentiation that characterizes ancient pairs of sex chromosomes. In addition, the X chromosome carrying an M factor

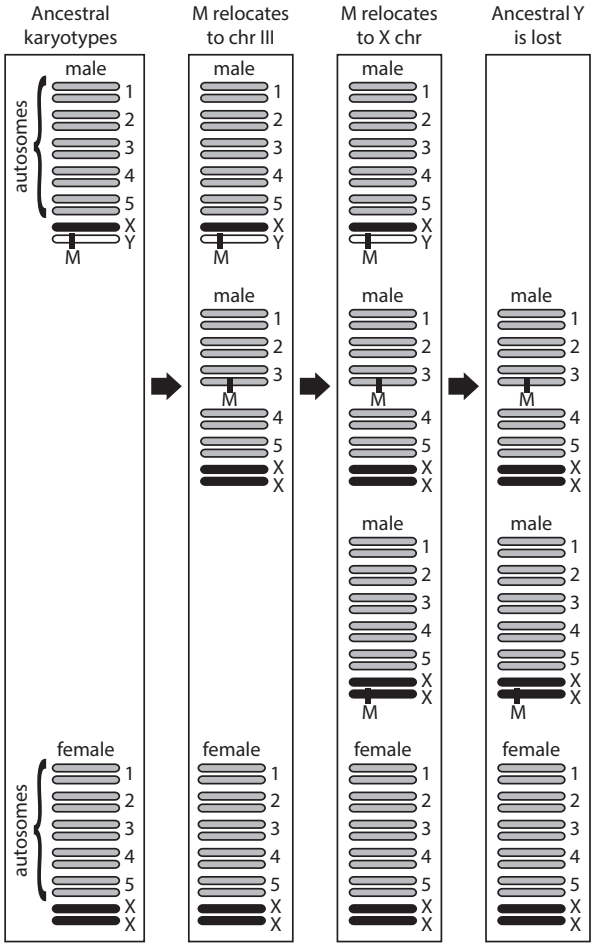

Figure 7: Model for the evolution of the house fly sex chromosomes via relocation of the M factor.

(X^M) was thought to be different from the Y chromosome (Hamm et al., 2015), but our results suggest that the X^M and Y chromosomes are one and the same. Second, no sex-linked genetic markers have been identified on the ancestral house fly sex chromosomes other than M (Hamm et al., 2015), suggesting that there are no X-specific genes or genetic variants. Third, III^M males that are classified as XX are fertile (Bull, 1983; Hamm et al., 2015), demonstrating that no essential male fertility genes are unique to the Y chromosome apart from the M factor. Fourth, house flies that carry only a single copy of either the X or Y chromosome (i.e., XO or YO flies) are viable and fertile (Bull, 1983; Hediger et al., 1998a), indicating that no essential genes are uniquely found on the X and missing from the Y chromosome and vice versa.

Our results provide the first evidence, to our knowledge, of the conversion of an existing X

chromosome into a Y chromosome (or Z into W), recycling a differentiated sex chromosome pair into nascent sex chromosomes without any evidence of fusion to an autosome. In comparison, most previously documented sex chromosome transitions involved autosomes transforming into sex chromosomes through either the evolution of a novel sex determining locus on the autosome or a fusion of the autosome with a sex chromosome (e.g., Patterson and Stone, 1952; Steinemann and Steinemann, 1998; Filatov et al., 2000; Liu et al., 2004; Veyrunes et al., 2004; Carvalho and Clark, 2005; Vallender and Lahn, 2006; Ross et al., 2009; Vicoso and Bachtrog, 2013; Bachtrog et al., 2014; Beukeboom and Perrin, 2014; Vicoso and Bachtrog, 2015). There are other examples of sex chromosome transformations involving only X, Y, Z, and W chromosomes (i.e., no autosomes) in platyfish, *Rana rugosa*, and *Xenopus tropicalis* (Kallman, 1984; Miura, 2007; Roco et al., 2015). These X/Y/Z/W transformations in fish and frogs involve nascent sex chromosomes, not ancient sex chromosomes as in house fly. Moreover, the sex chromosome transitions in platyfish, *R. rugosa*, and *X. tropicalis* all involve a change in the heterogametic sex, whereas the house fly X and Y chromosomes did not switch to a Z and W.

The X-to-Y conversion in house fly was possible because the sex-determining locus relocated from the Y to the X (Hamm et al., 2015). Relocating sex determining loci are rare and do not typically include long-established sex chromosomes (Traut and Willhoeft, 1990; Woram et al., 2003; Faber-Hammond et al., 2015), suggesting that X-to-Y (or Z-to-W) conversion similar to house fly may not be observed in other taxa. However, there is rampant gene traffic to and from long-established Y chromosomes (Koerich et al., 2008; Hughes et al., 2015), providing a possible mechanism for the Y-to-X (or W-to-Z) relocation of a sex determining locus in other taxa even if the sex determiner does not exhibit a high rate of translocation on its own. The fact that the neo-Y chromosome in house fly remained undetected despite decades of work on this system (Dübendorfer et al., 2002) suggests that X-to-Y transitions may have occurred in other taxa and remain cryptic because the karyotype has remained unchanged.

4. Methods

4.1. Fly strains

We attempted to identify X- and Y-linked sequences in five house fly strains. One strain, Cornell susceptible (CS), has been reported to have X/X;III^{M}/III males (Scott et al., 1996; Hamm et al., 2005; Meisel et al., 2015). The other four strains have previously been characterized as having males with the XY karyotype: aabys, A3, LPR, and CSaY. The genome strain, aabys, has recessive phenotypic markers on each of the five autosomes (chromosomes I–V) and had been cytologically determined to have XY males (Wagoner, 1967; Tomita and Wada, 1989; Scott et al., 2014). The A3 strain was generated by crossing XY males from a pyrethroid-resistant strain (ALHF) with aabys females (Liu and Yue, 2001). The LPR strain is a pyrethroid resistant strain that was previously determined to have XY males (Scott and Georghiou, 1985; Scott et al., 1996). Finally, the CSaY strain was created by crossing aabys males (XY) with CS females, and then backcrossing the male progeny to CS females to create a strain with the aabys Y chromosome on the CS background (Meisel et al., 2015). We validated that the M-factor is not on an autosome in the A3, LPR, and CSaY strains by crossing males of each strain to aabys females, and then we backcrossed the male progeny to aabys females. We did not observe sex-linked inheritance of any of the aabys phenotypic markers, confirming that the M-factor is not on chromosomes I–V in A3, LPR, or CSaY. Females of all strains were expected to be XX.

4.2. Genome sequencing, mapping, and assembly

The house fly genome consortium sequenced, assembled, and annotated the genome using DNA from female flies of the aabys strain, a line with XX females and XY males (Scott et al., 2014). The annotation includes both predicted genes and inferred homology relationships with *D. melanogaster* genes, and we used the orthology calls from annotation release 100 (version 2.0.2) to assign house fly genomic scaffolds to chromosome arms using a majority rule as described previously (Meisel et al., 2015). We independently sequenced genomic DNA (gDNA) from aabys male and female heads with 150 bp paired-end reads on an Illumina NextSeq 500 at the University of Houston genome sequencing core. Three replicate libraries

of each sex were prepared using the Illumina TruSeq DNA PCR-free kit, and the six libraries were pooled and sequenced in a single run of the machine (**Accession TBD**). We also sequenced gDNA from three replicates of male and female heads from A3 and LPR flies (12 samples total) in a single run on the NextSeq 500 using 75 bp paired-end reads (**Accession TBD**). Illumina sequencing reads were mapped to the assembled house fly genome using BWA-MEM with the default parameters (Li and Durbin, 2009; Li, 2013), and we only included uniquely mapping reads where both ends of a sequenced fragment mapped to the same scaffold in the reference genome.

We additionally assembled the reads from aabys male samples using SOAPdenovo2 (Luo et al., 2012) to construct a reference genome that contains Y-linked sequences. Mapping our sequence data to the reference genome revealed that our average insert size was 370 bp (**Fig S3**), which was used as a parameter in the SOAPdenovo2 genome assembly. A pair number cutoff of 3 and a minimum alignment length 32 bp were also used for the assembly.

4.3. Identifying X- and Y-linked sequences

We used four differential coverage approaches to identify candidate X- and Y-linked sequences in the house fly genome. The first approach identifies X-linked genes or sequences by testing for 2-fold higher abundance in females relative to males (Vicoso and Bachtrog, 2013). To do this, we used DESeq2 to calculate the $\log_2$ relative coverage within individual genes and 1 kb windows between the three male and female derived libraries (Love et al., 2014). We also used DESeq2 to calculate P-values for differential coverage between females and males.

The second approach was used to identify Y-linked sequences by searching for scaffolds in the male genome assembly that are missing from the female sequencing reads. We only considered assembled scaffolds from the male genome that were $\geq$1 kb. We implemented a k-mer comparison approach to identify male-specific sequences (Carvalho and Clark, 2013). In our implementation, we used a k-mer size of 15 bp and the options described by Carvalho and Clark (2013) for identifying Y-linked sequences in *Drosophila* genomes.

In the third approach, we analyzed gDNA sequencing reads from aabys males and females to identify k-mers with sexually dimorphic abundances. We used the k-Seek method to count the abundance of 2–10mers in the three male and three female aabys sequencing libraries

(Wei et al., 2014). We normalized the k-mer counts by multiplying the count by the length of the k-mer and dividing by the number of reads in the library.

The fourth approach identifies nascent sex chromosomes because they have elevated heterozygosity in the heterogametic sex (Vicoso and Bachtrog, 2015). We implemented this approach using both gDNA- and mRNA-Seq data. For the gDNA-Seq, we used the Genome Analysis Toolkit (GATK), following the best practices provided by the software developers (McKenna et al., 2010). Starting with the male and female mapped reads from the aabys strain described above, we identified duplicate reads. Insertions and deletions (indels) were identified and realigned using RealignerTargetCreator and IndelRealigner, respectively. We then called variants in each of the six aabys sequencing libraries using HaplotypeCaller, and we selected the highest quality SNPs and indels using SelectVariants and VariantFiltration (for SNPs: QD < 2, MQ < 40, FS > 60, SOR > 4, MQRankSum < -12.5, ReadPosRankSum < -8; for indels: QD < 2, ReadPosRankSum < -20, FS > 200, SOR > 10). The high quality SNPs and indels were next used for recalibration of the base calls with BaseRecalibrator and PrintReads. The process of variant calling and base recalibration was performed three times, at which point there were no benefits of additional base recalibration as validated with AnalyzeCovariates. We next used the recalibrated reads from all three replicates of each sex to call variants in males and females using HaplotypeCaller with emission and calling confidence thresholds of 20. We filtered those variants using VariantFiltration with a cluster window size of 35 bp, cluster size of 3 SNPs, FS > 20, and QD < 2. We used the variant calls to identify heterozygous SNPs within genes using the coordinates from the genome sequencing project (Scott et al., 2014).

When we implemented the GATK pipeline for variant calling of the mRNA-Seq data (accession: GSE67065 Meisel et al., 2015), we used STAR to align reads from 6 XY male libraries and 6 III^{M} male libraries separately (Dobin et al., 2013). After aligning reads to the reference genome, we used the aligned reads to create a new reference genome index from the inferred spliced junctions in the first alignment, and then we performed a second alignment with the new reference. We next marked duplicate reads and used SplitNCigarReads to reassign mapping qualities to 60 with the ReassignOneMappingQuality read filter for alignments with a mapping quality of 255. Indels were realigned and three rounds of variant calling and

base recalibration were performed as described above for the gDNA-Seq data. We applied GenotypeGVCFs to the variant calls from the 2 strains for joint genotyping of all samples, and then we used the same filtering parameters as used in the gDNA-Seq to extract high quality SNPs and indels from our variant calls.

5. Data Access

All sequence data have been submitted to GenBank under accessions XXXXXX.

6. Acknowledgements

This project was initiated during discussions with Andy Clark and Rob Unckless, who provided valuable comments throughout the completion of this work. Illumina sequencing was performed by the University of Houston Sequencing Core, with the assistance of Yinghong Pan and Utpal Pandya. Computational analyses were performed at the University of Houston Center for Advanced Computing and Data Systems, with some assistance from Adrian Garcia and Shuo Zhang. We thank Erin Kelleher for feedback on the preparation of this manuscript. This work was supported by start-up funds from the University of Houston.

Supplementary Figures

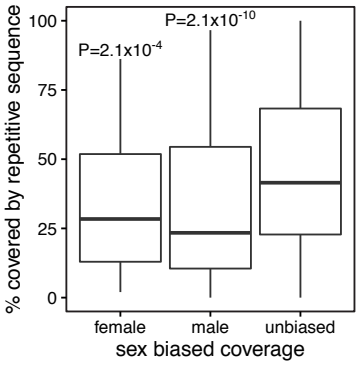

Figure S1: Boxplots show the distribution of the percent of 1 kb windows that contain predicted repetitive sequence. Three different types of 1 kb windows are plotted: those with female-biased read mapping coverage ($\log_2 \frac{M}{F} < -1$), those with male-biased coverage ($\log_2 \frac{M}{F} > 1$), and those with insignificant differences in coverage (unbiased). P values comparing the female- and male-biased windows with the unbiased windows from a Mann-Whitney test are shown.

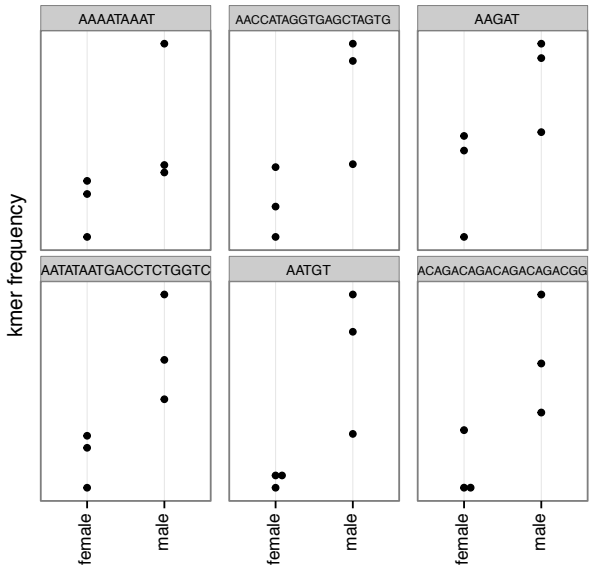

Figure S2: The frequency of the six k-mers over-represented in males is plotted for each of the 3 female and 3 male libraries.

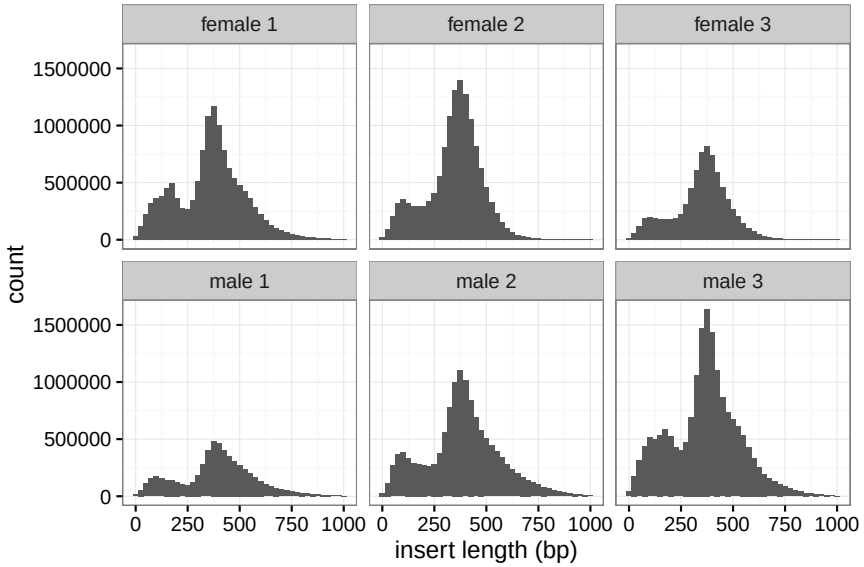

Figure S3: Distributions of insert sizes for the 6 male and female aabys libraries.

References

Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**: 113–124.

Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman TL, Hahn MW, Kitano J, Mayrose I, Ming R, et al.. 2014. Sex determination: why so many ways of doing it? *PLoS Biol* **12**: e1001899.

Baker RH and Wilkinson GS. 2010. Comparative genomic hybridization (CGH) reveals a neo-X chromosome and biased gene movement in stalk-eyed flies (genus *Teleopsis*). *PLoS Genet.* **6**: e1001121.

Beukeboom L and Perrin N. 2014. *The Evolution of Sex Determination.* Oxford University Press, New York, NY.

Bopp D, Saccone G, and Beye M. 2014. Sex determination in insects: variations on a common theme. *Sex. Dev.* **8**: 20–28.

Boyes JW, Corey MJ, and Paterson HE. 1964. Somatic chromosomes of higher diptera IX. Karyotypes of some muscid species. *Can J Cytol* **42**: 1025–1036.

Boyes JW and Van Brink JM. 1965. Chromosomes of calyptrate diptera. *Can. J. Genet. Cytol.* **7**: 537–550.

Bull JJ. 1983. *Evolution of Sex Determining Mechanisms.* Benjamin/Cummings, Menlo Park, CA.

Cakir S and Kence A. 1996. The distribution of males having XY and XX chromosomes in housefly populations (Diptera: Muscidae) of Turkey. *Genetica* **98**: 205–210.

Carvalho AB and Clark AG. 2005. Y chromosome of D. pseudoobscura is not homologous to the ancestral Drosophila Y. *Science* **307**: 108–110.

Carvalho AB and Clark AG. 2013. Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. *Genome Res.* **23**: 1894–1907.

Charlesworth B. 1991. The evolution of sex chromosomes. *Science* **251**: 1030–1033.

Charlesworth B. 1996. The evolution of chromosomal sex determination and dosage compensation. *Curr. Biol.* **6**: 149–162.

Charlesworth D, Charlesworth B, and Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95**: 118–128.

Denholm I, Franco MG, Rubini PG, and Vecchi M. 1983. Identification of a male determinant on the X chromosome of housefly (*Musca domestica* L.) populations in South-East England. *Genet. Res.* **42**: 311–322.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

Dübendorfer A, Hediger M, Burghardt G, and Bopp D. 2002. *Musca domestica*, a window on the evolution of sex-determining mechanisms in insects. *Int J Dev Biol* **46**: 75–79.

Ellegren H. 2011. Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat. Rev. Genet.* **12**: 157–166.

Faber-Hammond JJ, Phillips RB, and Brown KH. 2015. Comparative analysis of the shared sex-determination region (SDR) among salmonid fishes. *Genome Biol. Evol.* **7**: 1972–1987.

Filatov DA, Moneger F, Negrutiu I, and Charlesworth D. 2000. Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. *Nature* **404**: 388–390.

Foster GG, Whitten MJ, Konovalov C, Arnold JTA, and Maffi G. 1981. 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.). *Genet. Res.* **37**: 55–69.

Graves JAM. 2005. Recycling the Y chromosome. *Science* **307**: 50–51.

Hamm RL, Meisel RP, and Scott JG. 2015. The evolving puzzle of autosomal versus Y-linked male determination in *Musca domestica*. *G3* **5**: 371–384.

Hamm RL, Shono T, and Scott JG. 2005. A cline in frequency of autosomal males is not associated with insecticide resistance in house fly (Diptera: Muscidae). *J. Econ. Entomol.* **98**: 171–176.

Hediger M, Minet AD, Niessen M, Schmidt R, Hilfiker-Kleiner D, Cakir S, Nothiger R, and Dubendorfer A. 1998a. The male-determining activity on the Y chromosome of the housefly (*Musca domestica* L.) consists of separable elements. *Genetics* **150**: 651–661.

Hediger M, Niessen M, Müller-Navia J, Nöthiger R, and Dübendorfer A. 1998b. Distribution of heterochromatin on the mitotic chromosomes of *Musca domestica* L. in relation to the activity of male-determining factors. *Chromosoma* **107**: 267–271.

Hughes JF, Skaletsky H, Koutseva N, Pyntikova T, and Page DC. 2015. Sex chromosome-to-autosome transposition events counter Y-chromosome gene loss in mammals. *Genome Biol.* **16**: 1–9.

Kallman KD. 1984. A new look at sex determination in poeciliid fishes. In *Evolutionary Genetics of Fishes* (ed. BJ Turner), pp. 95–171. Springer US, Boston, MA.

Koerich LB, Wang X, Clark AG, and Carvalho AB. 2008. Low conservation of gene content in the Drosophila Y chromosome. *Nature* .

Larracuente AM, Noor MAF, and Clark AG. 2010. Translocation of Y-linked genes to the dot chromosome in *Drosophila pseudoobscura*. *Mol. Biol. Evol.* **27**: 1612–1620.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **1303.3997v2**.

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

Linger RJ, Belikoff EJ, and Scott MJ. 2015. Dosage compensation of X-linked Muller element F genes but not X-linked transgenes in the Australian sheep blowfly. *PLoS ONE* **10**: e0141544.

Liu N and Yue X. 2001. Genetics of pyrethroid resistance in a strain (ALHF) of house flies (Diptera: Muscidae). *Pestic. Biochem. Physiol.* **70**: 151–158.

Liu Z, Moore PH, Ma H, Ackerman CM, Ragiba M, Yu Q, Pearl HM, Kim MS, Charlton JW, Stiles JI, et al.. 2004. A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature* **427**: 348–352.

Love M, Huber W, and Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**: 550.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al.. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**: 18.

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

Meisel RP, Malone JH, and Clark AG. 2012. Disentangling the relationship between sex-biased gene expression and X-linkage. *Genome Res.* **22**: 1255–1265.

Meisel RP, Scott JG, and Clark AG. 2015. Transcriptome differences between alternative sex determining genotypes in the house fly, *Musca domestica*. *Genome Biol. Evol.* **7**: 2051–2061.

Miura I. 2007. An evolutionary witness: the frog *Rana rugosa* underwent change of heterogametic sex from XY male to ZW female. *Sex. Dev.* **1**: 323–331.

Muller HJ. 1940. Bearings of the 'Drosophila' work on systematics. In *The New Systematics* (ed. J Huxley), pp. 185–268. Clarendon Press, Oxford, Oxford.

Patterson JT and Stone WS. 1952. *Evolution in the Genus Drosophila*. The Macmillan Company, New York.

Rice WR. 1996. Evolution of the Y sex chromosome in animals. *Bioscience* **46**: 331–343.

Roco ÁS, Olmstead AW, Degitz SJ, Amano T, Zimmerman LB, and Bullejos M. 2015. Coexistence of Y, W, and Z sex chromosomes in *Xenopus tropicalis*. *Proc. Natl. Acad. Sci. U.S.A.* **112**: E4752–E4761.

Ross JA, Urton JR, Boland J, Shapiro MD, and Peichel CL. 2009. Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). *PLoS Genet.* **5**: e1000391.

Schaeffer SW, Bhutkar A, McAllister BF, Matsuda M, Matzkin LM, O'Grady PM, Rohde C, Valente VLS, Aguade M, Anderson WW, et al.. 2008. Polytene chromosomal maps of 11 Drosophila species: the order of genomic scaffolds inferred from genetic and physical maps. *Genetics* **179**: 1601–1655.

Scott JG and Georghiou GP. 1985. Rapid development of high-level permethrin resistance in a field-collected strain of the house fly (Diptera: Muscidae) under laboratory selection. *J. Econ. Entomol.* **78**: 316–319.

Scott JG, Sridhar P, and Liu N. 1996. Adult specific expression and induction of cytochrome $P450_{tpr}$ in house flies. *Arch. Insect Biochem. Physiol.* **31**: 313–323.

Scott JG, Warren WC, Beukeboom LW, Bopp D, Clark AG, Giers SD, Hediger M, Jones AK, Kasai S, Leichter CA, et al.. 2014. Genome of the house fly, *Musca domestica* L., a global vector of diseases with adaptations to a septic environment. *Genome Biol.* **15**: 466.

Smith CD, Shu S, Mungall CJ, and Karpen GH. 2007. The Release 5.1 annotation of *Drosophila melanogaster* heterochromatin. *Science* **316**: 1586–1591.

Steinemann M and Steinemann S. 1998. Enigma of Y chromosome degeneration: neo-Y and neo-X chromosomes of Drosophila miranda a model for sex chromosome evolution. *Genetica* **102-103**: 409–420.

Stöck M, Horn A, Grossen C, Lindtke D, Sermier R, Betto-Colliard C, Dufresnes C, Bonjour E, Dumas Z, Luquet E, et al.. 2011. Ever-young sex chromosomes in european tree frogs. *PLoS Biol.* **9**: e1001062.

Sved JA, Chen Y, Shearman D, Frommer M, Gilchrist AS, and Sherwin WB. 2016. Extraordinary conservation of entire chromosomes in insects over long evolutionary periods. *Evolution* **70**: 229–234.

Tomita T and Wada Y. 1989. Multifactorial sex determination in natural populations of the housefly (*Musca domestica*) in Japan. *Jpn. J. Genet.* **64**: 373–382.

Traut W and Willhoeft U. 1990. A jumping sex determining factor in the fly *Megaselia scalaris*. *Chromosoma* **99**: 407–412.

Vallender EJ and Lahn BT. 2006. Multiple independent origins of sex chromosomes in amniotes. *Proc. Natl. Acad. Sci. U.S.A.* **103**: 18031–18032.

Veyrunes F, Catalan J, Sicard B, Robinson TJ, Duplantier JM, Granjon L, Dobigny G, and Britton-Davidian J. 2004. Autosome and sex chromosome diversity among the African pygmy mice, subgenus *Nannomys* (Murinae; Mus). *Chromosome Res.* **12**: 369–382.

Vicoso B and Bachtrog D. 2013. Reversal of an ancient sex chromosome to an autosome in *Drosophila*. *Nature* **499**: 332–335.

Vicoso B and Bachtrog D. 2015. Numerous transitions of sex chromosomes in Diptera. *PLoS Biol.* **13**: e1002078.

Vicoso B, Kaiser VB, and Bachtrog D. 2013. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proc. Natl. Acad. Sci. U.S.A.* **110**: 6453–6458.

Wagoner DE. 1967. Linkage group-karyotype correlation in the house fly determined by cytological analysis of X-ray induced translocations. *Genetics* **57**: 729–739.

Wei KHC, Grenier JK, Barbash DA, and Clark AG. 2014. Correlated variation and population differentiation in satellite DNA abundance among lines of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* **111**: 18793–18798.

Weller GL and Foster GG. 1993. Genetic maps of the sheep blowfly *Lucilia cuprina*: linkage-group correlations with other dipteran genera. *Genome* **36**: 495–506.

Woram RA, Gharbi K, Sakamoto T, Hoyheim B, Holm LE, Naish K, McGowan C, Ferguson MM, Phillips RB, Stein J, et al.. 2003. Comparative genome analysis of the primary sex-determining locus in salmonid fishes. *Genome Res.* **13**: 272–280.

Yazdi HP and Ellegren H. 2014. Old but not (so) degenerated—slow evolution of largely homomorphic sex chromosomes in ratites. *Molecular Biology and Evolution* **31**: 1444–1453.