

# The house fly Y Chromosome is young and minimally differentiated from its ancient X Chromosome partner

Richard P. Meisel,<sup>1</sup> Christopher A. Gonzales,<sup>1,2</sup> and Hoang Luu<sup>1</sup>

<sup>1</sup>Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204, USA

Canonical ancient sex chromosome pairs consist of a gene rich X (or Z) Chromosome and a male-limited (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 can arise if an existing sex chromosome fuses to an autosome or an autosome acquires a new sex-determining locus/allele. Sex chromosomes often differ between closely related species and can even be polymorphic within species, suggesting that nascent sex chromosomes arise frequently over the course of evolution. 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 resemble the ancestral fly karyotype that originated ~100 million yr ago; therefore, the house fly is expected to have X and Y Chromosomes with different gene content. We tested this hypothesis using whole-genome sequencing and transcriptomic data, and we discovered little evidence for genetic differentiation between the X and Y in house fly. We propose that the house fly has retained the ancient X Chromosome, but the ancestral Y was replaced by an X Chromosome carrying a new male determining gene. Our proposed hypothesis provides a mechanism for how one member of a sex chromosome pair can experience evolutionary turnover while the other member remains unaffected.

[Supplemental material is available for this article.]

In organisms in which sex is determined by heritable genetic factors, sex determining loci can reside on sex chromosomes. Sex chromosome systems are 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 predicted and observed differences in gene content and evolutionary rates between the X (or Z) and autosomes (Rice 1984; Charlesworth et al. 1987; Vicoso and Charlesworth 2006; Sturgill et al. 2007; Ellegren 2011; Meisel et al. 2012; Meisel and Connallon 2013). In contrast, Y (Z) Chromosomes tend to contain a small number of genes with male-specific (female-specific) functions and are often enriched with repetitive DNA as a result of male-specific (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 an undifferentiated autosomal pair (Bull 1983; Charlesworth 1991). Many species harbor undifferentiated sex chromosomes because they are either of recent origin or noncanonical 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 often 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 different autosome becomes a new sex chromosome (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 (Brachycera), 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 heterochromatic, gene-poor sex chromosome pair (element F is the X Chromosome); this genomic arrangement has been conserved for ~100 million yr in some lineages (Muller 1940; Foster et al. 1981; Weller and Foster 1993; Vicoso and Bachtrog 2013). 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). Many sex chromosome transitions have occurred across Brachycera, including fusions of ancestral autosomes with the X Chromosome, autosomes transitioning into sex chromosomes, and complete reversions of the ancestral X to an autosome (Carvalho and Clark 2005; Baker and Wilkinson 2010; Larracuente et al. 2010; Vicoso and Bachtrog 2013, 2015).

<sup>2</sup>Present address: Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

Corresponding author: [rpmisel@uh.edu](mailto:rpmisel@uh.edu)

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.215509.116>.

© 2017 Meisel et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

The house fly (*Musca domestica*) is a classic model system for studying sex determination and sex chromosomes, because it harbors multiple natural and laboratory variants in sex determining genes and sex chromosomes (Dübendorfer et al. 2002). The house fly karyotype resembles that of the ancestral brachyceran, with five large euchromatic elements and a heterochromatic sex chromosome pair (Boyes et al. 1964). As in other species with that ancestral karyotype, the house fly X and Y Chromosomes can be distinguished based on their length in cytological preparations (Boyes and Van Brink 1965; Denholm et al. 1983; Cakir and Kence 1996; Hediger et al. 1998b). In some close relatives of the house fly that have the ancestral karyotype (e.g., *Lucilia* blow flies) the ancient X and Y Chromosomes are highly differentiated in gene content (Linger et al. 2015; Vicoso and Bachtrog 2015).

Despite the cytological similarities between the house fly and ancestral karyotypes, there are multiple reasons to suspect that house fly has a Y Chromosome that is not differentiated from the X. First, the house fly M factor (*Mdmd*) is a recently arisen duplication of the gene encoding the spliceosome-associated protein CWC22 (Sharma et al. 2017), not an ancient gene as would be expected if it were the ancestral male-determining locus of brachycerans. *Mdmd* has been mapped to the autosomes as well as the Y Chromosome (Sharma et al. 2017), suggesting that the house fly Y and the autosomes harboring *Mdmd* are all recently derived neo-Y Chromosomes. Second, no sex-linked genetic markers have been identified on the house fly X or Y Chromosomes other than *Mdmd* (Hamm et al. 2015), suggesting that there are no X-specific genes or genetic variants that are not found on the Y Chromosome. Third, males with an autosomal *Mdmd* that do not carry a Y Chromosome (XX males) are fertile (Bull 1983; Hamm et al. 2015), suggesting that no essential male fertility genes are unique to the Y Chromosome apart from *Mdmd*. 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. We used whole-genome and transcriptome sequencing to test the hypothesis that the house fly Y Chromosome is young and minimal differentiated from its X Chromosome partner.

## Results

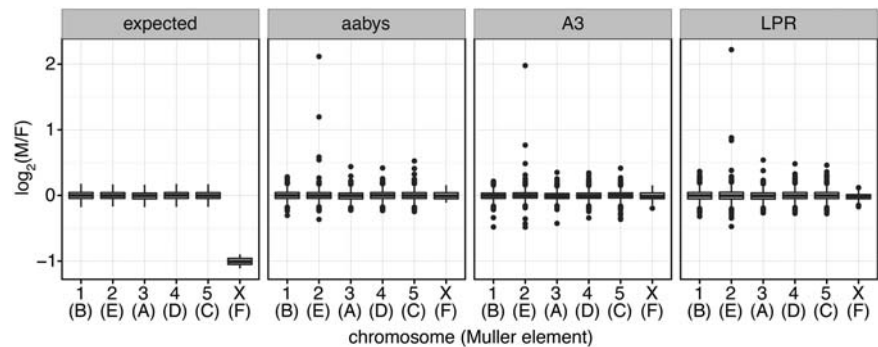
### Very few X-specific sequences in the house fly genome

Our first goal was to identify house fly X Chromosome sequences not found on the Y (X-specific sequences), which would be consistent with the hypothesis that house flies have a differentiated sex chromosome pair. The genome sequencing project used DNA from female flies (XX genotype) to produce the assembly and annotation (Scott et al. 2014), which means there are no Y-specific sequences in the reference. 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-specific genes, we used the Illumina technology to sequence genomic DNA (gDNA) separately

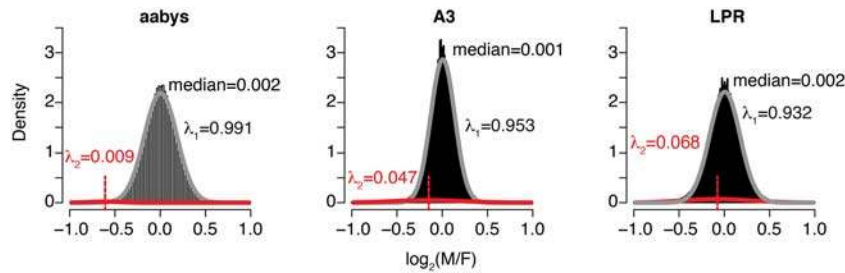
from male (XY) and female (XX) aabys flies (three replicates of each sex), and we aligned the reads to the annotated genome (for read counts, see Supplemental Data S1). If house fly males have a Y Chromosome that is fully differentiated from the X, we expect females to have twice the sequencing coverage ( $\log_2 \frac{M}{F} = -1$ ) within genes on Muller element F (the ancestral X Chromosome) as males (Vicoso and Bachtrog 2013). We instead find that the average sequencing coverage in males and females is almost identical ( $\log_2 \frac{M}{F} \approx 0$ ) for genes on all six chromosomes, and no genes have  $\log_2 \frac{M}{F} \leq -1$  (Fig. 1).

We tested if the lack of X-specific genes is common to two other strains of house fly previously reported to have XY males: A3 and LPR (Scott and Georgiou 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 aabys female reference genome (for read counts, see Supplemental Data S2, S3). Consistent with the results from the aabys strain, the average relative male-to-female sequencing coverage within genes in both A3 and LPR is similar across all six chromosomes (Fig. 1). Because we fail to find any genes with a twofold enrichment in females ( $\log_2 \frac{M}{F} \leq -1$ ), our results suggest that there are no genes on the house fly X Chromosome that are not present on the Y Chromosome. However, LOC101893103 (the ortholog of muscarinic acetylcholine receptor Dm1) has the most “female-biased” coverage across all three strains ( $-0.707 > \log_2 \frac{M}{F} > -0.852$ ), suggesting that it is the best candidate X-specific gene in the house fly genome.

To ensure that our results are not an artifact of poor annotations of house fly X Chromosome genes, we calculated  $\log_2 \frac{M}{F}$  coverage across nonoverlapping 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 twofold enrichment of X Chromosome sequences in females. We do indeed observe that the distributions of  $\log_2 \frac{M}{F}$  are centered near zero for all three house fly strains in our analysis, but there is no obvious secondary peak at  $\log_2 \frac{M}{F} = -1$  in any of the distributions (Fig. 2). To test for a secondary peak at  $\log_2 \frac{M}{F} = -1$ , we fit a mixture of two normal distributions to our data using an expectation-maximization algorithm with starting values of  $\log_2 \frac{M}{F} = \{-1, 0\}$  for the means of the two distributions



**Figure 1.** Expected sequencing coverage in males relative to females ( $\log_2 \frac{M}{F}$ ) in an XY system with a degenerated Y Chromosome (left), and observed coverage in three house fly strains (aabys, A3, and LPR) for each house fly chromosome (Muller elements in parentheses). Chromosome assignments are based on orthology relationships with *Drosophila melanogaster*. Box plots show the median and quartiles, with outliers indicated as points.



**Figure 2.** Histograms are plotted of  $\log_2 \frac{M}{F}$  for 1-kb intervals across three strains. The medians of the distributions are shown. The gray curve shows the normal distribution that fits the majority of intervals (mean  $\approx 0$ ), and the red curve shows the normal distribution that fits the remaining scaffolds (mean  $< 0$ ). The  $\lambda_1$  values are the proportion of observed data estimated to be part of the gray normal distribution centered near zero, and  $\lambda_2$  is the proportion estimated to be part of the red distribution with a mean  $< 0$ . The red vertical lines show the means of the  $\lambda_2$  distributions.

(Benaglia et al. 2009). Most of the 1-kb intervals (93%–99%) are assigned to distributions that are centered near zero, and the remainder of the intervals are assigned to secondary distributions with means less than zero (Fig. 2). However, those secondary distributions all have a mean greater than  $-1$ , suggesting that there are few sequences present in XX females at twice the abundance as in XY males.

#### Very few Y-specific sequences but some autosome-to-Y duplicates in the house fly genome

We next sought to identify Y Chromosome sequences that are absent from the X Chromosome (i.e., the reciprocal of the analyses described above). Alignment to a female (XX) reference genome cannot identify Y-specific sequences because they would be absent from the reference genome. However, we can identify recent Y Chromosome duplicates of autosomal genes because those genes should have 3:2 male:female coverage (1.5 $\times$ ) when aligned to an XX female reference genome (two autosomal copies plus a Y Chromosome copy in males as compared to only two autosomal copies in females). There are five genes with  $\geq 1.5\times$  male-biased coverage ( $\log_2 \frac{M}{F} \geq 0.585$ ) across all three strains (Table 1). Two of the genes are members of large gene families, suggesting possible Y-specific expansions. One of the other three genes is *Md-ncm* (LOC101896466), which is the ancestral autosomal paralog of the male-determining gene *Mdmd* (Sharma et al. 2017). Our results therefore demonstrate that screening for genes with  $1.5 \times \frac{M}{F}$  coverage can identify recent autosome-to-Y duplications.

To identify Y-specific sequences without autosomal paralogs, we first used the male sequencing reads from the aabys strain to assemble a genome that contains a Y Chromosome using SOAPdenovo2 (Luo et al. 2012). It was necessary to assemble a male genome because the genome project sequenced gDNA from

XX 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). We obtain similar results when we use a male genome assembled with ABySS (Simpson et al. 2009) or if we assemble the male genome with SOAPdenovo2 using only male reads that do not align to the female as-

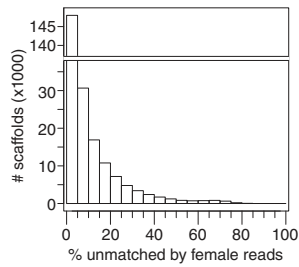
sembly (Supplemental Fig. S1). In contrast, when this approach was used to identify Y Chromosome scaffolds in species with differentiated sex chromosomes (*Drosophila* and humans), a substantial number of Y Chromosome scaffolds were completely unmatched by female sequencing reads (Carvalho and Clark 2013). These results suggest that there are not large segments of the house fly Y Chromosome that are unique from the X Chromosome or autosomes.

We examined the scaffolds from the male house fly genome with a high percentage of sequence unmatched by female reads. We performed blastx searches of the 50 scaffolds with the highest percentage of unmatched sequence (79.7%–96.6%) against the NCBI nonredundant protein database (Altschul et al. 1997). Only six of 50 scaffolds had hits to annotated house fly genes, whereas 27 had hits to transposable element (TE) sequences, three hit other sequences from other species, and 14 had no hits in the database (Supplemental Data S4, S5). The scaffold with the highest percentage of unmatched sequence (96.6%) is 1143 nt long and contains a 219-bp segment that matches an annotated house fly gene on Chromosome 1 that is homologous to a *D. melanogaster* gene with a predicted membrane associated GRAM domain (CG34392). This scaffold does not have any blastn or blastx hits to other sequences in the database. In addition, there are 22 scaffolds that are both  $>5$  kb and  $>50\%$  unmatched by female reads (Supplemental Data S4, S6). We also performed blastx of those scaffolds against the NCBI database, and we found that 15 of 22 hit a TE, four hit an annotated house fly gene, and three hit another sequence from a different species. None of the annotated house fly genes hit by these 72 scaffolds are predicted to be on element F (the house fly X Chromosome). Most of these scaffolds (42/72) have sequence similarity with a TE, suggesting that the Y Chromosome may contain unique repetitive sequences or be enriched for particular repeat classes. Notably, none of the scaffolds

**Table 1.** Genes with  $\geq 1.5\times$  read mapping coverage in males relative to females in all three strains

House fly gene	min $\log_2 \frac{M}{F}$	Predicted <i>D. melanogaster</i> homolog(s)	Gene family
LOC101901673	1.97	FBgn0031850	1:1 ortholog
LOC101890714	1.43	Cytochrome P450	Large family
LOC101892987	0.82	Activator of 90 kDa HSP ATPase	Unknown
LOC101896466	0.76	FBgn0086707	1:1 ortholog
LOC101900982	0.69	FBgn0033928 and FBgn0033926	Large family

The lowest  $\log_2 \frac{M}{F}$  across the three strains is shown (min  $\log_2 \frac{M}{F}$ ). The “gene family” column shows the homology relationship of the annotated autosomal gene with *D. melanogaster* (Scott et al. 2014).



**Figure 3.** Histogram of the percentage of each scaffold in the male genome assembly that is unmatched by female reads.

with a high percentage of sequence unmatched by female reads contain any sequences that resemble *Mdmd/Md-ncm*, suggesting that this *k*-mer comparison approach is not effective at identifying very recent autosome-to-Y duplications.

### Moderate differences in sequence abundance between house fly males and females

We next examined whether housefly X and Y Chromosomes 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-derived sequences to the reference (XX female) genome. As the input into the PC analysis, we used the number of reads from each of the XY male and XX female sequencing libraries that mapped to each nonoverlapping 1-kb interval. 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.

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 of these “sex-biased” windows: 63 are >twofold enriched in females, and 151 are >twofold enriched in males (Supplemental Data S7). 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., TEs 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 two of 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 of 151 (39.1%) male-enriched windows are within a scaffold that maps to a chromosome (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 chromosomes (2033/3096 windows with  $P > 0.05$  and  $|\log_2 \frac{M}{F}| < 0.01$ ). These “unbiased” windows

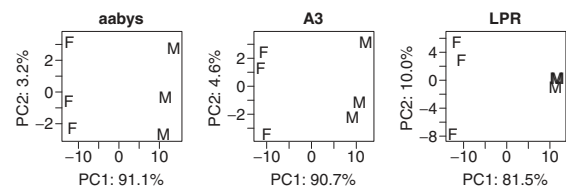
are more likely to be mapped to a chromosome than the sex-biased windows ( $P < 10^{-15}$  in Fisher’s exact test), providing some evidence that differential coverage between males and females could be driven by repeat content differences between the X and Y Chromosomes.

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 (Supplemental Data S7). Similarly, nearly all (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 (Supplemental Fig. S2). However, these analyses are limited because most ( $\geq 52\%$ ) of the assembled 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.

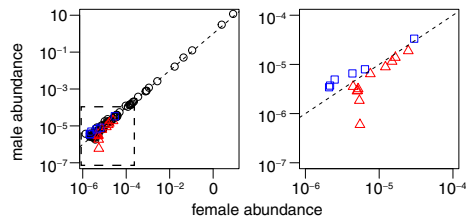
As a second approach to identify candidate X- or Y-enriched sequences, we first determined the abundances of all possible 2–10 mers in the male and female aabys sequencing reads. This approach will identify smaller sequence motifs that may differentiate the X and Y Chromosomes than the analysis described above, and it does not require any a priori repeat annotations. However, 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 overrepresented 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 overrepresented in males using this cutoff, but they are all <twofold enriched in males (Fig. 5; Supplemental Fig. S3). These results suggest that short sequence repeats do not predominantly differentiate the X and Y Chromosomes.

### Relative heterozygosity in males and females suggests that the house fly Y Chromosome is very young

Our data suggest that there are very few X- or Y-specific sequences in the house fly genome. We therefore hypothesize that the house fly Y Chromosome is actually an ancestral brachyceran X Chromosome that recently acquired the male-determining *Mdmd* gene (Sharma et al. 2017). Although recently derived neo-Y Chromosomes may not differ much in gene content from the gametologous X Chromosome, modest sequence-level X-Y



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



**Figure 5.** Minimal differences between  $k$ -mer abundances in male and female genomes. The *left* plot shows the abundances of the 100 most common  $k$ -mers in the male and female sequencing reads averaged across the three libraries from each sex in the aabys strain. The dashed box in the *left* graph indicates the subset of the range plotted on the *right* graph, which contains only  $k$ -mers in which abundances in all three libraries from one sex are greater than the three libraries from the other sex. Red triangles indicate  $k$ -mers in which abundances in all three female libraries are greater than the three male libraries; blue squares indicate  $k$ -mers that are more abundant in male libraries. The dashed line indicates equal representation in males and females.

differentiation can result in elevated heterozygosity within sex chromosome genes in males (Vicoso and Bachtrog 2015). We tested for elevated heterozygosity by first identifying polymorphic sites (SNPs) within genes in aabys males and females. Heterozygosity is elevated in X Chromosome (element F) genes relative to autosomes in both males and females (Supplemental Fig. S5). However, when we compare the proportion of heterozygous SNPs in males relative to females for genes on each chromosome (Fig. 6A), genes on the X Chromosome resemble autosomal genes with equivalent heterozygosity in males and females ( $P=0.45$  in a Mann-Whitney  $U$  test comparing male:female heterozygosity on element F with the other chromosomes). This result demonstrates that the house fly Y is so young that Y Chromosome genes have not yet accumulated modest sequence differences from the X Chromosome.

Some house fly males carry *Mdmd* on the third chromosome ( $III^M$ ), and  $III^M$  is expected to be a young neo-Y (Hamm et al. 2015; Sharma et al. 2017). Males that are heterozygous for  $III^M$  and a standard third chromosome (hereafter  $III^M$  males) can have two copies of the X Chromosome (Hamm et al. 2015). If the  $III^M$  chromosome is a young neo-Y, we expect that  $III^M$  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 within genes in  $III^M$  males relative to XY males (Fig. 6B). 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  $U$  test comparing Chromosome III with the other autosomes).  $III^M$  males additionally have an elevated number of strain-specific SNPs on the third chromosome (Supplemental Fig. S6). Surprisingly, there is also increased heterozygosity on the X Chromosome in  $III^M$  males relative to XY males ( $P=10^{-4}$  in a Mann-Whitney  $U$  test comparing the X Chromosome with Chromosomes 1, 2, 4, and 5) even though  $III^M$  males have the XX genotype. We also observe elevated heterozygosity on the third chromosome and X Chromosome in  $III^M$  males relative to females of the same strain, but not on the X Chromosome in XY males relative to XX females (Supplemental Fig. S4). These results further support our conclusion that, other than a handful of autosome-to-Y duplications, house fly Y Chromosome genes are not differentiated from X Chromosome genes. In contrast, the  $III^M$  chromosome harbors evidence that it is partially differentiated from the non-M-bearing third chromosome.

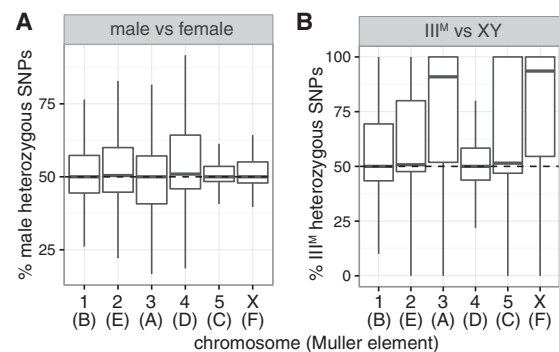
## Discussion

We found very little evidence for differentiation between the X and Y Chromosomes in house fly, despite the fact that the house fly has a karyotype that resembles the  $\sim 100$  million-yr-old ancestral brachyceran karyotype (Boyes et al. 1964; Foster et al. 1981; Weller and Foster 1993; Vicoso and Bachtrog 2013). There are few sequences unique to the X or Y (Figs. 1–3), little evidence for differential abundance of sequences on the X and Y (Fig. 5, but see Fig. 4), and no elevated heterozygosity within X Chromosome genes in XY males (Fig. 6). The strongest evidence for Y Chromosome genes not found on the X are candidate autosome-to-Y duplications, including the male-determining gene *Mdmd* (Table 1). We conclude that the house fly X Chromosome has few genes (if any) not found on the Y Chromosome, and there are only a handful of recently acquired Y-specific genes in addition to *Mdmd*. This is consistent with previous experiments that failed to identify sex-linked markers and found that XX, XO, and YO flies are viable and fertile (Bull 1983; Hediger et al. 1998a; Hamm et al. 2015). Additionally, XY males have equal or greater expression of X Chromosome genes when compared to XX ( $III^M$ ) males (Meisel et al. 2015), providing further evidence that XY males do not have a haploid X Chromosome dose. Our results suggest that the house fly Y Chromosome is an ancestral brachyceran X Chromosome that very recently acquired *Mdmd* via the duplication of *Md-ncm* (Sharma et al. 2017).

### X-Y differentiation in house fly

We detect two forms of minimal X-Y differentiation in the house fly genome. First, there are a few candidate recent autosome-to-Y duplications, including *Mdmd* (Table 1). Autosome-to-Y duplications are a well-documented source of new Y Chromosome genes across animals (Carvalho et al. 2000, 2001, 2015; Matsuda et al. 2002; Nanda et al. 2002; Hattori et al. 2012; Yano et al. 2012; Hall et al. 2013). Second, there is some evidence for differential abundance of sequences between the X and Y Chromosomes (Fig. 4), but we are unable to identify the specific sequences that are differentially represented (Fig. 5).

Our results are consistent with cytological examinations that have identified morphological differences between the house fly X



**Figure 6.** There is 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 percentage of heterozygous SNPs within genes on each chromosome in either XY aabys males relative to XX aabys females using genomic DNA sequences (A) or  $III^M$  males relative to XY males using RNA-seq data (B). Values  $>50\%$  indicate elevated heterozygosity in XY males or  $III^M$  males. The median across all autosomes is indicated by a dashed line.

and Y Chromosomes (Boyes et al. 1964; Hediger et al. 1998b). These morphological differences may be the result of differentially abundant repetitive sequences between the X and Y (Fig. 4). Additionally, in situ hybridizations of chromosomal dissections to mitotic chromosomes detected Y-specific, but not X-specific, segments of the house fly genome (Hediger et al. 1998b). The sequences of these chromosomal segments are unknown, but they presumably include *Mdmd* and possibly a handful of other recent autosome-to-Y duplications (Table 1). Similarly, cytological analyses previously characterized separate X Chromosomes carrying *Mdmd* ( $X^M$ ) that were thought to be different from the Y Chromosome (Denholm et al. 1983; Cakir and Kence 1996). Our results suggest that the Y and  $X^M$  Chromosomes are morphological variants of neo-Y Chromosomes that arose when at least one ancestral X Chromosome recently acquired *Mdmd*. Such morphological or repeat content variation in Y Chromosomes has been previously documented in *Drosophila* and other dipterans (Dobzhansky 1935; Miller and Stone 1962; Miller and Roy 1964; Lyckegaard and Clark 1989; Lemos et al. 2008; Hall et al. 2016).

**Creation of house fly neo-Y Chromosomes**

We hypothesize that the recent duplication of *Mdncm* that created *Mdmd* (Sharma et al. 2017) arose on an X Chromosome, transforming it into a neo-Y (Fig. 7). The ancestral fly Y Chromosome would have subsequently been lost from house fly populations if it did not contain any essential genes (other than the ancestral male determiner), or it could have fused to an autosome (Carvalho and Clark 2005; Larracuente et al. 2010). Alternatively, the house fly Y Chromosome could have arisen through the fusion of the ancestral Y and X Chromosomes. However, after an X-Y fusion, the neo-Y should retain ancestral Y-specific sequences, which we fail to detect.

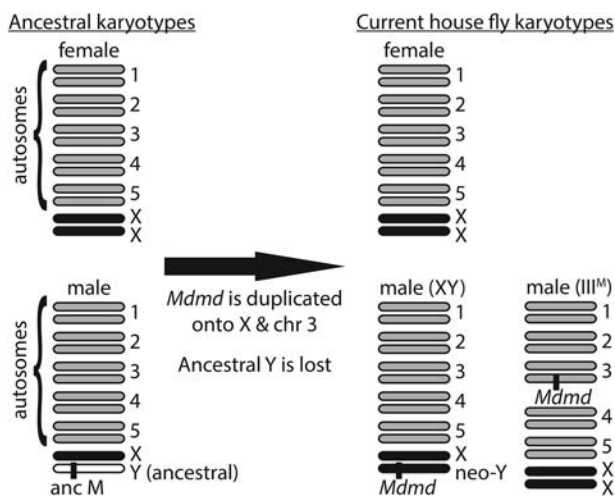
The III<sup>M</sup> Chromosome is also a neo-Y that arose when *Mdmd* was duplicated onto a standard third chromosome (Sharma et al. 2017). Curiously, III<sup>M</sup> males have elevated heterozygosity in X Chromosome genes relative to XX females and XY males (Fig. 6; Supplemental Fig. S4). One possible cause of this elevated heterozygosity is that some X Chromosome genes were translocated onto III<sup>M</sup> along with *Mdmd*. The elevated X Chromosome heterozygosity

we detect in III<sup>M</sup> males would therefore be the result of those males being triploid for X Chromosome genes. No nullo-X/Y flies carrying III<sup>M</sup> have been identified to our knowledge, suggesting that the X and Y Chromosomes contain some essential genes not translocated to III<sup>M</sup>. Additional work is necessary to determine the nature of the translocations or duplications of *Mdmd* that created the Y and III<sup>M</sup> Chromosomes.

Our results are also suggestive of the order of events that created the Y and III<sup>M</sup> Chromosomes. III<sup>M</sup> males have elevated heterozygosity on the third chromosome, whereas XY males do not have elevated heterozygosity on element F (the ancestral X Chromosome) (Fig. 6). This suggests that the Y Chromosome is a younger neo-Y than III<sup>M</sup>, but there is an alternative explanation for the patterns of polymorphism. Differentiation between nascent X and Y Chromosomes is accelerated by suppressed recombination in XY males (Charlesworth 1991; Rice 1996; Charlesworth et al. 2005; Bachtrog 2013). Lack of recombination can be an inherent property of male meiosis (as in *Drosophila*) or arise via Y Chromosome inversions that suppress crossing over between the X and Y. There is evidence for male recombination in house flies (Feldmeyer et al. 2010), suggesting that chromosomal inversions would be required for recombination suppression between the house fly X and Y. If the III<sup>M</sup> Chromosome carries inversions and the Y Chromosome does not, then the elevated heterozygosity in III<sup>M</sup> males but not XY males could be a result of inversions accelerating the rate of divergence between the III<sup>M</sup> Chromosome and the standard third chromosome (Navarro et al. 2000). However, element F in *Drosophila* does not experience crossing over in meiosis, and estimates of element F recombination rates from population genetic data are extremely low (Wang et al. 2002; Arguello et al. 2010). This suggests that Y Chromosome inversions might not be necessary to suppress X-Y recombination in house fly. Additional sequencing of XY and III<sup>M</sup> males is needed to test for inversions or other recombination suppressors, which could shed light on the lack of X-Y differentiation.

**Translocating sex-determining loci can cause sex chromosome recycling and create cryptic neo-sex chromosomes**

Our results provide the first evidence, to our knowledge, of the “recycling” of a sex chromosome pair through the creation of a nascent Y (or W) from an ancient X (or Z) Chromosome (Graves 2005). The recycling of differentiated sex chromosomes into a neo-sex chromosome pair in house fly appears to have happened without a 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, 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, however, involve nascent sex chromosomes, not ancient sex chromosomes as in house fly. Moreover, the sex chromosome transitions in fish and amphibians all involve a change in the heterogametic sex (i.e., XY males to ZW females, or vice versa), whereas the house fly X and Y Chromosomes did not switch to a Z and W.



**Figure 7.** Ancestral karyotypes and hypothesized derived house fly karyotypes.

We hypothesize that the X-to-Y conversion in house fly occurred because *Mdmd* was duplicated onto an X Chromosome (Sharma et al. 2017). Gene duplications have given rise to new male-determining loci on neo-Y Chromosomes in other taxa (Matsuda et al. 2002; Nanda et al. 2002; Hattori et al. 2012), translocating sex determining loci have been observed in other animals (Traut and Willhoeft 1990; Woram et al. 2003; Faber-Hammond et al. 2015), and there is rampant gene traffic to and from other long-established Y Chromosomes (Koerich et al. 2008; Hughes et al. 2015). These different forms of gene movement suggest that new male-determining (female-determining) loci may arise on established X (Z) Chromosomes in other evolutionary lineages, causing X-to-Y (or W-to-Z) transitions. 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 (or W-to-Z) transitions may have occurred in other taxa and remain cryptic because the karyotype has remained unchanged.

## Methods

### Fly strains

We used five house fly strains to identify X and Y Chromosome sequences. One strain, Cornell susceptible (CS), has X/X; III<sup>M</sup>/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 observed equivalent inheritance in males and females of all aabys phenotypic markers (i.e., no sex-linked inheritance), 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.

### 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 (assembly version 2.0.2) to assign house fly genomic scaffolds to chromosome arms using a majority rule as described previously (Meisel et al. 2015). Briefly, scaffolds were assigned to a Muller element if the majority of genes on the scaffold with 1:1 *D. melanogaster* orthologs have orthologs on the same *D. melanogaster* element. In total, 62 house fly genes have 1:1 *D. melanogaster* orthologs on Muller element F, which amounts to roughly three-quarters of the approximately 80 genes on *Drosophila* element F (Leung et al. 2010). We used these 1:1 orthologs to assign seven house fly scaffolds to element F (the X Chromosome), and those seven scaffolds contain 51 genes.

We repeated all of the analyses described in “Results” using only genes with 1:1 *D. melanogaster* orthologs and obtained qualitatively similar results as when we used scaffold-level Muller element assignments.

We sequenced 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 high-output run of the machine. We also sequenced gDNA from three replicates of male and female heads from A3 and LPR flies (12 samples total) in another single high-output run on the NextSeq 500 using 75-bp paired-end reads. For each of the 18 sequencing libraries, DNA was extracted from separate pools of fly heads using the QIAGEN DNeasy blood and tissue kit. 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 in which both ends of a sequenced fragment mapped to the same scaffold in the reference genome. Reads that failed to meet these criteria were considered unmapped for the male genome assembly described below. Mapping statistics are presented in Supplemental Table S1. Mapped reads were assigned to annotated genes if mapping coordinates overlap with at least 1 bp of the coordinates of an annotated gene (from beginning to end point of annotation). Reads could therefore map to exons or introns.

We additionally assembled the reads from aabys male samples using SOAPdenovo2 (Luo et al. 2012) and ABySS (Simpson et al. 2009) to construct a reference genome that contains Y Chromosome sequences. Mapping our sequence data to the reference genome revealed that our average insert size was 370 bp (Supplemental Fig. S7), which was used as a parameter in the SOAPdenovo2 genome assembly, along with a pair number cutoff of 3 and a minimum alignment length 32 bp. For the ABySS assembly, we used a *k*-mer pair span (*k*) of 64. We also assembled a genome from only male reads that did not align to the female genome reference assembly using SOAPdenovo2 (Luo et al. 2012). For downstream analyses, we only retained scaffolds with a length  $\geq 1000$  bp in each assembly. Assembly statistics are presented in Supplemental Table S2.

### Identifying X and Y Chromosome sequences

We used four differential coverage approaches to identify candidate X and Y Chromosome sequences in the house fly genome. The first approach identifies X Chromosome genes or sequences by testing for twofold higher abundance in females relative to males (Vicoso and Bachtrog 2013). To do this, we used DESeq2 to calculate  $\log_2 \frac{M}{F}$  within individual genes and 1-kb windows across the three male and female derived libraries for each strain (Love et al. 2014). We also used DESeq2 to calculate *P*-values for differential coverage between females and males. This approach was also used to identify candidate autosome-to-Y duplicates with  $\geq 1.5\times$  coverage in males relative to females.

The second approach was used to identify Y Chromosome 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, used the male sequencing reads to construct a validating bit-array, and implemented the options described by Carvalho and Clark (2013) for identifying Y Chromosome sequences in *Drosophila* genomes (Supplemental Methods S1, S2).

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 2–10 mers 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-seq 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 in three consecutive iterations, 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). An example script with our SNP calling pipeline is available in Supplemental Methods S3.

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<sup>M</sup> 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 two 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.

## Data access

All data generated in this study have been submitted to NCBI BioProject (<https://www.ncbi.nlm.nih.gov/bioproject>) under the umbrella accession number PRJNA383366. Raw sequencing reads have been submitted to the NCBI Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) under accession numbers SRX2154714–SRX2154719. The male genome assembly has been submitted to the NCBI Genome database ([\[ncbi.nlm.nih.gov/genome/\]\(https://www.ncbi.nlm.nih.gov/genome/\)\) under accession number NDYK00000000. Variant calls have been submitted to dbSNP \(<https://www.ncbi.nlm.nih.gov/snp>\) under the submitter handle MEISEL and associated with BioProject PRJNA382546.](https://www.</a></p>
</div>
<div data-bbox=)

## Acknowledgments

This project was initiated during discussions with Andy Clark and Rob Unckless, who provided valuable comments throughout the completion of this work. Jeff Scott kindly supplied the A3 and LPR flies. Daniel Bopp and Leo Beukeboom shared their results identifying the *Mdmd* gene. 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 and three anonymous reviewers for feedback on the preparation and revisions of this manuscript. This work was supported by start-up funds from the University of Houston.

## References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Arguello JR, Zhang Y, Kado T, Fan C, Zhao R, Innan H, Wang W, Long M. 2010. Recombination yet inefficient selection along the *Drosophila melanogaster* subgroup's fourth chromosome. *Mol Biol Evol* **27**: 848–861.
- 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, 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.
- Benaglia T, Chauveau D, Hunter DR, Young DS. 2009. mixtools: an R package for analyzing mixture models. *J Stat Softw* **32**: 6.
- Beukeboom L, Perrin N. 2014. *The evolution of sex determination*. Oxford University Press, New York.
- Bopp D, Saccone G, Beye M. 2014. Sex determination in insects: variations on a common theme. *Sex Dev* **8**: 20–28.
- Boyes JW, Van Brink JM. 1965. Chromosomes of calyptrate diptera. *Can J Genet Cytol* **7**: 537–550.
- Boyes JW, Corey MJ, Paterson HE. 1964. Somatic chromosomes of higher diptera: IX. Karyotypes of some muscid species. *Can J Cytol* **42**: 1025–1036.
- Bull JJ. 1983. *Evolution of sex determining mechanisms*. Benjamin/Cummings, Menlo Park, CA.
- Cakir S, 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, Clark AG. 2005. Y chromosome of *D. pseudoobscura* is not homologous to the ancestral *Drosophila* Y. *Science* **307**: 108–110.
- Carvalho AB, Clark AG. 2013. Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. *Genome Res* **23**: 1894–1907.
- Carvalho AB, Lazzaro BP, Clark AG. 2000. Y chromosomal fertility factors *kl-2* and *kl-3* of *Drosophila melanogaster* encode dynein heavy chain polypeptides. *Proc Natl Acad Sci* **97**: 13239–13244.
- Carvalho AB, Dobo BA, Vibranovski MD, Clark AG. 2001. Identification of five new genes on the Y chromosome of *Drosophila melanogaster*. *Proc Natl Acad Sci* **98**: 13225–13230.
- Carvalho AB, Vicoso B, Russo CAM, Swenor B, Clark AG. 2015. Birth of a new gene on the Y chromosome of *Drosophila melanogaster*. *Proc Natl Acad Sci* **112**: 12450–12455.
- 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 B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and autosomes. *Am Nat* **130**: 113–146.
- Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95**: 118–128.



- Denholm I, Franco MG, Rubini PG, 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, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.
- Dobzhansky T. 1935. The Y chromosome of *Drosophila pseudoobscura*. *Genetics* **20**: 366–376.
- Dübendorfer A, Hediger M, Burghardt G, 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, Brown KH. 2015. Comparative analysis of the shared sex-determination region (SDR) among salmonid fishes. *Genome Biol Evol* **7**: 1972–1987.
- Feldmeyer B, Pen I, Beukeboom LW. 2010. A microsatellite marker linkage map of the housefly, *Musca domestica*: evidence for male recombination. *Insect Mol Biol* **19**: 575–581.
- Filatov DA, Moneger F, Negrutiu I, 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, Kononov C, Arnold JTA, 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.
- Hall AB, Qi Y, Timoshevskiy V, Sharakhova MV, Sharakhov IV, Tu Z. 2013. Six novel Y chromosome genes in *Anopheles* mosquitoes discovered by independently sequencing males and females. *BMC Genomics* **14**: 273–273.
- Hall AB, Papathanos PA, Sharma A, Cheng C, Akbari OS, Assour L, Bergman NH, Cagnetti A, Crisanti A, Dottorini T, et al. 2016. Radical remodeling of the Y chromosome in a recent radiation of malaria mosquitoes. *Proc Natl Acad Sci* **113**: E2114–E2123.
- Hamm RL, Shono T, 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.
- Hamm RL, Meisel RP, Scott JG. 2015. The evolving puzzle of autosomal versus Y-linked male determination in *Musca domestica*. *G3* **5**: 371–384.
- Hattori RS, Murai Y, Oura M, Masuda S, Majhi SK, Sakamoto T, Fernandez JJ, Somoza GM, Yokota M, Strüssmann CA. 2012. A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proc Natl Acad Sci* **109**: 2955–2959.
- Hediger M, Minet AD, Niessen M, Schmidt R, Hilfinger-Kleiner D, Cakir S, Nöthiger R, Dübendorfer 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, 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, 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. Turner BJ), pp. 95–171. Springer US, Boston, MA.
- Koerich LB, Wang X, Clark AG, Carvalho AB. 2008. Low conservation of gene content in the *Drosophila* Y chromosome. *Nature* **456**: 949–951.
- Larracuente AM, Noor MA, Clark AG. 2010. Translocation of Y-linked genes to the dot chromosome in *Drosophila pseudoobscura*. *Mol Biol Evol* **27**: 1612–1620.
- Lemos B, Araripe LO, Hartl DL. 2008. Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. *Science* **319**: 91–93.
- Leung W, Shaffer CD, Cordonnier T, Wong J, Itano MS, Slawson Tempel EE, Kellmann E, Desruisseau DM, Cain C, Carrasquillo R, et al. 2010. Evolution of a distinct genomic domain in *Drosophila*: comparative analysis of the dot chromosome in *Drosophila melanogaster* and *Drosophila virilis*. *Genetics* **185**: 1519–1534.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 1303.3997v2.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760.
- Linger RJ, Belikoff EJ, 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, 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 JL, et al. 2004. A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature* **427**: 348–352.
- Love M, Huber W, 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.
- Lyckegaard EM, Clark AG. 1989. Ribosomal DNA and Stellate gene copy number variation on the Y chromosome of *Drosophila melanogaster*. *Proc Natl Acad Sci* **86**: 1944–1948.
- Matsuda M, Nagahama Y, Shinomiya A, Sato T, Matsuda C, Kobayashi T, Morrey CE, Shibata N, Asakawa S, Shimizu N, et al. 2002. *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* **417**: 559–563.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky 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, Connallon T. 2013. The faster-X effect: integrating theory and data. *Trends Genet* **29**: 537–544.
- Meisel RP, Malone JH, Clark AG. 2012. Disentangling the relationship between sex-biased gene expression and X-linkage. *Genome Res* **22**: 1255–1265.
- Meisel RP, Scott JG, Clark AG. 2015. Transcriptome differences between alternative sex determining genotypes in the house fly, *Musca domestica*. *Genome Biol Evol* **7**: 2051–2061.
- Miller DD, Roy R. 1964. Further data on Y chromosome types in *Drosophila athabasca*. *Can J Genet Cytol* **6**: 334–348.
- Miller DD, Stone LE. 1962. Reinvestigation of karyotype in *Drosophila affinis* Sturtevant and related species. *J Hered* **53**: 12–24.
- 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. Huxley J), pp. 185–268. Clarendon Press, Oxford, UK.
- Nanda I, Kondo M, Hornung U, Asakawa S, Winkler C, Shimizu A, Shan Z, Haaf T, Shimizu N, Shima A, et al. 2002. A duplicated copy of *DMRT1* in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proc Natl Acad Sci* **99**: 11778–11783.
- Navarro A, Barbadilla A, Ruiz A. 2000. Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in *Drosophila*. *Genetics* **155**: 685–698.
- Patterson JT, Stone WS. 1952. *Evolution in the genus Drosophila*. The Macmillan Company, New York.
- Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* **38**: 735–742.
- Rice WR. 1996. Evolution of the Y sex chromosome in animals. *Bioscience* **46**: 331–343.
- Roco AS, Olmstead AW, Degitz SJ, Amano T, Zimmerman LB, Bullejos M. 2015. Coexistence of Y, W, and Z sex chromosomes in *Xenopus tropicalis*. *Proc Natl Acad Sci* **112**: E4752–E4761.
- Ross JA, Urton JR, Boland J, Shapiro MD, 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 VL, Aguadé M, Anderson WW, et al. 2008. Polytenic chromosomal maps of 11 *Drosophila* species: the order of genomic scaffolds inferred from genetic and physical maps. *Genetics* **179**: 1601–1655.
- Scott JG, 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, Liu N. 1996. Adult specific expression and induction of cytochrome P450<sub>trp</sub> 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.
- Sharma A, Heinze SD, Wu Y, Kohlbrenner T, Morilla I, Brunner C, Wimmer EA, van de Zande L, Robinson MD, Beukeboom LW, et al. 2017. Male sex in houseflies is determined by *Mdmx*, a paralog of the generic splice factor gene *CWC22*. *Science* **356**: 642–645.
- Simpson JT, Wong K, Jackman SD, Schein JE, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res* **19**: 1117–1123.
- Smith CD, Shu S, Mungall CJ, Karpen GH. 2007. The release 5.1 annotation of *Drosophila melanogaster* heterochromatin. *Science* **316**: 1586–1591.

- Steinemann M, 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.
- Sturgill D, Zhang Y, Parisi M, Oliver B. 2007. Demasculinization of X chromosomes in the *Drosophila* genus. *Nature* **450**: 238–241.
- Tomita T, Wada Y. 1989. Multifactorial sex determination in natural populations of the housefly (*Musca domestica*) in Japan. *Jpn J Genet* **64**: 373–382.
- Traut W, Willhoeft U. 1990. A jumping sex determining factor in the fly *Megaselia scalaris*. *Chromosoma* **99**: 407–412.
- Vallender EJ, Lahn BT. 2006. Multiple independent origins of sex chromosomes in amniotes. *Proc Natl Acad Sci* **103**: 18031–18032.
- Veyrunes F, Catalan J, Sicard B, Robinson TJ, Duplantier JM, Granjon L, Dobigny G, 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, Bachtrog D. 2013. Reversal of an ancient sex chromosome to an autosome in *Drosophila*. *Nature* **499**: 332–335.
- Vicoso B, Bachtrog D. 2015. Numerous transitions of sex chromosomes in Diptera. *PLoS Biol* **13**: e1002078.
- Vicoso B, Charlesworth B. 2006. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet* **7**: 645–653.
- Vicoso B, Kaiser VB, Bachtrog D. 2013. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proc Natl Acad Sci* **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.
- Wang W, Thornton K, Berry A, Long M. 2002. Nucleotide variation along the *Drosophila melanogaster* fourth chromosome. *Science* **295**: 134–137.
- Wei KHC, Grenier JK, Barbash DA, Clark AG. 2014. Correlated variation and population differentiation in satellite DNA abundance among lines of *Drosophila melanogaster*. *Proc Natl Acad Sci* **111**: 18793–18798.
- Weller GL, 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.
- Yano A, Guyomard R, Nicol B, Jouanno E, Quillet E, Klopp C, Cabau C, Bouchez O, Fostier A, Guiguen Y. 2012. An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. *Curr Biol* **22**: 1423–1428.
- Yazdi HP, Ellegren H. 2014. Old but not (so) degenerated—slow evolution of largely homomorphic sex chromosomes in ratites. *Mol Biol Evol* **31**: 1444–1453.

Received September 1, 2016; accepted in revised form June 8, 2017.