Submission for Article
Fit in the Discoveries

Gene-level, but not chromosome-wide, divergence between a very young house fly Y chromosome and its homologous X chromosome

Jae Hak Son[1, 2, 3] and Richard P. Meisel[1, 4]

1. Department of Biology and Biochemistry, University of Houston, Houston, TX, USA

2. Current Address: Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, USA

3. Jae Hak Son: jaehak.son@yale.edu

4. Richard P. Meisel: rpmeisel@uh.edu

Abstract

X and Y chromosomes are derived from a pair of homologous autosomes, which then diverge from each other over time. Although Y-specific features have been characterized in sex chromosomes of various ages, the earliest stages of Y chromosome evolution are poorly understood. In particular, we do not know whether early stages of Y chromosome evolution consist of changes to individual genes or happen via chromosome-scale divergence from the X. To address this question, we used house fly, *Musca domestica*, as a model because it has very young sex chromosomes that are still segregating as polymorphisms within natural populations. To identify early differentiation between the very young X and Y chromosomes, we compared genotypic (XY) and sex-reversed (XX) males in gene sequence and gene expression using RNA-seq and Oxford Nanopore sequencing data. There is an excess of genes with divergent expression between the X and Y copies, but the number of genes is small. This suggests that individual Y genes, but not the entire Y chromosome, have diverged from their homologous X-linked alleles. We identified one gene, encoding an axonemal dynein assembly factor (which functions in sperm motility), that has higher expression in the abdomens of XY males than XX males because of a disproportionate contribution of the Y allele to gene expression. The up-regulation of the Y-linked copy of this gene may be favored in males because of its function in spermatogenesis, consistent with sexually antagonistic selection affecting the expression evolution of individual genes during early Y chromosome evolution.

Introduction

In many organisms with two separate sexes, a gene on a sex chromosome determines whether an individual develops into a male or female. In XX/XY sex chromosome systems, males are the heterogametic sex (XY genotype), and females are homogametic with the XX genotype (Bull 1983). All X and Y chromosomes are ultimately derived from a pair of ancestral autosomes that became sex chromosomes when one homolog obtained a sex-determining locus, such as a male-determining gene on a Y chromosome. The X and Y chromosomes then diverge from each other over time (Bull 1983; Charlesworth et al. 2005). Sex chromosomes have originated and diverged from each other in multiple independent evolutionary lineages (Bachtrog et al. 2014; Beukeboom and Perrin 2014)

Despite their independent origins, non-homologous Y chromosomes share many common features across species (Charlesworth et al. 2005). First, "masculinization" occurs because male-limited inheritance of the Y chromosome favors the fixation of male-beneficial genetic variants (Rice 1996). Second, suppressed recombination between the X and Y chromosomes evolves, possibly because selection favors co-inheritance of the male-beneficial alleles and the male-determining locus (Charlesworth 2018). Third, "degeneration" occurs in nonrecombining regions; functional genes that were present on ancestral autosomes become pseudogenes on the Y chromosome because suppressed recombination between the X and Y inhibits the purging of deleterious mutations in Y-linked genes (Muller's ratchet) and enhances the effects of hitchhiking (Charlesworth and Charlesworth 2000; Bachtrog 2013; Vicoso 2019). Other common features of Y chromosomes are repetitive sequences and enlarged heterochromatic regions due to reduced effectiveness of purifying selection caused by suppressed recombination and a small effective population size (Skaletsky et al. 2003). In some cases, a mechanism evolves to compensate for the haploid dosage of X-linked genes in males, but this is not always the case (Mank 2013; Gu and Walters 2017).

Many features of Y chromosomes are thought to emerge shortly after an autosome becomes Y-linked. For example, recombination suppression has been considered to come after the emergence of a new sex-determining locus on a Y chromosome to favor the co-inheritance of the sex-determining locus and male-beneficial/female-detrimental sexually antagonistic alleles (Orzack et al. 1980; van Doorn and Kirkpatrick 2007; Roberts et al. 2009; van Doorn and Kirkpatrick 2010). As additional sexually antagonistic alleles accumulate on a Y chromosome, this is predicted to trigger progressive spread of the nonrecombining region along the Y chromosome (Rice 1987; van Doorn and Kirkpatrick 2007). Although these features have been characterized in sex chromosomes of various ages and degeneration levels (Bachtrog 2013; Zhou et al. 2014), the very first stages of Y chromosome evolution are poorly understood because of a lack of extremely young sex chromosome systems. Most of the best studied young sex chromosomes have accumulated multiple types of X-Y differentiation, including suppressed recombination, Y chromosome gene loss, or X chromosome dosage compensation (Bergero et al. 2013; Mahajan et al. 2018; Darolti et al. 2019; Krasovec et al. 2019). It is therefore unclear the extent to which the early evolution of sex chromosomes is dominated by chromosome-wide X-Y divergence or gene-level differences (Kaiser et al. 2011; Zhou and Bachtrog 2012). This study addresses that shortcoming by determining how a young "proto-Y" chromosome has differentiated from its homologous proto-X chromosome shorty after its emergence.

We are especially interested in how gene expression differences accumulate between the proto-Y and proto-X chromosomes. As the proto-Y and proto-X chromosomes diverge, it is expected that alleles on the proto-Y chromosome are up- or down-regulated because of *cis*-regulatory sequence differences that contribute to proto-Y gene expression. These *cis*-regulatory effects may be especially important for the expression of sexually antagonistic (male-beneficial/female-deleterious) alleles and degeneration of functional genes (Rice 1984; Zhou and Bachtrog 2012). How these gene expression differences accumulate during the very earliest stages of sex chromosome evolution is not well understood.

We used the house fly, *Musca domestica*, as a model system to study the early evolution of sex chromosomes because it has very young sex chromosomes that are still segregating as polymorphisms within natural populations (Hamm et al. 2015). The *Musca domestica male determiner* (*Mdmd*) can be found on the Y chromosome (Y^M) and on at least four autosomes (Sharma et al. 2017). Each chromosome carrying *Mdmd*, including Y^M, is a recently derived proto-Y chromosome (Meisel et al. 2017). The proto-Y and proto-X chromosomes show minimal sequence and morphological divergence, as well as similar gene content (Boyes et al. 1964; Hediger et al. 1998; Meisel et al. 2017), consistent with their recent origins. However, it is not clear the extent to which the proto-Y chromosomes are masculinized or degenerated. A previous study revealed a small, but significant, effect of the proto-Y chromosomes on gene expression (Son et al. 2019). However, it could not resolve if the expression differences are the result of changes in the expression of the proto-Y copies, proto-X copies, or both.

In this study, we tested if one proto-Y chromosome, the third chromosome carrying *Mdmd* (III^M), has evidence of differentiation from its homologous proto-X chromosome by evaluating gene expression and DNA sequence differences between proto-Y genes and their proto-X counterparts. We compared normal (genotypic) males carrying a III^M proto-Y chromosome to sex-reversed males with no proto-Y chromosome using RNA-seq data generated in a previous study (Son et al. 2019) and newly generated Nanopore long read sequencing data. The genotypic males contain one copy each of the proto-Y and proto-X, while the sex-reversed males carry two copies of the proto-X chromosome and no proto-Y. Using sex-reversed males allows us to control for the effect of sexually dimorphic gene expression on the inference of divergence between the proto-Y (III^M) and proto-X. We used these data to test if the early evolution of a Y chromosome is dominated by chromosome-wide or gene-by-gene changes in expression.

Results and Discussion

DNA sequence divergence between the proto-Y and proto-X chromosomes

We used RNA-seq data to identify genetic variants (SNPs and small indels) within genes in genotypic (III^M/III) and sex-reversed (III/III) male house flies (Son et al. 2019). We found that the genotypic males have an excess of heterozygosity on the third chromosome, relative to the sex-reversed males (Figure 1; $P < 10^{-16}$ in a Wilcoxon rank sum test comparing genes on the third chromosome with genes on the other chromosomes). This is consistent with a previous comparison between females and III^M males (Meisel et al. 2017), and it suggests that the

sequences of genes on the III^M proto-Y chromosome are differentiated from the copies on the proto-X (i.e., the standard third chromosome).

We would expect that there would be a similar level of heterozygosity on the ancestral X chromosome in the genotypic males (X/X; III^M/III) and sex-reversed males (X/X; III/III) due to the presence of two copies of the X chromosome in both genotypes. However, the III^M males have elevated heterozygosity on the X chromosome (Figure 1; $P = 8.32\times10^{-13}$ in a Wilcoxon rank sum test comparing the X chromosome with chromosomes I, II, IV, and V). This was also observed in a comparison between females and III^M males(Meisel et al. 2017), and the cause of the elevated X chromosome heterozygosity in III^M males remains unresolved.

Gene expression divergence between the proto-Y and proto-X chromosomes

Elevated heterozygosity on the third chromosome in genotypic (III^M/III) males relative to sex-reversed (III/III) males suggests that the DNA sequences of the house fly III^M proto-Y chromosome is differentiated from the standard third (proto-X) chromosome even if the proto-Y and proto-X chromosomes have similar morphology and gene content (Boyes et al. 1964; Hediger et al. 1998; Meisel et al. 2017). We hypothesized that X-Y sequence differences could contribute to expression differentiation between the proto-Y and the proto-X chromosomes. To test this hypothesis, we quantified differential gene expression between the proto-X and proto-Y chromosome copies of the third chromosome using the RNA-seq data described above and new Oxford Nanopore long read sequencing data.

We quantified ASE of genes in genotypic (III^M/III) and sex-reversed (III/III) males on the third (proto-sex) chromosome (Figure 2) and the other chromosomes (Supplementary Figure 1). We measured ASE as the proportion of iterations in an MCMC simulation in which the expression of a focal haplotype is estimated as >0.5 (see Materials and Methods). The proportions of iterations with focal haplotypes >0.5 were overrepresented at five points in the histograms (0, 0.25, 0.5, 0.75, and 1) both in genotypic and sex-reversed males (Figure 2 and Supplementary Figure 1). These proportions may be overrepresented because we only sampled two genotypes for our ASE analysis, which caused us to have a non-continuous distribution of proportions.

The proportion of iterations with a focal haplotype >0.5 gives a measure of ASE ranging from 0 (extreme ASE in favor of one allele) to 1 (extreme ASE in favor of another allele), with 0.5 indicating equal expression of both alleles. We divided our measures of ASE into five bins, with each bin capturing one of the five most common proportions (Figure 2 and Supplementary Figure 1): 1) extreme ASE with a value between 0 and 0.125, 2) moderate ASE with a value between 0.125 and 0.375, 3) no ASE with a value between 0.375 and 0.625, 4) moderate ASE with a value between 0.625 and 0.875, and 5) extreme ASE with a value between 0.875 and 1. In the analysis below, we considered a gene to have ASE if it falls into one of the two bins of extreme ASE, and genes in the no ASE bin were classified as not having no allele-specific expression. Genes with moderate ASE were excluded in order to be conservative about ASE assignment.

If the III^M proto-Y chromosome is differentiated in gene expression from its homologous III proto-X chromosome because of differences in *cis*-regulatory alleles across the entire third

chromosome, then we expect a higher fraction of genes with ASE on the third chromosome in the genotypic (III^M/III) males than in the sex-reversed (III/III) males. In contrast to that expectation, we did not find an excess of genes with ASE in genotypic males compared to ASE genes in sex-reversed males on the third chromosome relative to other chromosomes (Figure 3A and Supplementary Table 1; Fisher's exact test, $P = 0.6996$). This result suggests that the III^M proto-Y chromosome is not broadly differentiated in *cis*-regulatory alleles from the standard third (proto-X) chromosome. This provides evidence that the early stages of Y chromosome evolution do not involve chromosome-wide changes in gene regulation via an excess of *cis*-regulatory mutations.

We next identified individual genes with differences in ASE between genotypic (III^M/III) and sex-reversed (III/III) males. There are 95 genes with extreme ASE in the genotypic males and no ASE in the sex-reversed males on the third chromosome (Supplementary Table 2). These genes could have ASE in III^M males because of differences in *cis* regulatory sequences between the III^M and standard third chromosome. To test whether the observed number of third chromosome genes with ASE in genotypic males and no ASE in sex-reversed males is in excess of a null expectation, we determined the number of third chromosome genes with no ASE in genotypic males and extreme ASE in sex-reversed males (i.e., the opposite of what we did above to find the first set of 95 genes). There are 76 genes with no ASE in the genotypic males and extreme ASE in the sex-reversed males on the third chromosome (Supplementary Table 2). To test if there is a significant excess of genes with ASE on the third chromosome in III^M males, we also identified 241 genes on other chromosomes with ASE in genotypic males and no ASE in sex-reversed males, as well as 281 genes on other chromosomes with no ASE in genotypic males and ASE in sex-reversed males (Figure 3B and Supplementary Table 2). There is an excess of genes (95) with ASE in genotypic males and non-ASE in sex-reversed males compared to the number of genes (76) with non-ASE in genotypic males and ASE in sex-reversed males on the third chromosome, relative to the other chromosomes (Figure 3B and Supplementary Table 2; Fisher's exact test, $P = 0.03467$). These results suggest that, while the III^M proto-Y chromosome is not broadly differentiated in *cis*-regulatory sequences from the standard third (proto-X) chromosome, there is an excess of individual genes with *cis*-regulatory differences between the III^M proto-Y and its homologous proto-X chromosome.

Up-regulation of the Y-allele and sex-biased expression

Male-beneficial/female-detrimental sexually antagonistic alleles are expected to accumulate on a Y chromosome (Rice 1984). These sexually antagonistic polymorphisms can favor the inhibition of recombination between the male-determining gene and any loci with sexually antagonistic alleles, thereby causing X-Y divergence (van Doorn and Kirkpatrick 2007; van Doorn and Kirkpatrick 2010). One way for alleles to have sexually antagonistic effects is if they are expressed at a level that is beneficial to one sex and deleterious in the opposite sex (Parsch and Ellegren 2013). Alternatively, differential expression of the Y-linked allele from the homologous allele on the X chromosome could be favored after protein-coding sequence divergence between the X and Y alleles that resulted from the fixation of male-beneficial sexually antagonistic alleles in the coding sequence of the Y-linked copy (Mank 2017). In both cases, we expect differential expression between X and Y alleles.

To test for expression divergence between X and Y copies that are likely to have sex-specific effects, we simultaneously investigated ASE and sex-biased gene expression. Specifically, we tested if genes on the third chromosome with ASE in the genotypic (III^M/III) males and no ASE in the sex-reversed (III/III) males are differentially expressed between genotypic and sex-reversed males. We previously showed that genes with "discordant sex-biased expression" (i.e., opposite sex-biased expression) between the genotypic and sex-reversed males are over-represented on the third chromosome (Son et al. 2019), suggesting divergence of *cis*-regulatory alleles between the III^M (proto-Y) and standard third (proto-X) chromosomes. However, we did not previously determine which alleles (proto-Y or proto-X copies) are responsible for the expression differences between genotypic and sex-reversed males.

Using our ASE results, we found one gene (LOC101899975, encoding XM_011293910.2 and XM_011293909.2) with discordant sex-biased gene expression out of the 95 genes on the third chromosome with ASE in the genotypic (III^M/III) males and no ASE in the sex-reversed (III/III) males. This gene is homologous to *dynein assembly factor 5, axonemal* (human gene *DNAAF5* and *Drosophila melanogaster* gene *HEATR2*). The gene, which we refer to as *Musca domestica HEATR2* (*Md-HEATR2*), is expected to encode a protein that functions in flagellated sperm motility (Diggle et al. 2014), and it has strong testis-biased expression in *D. melanogaster* (Chintapalli et al. 2007). *Md-HEATR2* has male-biased expression in genotypic males and female-biased expression in sex-reversed males (Son et al. 2019), suggesting that expression differences between the III^M proto-Y and the standard third (proto-X) chromosome cause the male-biased expression of the gene in the genotypic males.

With haplotypes estimated by IDP-ASE, we identified three diagnostic variant sites for ASE within *Md-HEATR2* (Figure 4A), which are all synonymous SNPs. The genotypic (III^M/III) males are heterozygous and the sex-reversed (III/III) males are homozygous at all diagnostic sites. We inferred the allele on the standard third chromosome as the one in common between genotypic and sex-reversed males, and the III^M allele as the one unique to genotypic males at each diagnostic variant site. *Md-HEATR2* is expressed higher in III^M genotypic males than in sex-reversed males (Figure 4A). In the III^M genotypic males, the III^M (Y-linked) alleles are expressed higher than the X-linked alleles, indicating that the Y-linked alleles are associated with the up-regulation of the gene in III^M genotypic males relative to sex-reversed males (Figure 4A). The copy of *Md-HEATR2* on the III^M proto-Y chromosome is therefore up-regulated relative to the proto-X copy.

The evolutionary divergence of the proto-X and proto-Y copies of *Md-HEATR2* could constitute an early stage of X-Y differentiation before chromosome-wide X-Y differentiation occurs (Bachtrog 2013). Young Y chromosomes have very similar gene content as their ancestral autosomes, in contrast to old Y chromosomes that have only retained genes with male-specific functions or recruited genes associated with testis expression from other autosomes (Koerich et al. 2008; Kaiser et al. 2011; Mahajan and Bachtrog 2017). Our results suggest that changes in the expression of individual Y-linked genes that were retained from the ancestral autosome could have important phenotypic effects during early Y chromosome evolution, as opposed to chromosome-scale divergence of the proto-Y chromosome from its homologous proto-X chromosome (Zhou and Bachtrog 2012).

Using Nanopore long read sequencing data, we examined 1,273 base pairs upstream of *Md-HEATR2* to identify diagnostic sites that could be responsible for regulating the expression differences between the proto-X and proto-Y alleles. We chose that distance because it includes the first variable site we could identify on the scaffold containing *Md-HEATR2* in our Nanopore data (i.e., including a larger region would not provide any additional information). We found twelve variable sites with different variants (SNPs and small indels) between genotypic (III^M/III) and sex-reversed (III/III) males (Figure 4B). We next examined whether these sites are located within a potential transcription factor (TF) binding region. We found five TF binding regions predicted upstream of *Md-HEATR2* using the 'Tfsitescan' tool in the 'object-oriented Transcription Factors Database (ooTFD)' (Ghosh 1999). However, none of the twelve variable sites are found within any predicted TF binding regions (Figure 4C). Further work is needed to determine how the differential expression of the proto-X and proto-Y copies of *Md-HEATR2* is regulated.

If alleles have sexually antagonistic effects (e.g., beneficial to males and deleterious to females), then selection on these alleles could drive sex chromosome turnover if they are tightly linked to a new sex-determining gene(Orzack et al. 1980; van Doorn and Kirkpatrick 2007; Roberts et al. 2009; van Doorn and Kirkpatrick 2010). The genetic linkage between sexually antagonistic alleles and the new sex-determining locus could favor restricted or suppressed recombination between the proto-Y and proto-X chromosomes in that linked region, triggering additional X-Y differentiation (Bachtrog 2013). The expression of *Md-HEATR2* could be under sexually antagonistic selection because it functions in flagellated sperm motility (Diggle et al. 2014). *Md-HEATR2* has male-biased expression in abdominal tissue (Son et al. 2019), consistent with a function in spermatogenesis. Axonemal dynein is important for male fertility by affecting sperm motility in *Drosophila* (Kurek et al. 1998; Carvalho et al. 2000), suggesting that it may be beneficial to male fitness to have higher expression of *Md-HEATR2*. In addition, investment in expressing the gene in females could be costly, possibly because of the gene's other functions in mechanosensory neurons (Diggle et al. 2014). Therefore, the up-regulation of *Md-HEATR2* in XY males due to high expression of the proto-Y copy could be consistent with sexually antagonism playing an important role in the early stage of X-Y differentiation at individual genes in house fly.

Conclusions

We investigated gene sequence and expression differences between the III^M proto-Y and its homologous proto-X chromosome to determine how a very young Y chromosome has been differentiated from its homologous X chromosome shorty after it was formed. To those ends, we used genotypic (III^M/III) and sex-reversed (III/III) males because they are phenotypically almost the same but differ in the proto-sex chromosomes they carry (Hediger et al. 2010; Son et al. 2019). We found increased heterozygosity on the III^M proto-Y chromosome in genotypic males relative to sex-reversed males, consistent with sequence divergence between the proto-Y and proto-X (Figure 1). There is not an excess of genes with ASE on the proto-sex chromosome in genotypic males compared to genes with ASE on the same chromosome in sex-reversed males (Figure 3A and Supplementary Table 1). In contrast, we found an excess of individual genes with ASE in the genotypic males and no ASE in the sex-reversed males on the proto-sex chromosome relative to the other chromosomes (Figure 3B and Supplementary Table 2).

However, the number of genes with ASE on the third chromosome only in genotypic males, and not in sex-reverse males, is small (<100, which is likely less than 5% of the entire chromosome). We identified one gene on the third chromosome (*Md-HEATR2*) with ASE in genotypic males, no ASE in sex-reversed males, and discordant sex-biased expression between genotypic and sex-reversed males (Figure 4). We hypothesize that expression divergence of *Md-HEATR2* could be an example of very early X-Y differentiation of individual genes that results from sexually antagonistic selection. Therefore, the house fly III^M proto-Y chromosome is differentiated in gene sequence and expression from its homologous proto-X chromosome at individual genes, but not chromosome wide. This suggests that the earliest stages of Y chromosome evolution consist of gene-by-gene, rather than chromosome-scale, changes in gene expression.

Materials and Methods

Fly strain

We analyzed RNA-seq data and performed Oxford Nanopore sequencing on a house fly strain that allows for identification of genotypic (III^M/III) males and sex-reversed (III/III) males (Hediger et al. 2010). This is because the standard third chromosome (III) in this strain has the recessive mutations *pointed wing* (*pw*) and *brown body* (*bwb*). Sex-reversed males (and normal females) have both mutant phenotypes, whereas genotypic males are wild-type for both phenotypes because the III^M chromosome has the dominant wild-type alleles. The RNA-seq data that we analyzed (available at NCBI GEO accession GSE126689) comes from a previous study that used double-stranded RNA (dsRNA) targeting *Md-tra* to create sex-reversed phenotypic males that have a female genotype without a proto-Y chromosome (Son et al. 2019). We compared the sex-reversed males to genotypic males that received a sham treatment of dsRNA targeting GFP. This comparison allows us to investigate genes that have different sex-biased expression in genotypic and sex-reversed males (Son et al. 2019). We used genotypic males and sex-reversed males from the same strain that were subjected to the same treatment for genome sequencing using the Oxford Nanopore long read technology.

Oxford Nanopore sequencing

We performed Oxford Nanopore sequencing of one genotypic (III^M/III) male and one sex-reversed (III/III) male created in the same strain and using the same *Md-tra* dsRNA treatment as a previous RNA-seq study (Son et al. 2019). DNA was isolated with a phenol/chloroform protocol. A single genotypic male and a single sex-reversed male with detached wings were each transferred to a 1.5 mL Eppendorf tube with 0.5mL homogenization buffer (4.1 g sucrose, 15 mL 1M Tris-HCl pH 8.0, 0.5M EDTA, 100 mL dH_2O) and then homogenized using pestles set into a tissue grinder homogenizer. To each tube we added 40 uL of 10% SDS and 2.5 uL of 10mg/mL Proteinase K, and then we incubated the tube at 65°C for 30 min. We next added 2 uL of 4 mg/mL RNase to each tube and incubated at 37°C for 15 min. We added 48 uL of 5M KAc to each tube and placed on ice for 30 min. Then we centrifugated the tubes at 14000 rpm for 10 min at 4°C, and the supernatant was transferred into a new tube using a wide-bore pipette tip (all subsequent steps of transfer and mixing during DNA extraction were also done with wide-bore pipette tips to prevent DNA shearing). We added 250 uL phenol and 250 uL chloroform to the

extracted supernatant in the new tube, mixed briefly, spun at 14000 rpm for 15 min at 4°C, and then transferred the supernatant into a new tube. We next added 500 uL chloroform to the supernatant in the new tube, mixed well, and spun a 14000 rpm for 5 min at 4°C. We then transferred the supernatant into a new tube. We added 40 uL of 3M NaAc and 800 uL of 95% ethanol to the supernatant in the new tube, mixed briefly, spun at 14000 rpm for 15 min at 4°C, and then carefully poured off all supernatant. We next added 800 uL of 70% ethanol to the remaining pellet, mixed briefly to wash the pellet, spun at 14000 rpm for 15 min at 4°C, removed the supernatant, and then resuspended the pellet in 30 uL of nuclease-free water.

Oxford Nanopore Sequencing libraries were prepared with the 1D genomic DNA Ligation kit (SQK-LSK109, Oxford Nanopore), following the manufacturer's protocol. DNA from the genotypic male and sex-reversed male (see above) was used to create a separate sequencing library for each genotype. Following the manufacturer's protocol, 15 uL of each library, along with sequencing buffer and loading beads (totaling 75 uL), were loaded onto a R9.4 flow cell until no pores were available on a MinION sequencer (Oxford Nanopore). The libraries from the genotypic and sex-reversed males were run on separate flow cells. Base calling was performed using the Guppy pipeline software version 3.1.5 (Oxford Nanopore) with parameters (--calib_detect --qscore_filtering --min_qscore 10). The base called reads were aligned to the house fly genome assembly v2.0.2 (Scott et al. 2014) using Minimap2 version 2.17 with the "-ax map-ont" parameter (Li 2018).

Variant calling

We used available RNA-seq data (Son et al. 2019) to identify genetic variants (SNPs and small insertions/deletions) that differentiate the III^M proto-Y chromosome from the standard third (proto-X) chromosome, and then we tested if III^M males have elevated heterozygosity on the third chromosome as compared to sex-reversed males (Meisel et al. 2017). We used the Genome Analysis Toolkit (GATK) pipeline for calling variants in the RNA-seq data from the *Md-tra* RNAi experiment in (Son et al. 2019), following the best practices for SNP and insertion/deletion (indel) calling on RNA-seq data (McKenna et al. 2010; Meisel et al. 2017). We used STAR (Dobin et al. 2013) to align reads from three genotypic (III^M/III) male libraries and three sex-reversed (III/III) male libraries to the reference assembly v2.0.2 (Scott et al. 2014). The aligned reads were used to generate a new reference genome index from the detected splice junctions in the first alignment run, and then a second alignment was performed with the new reference. We next marked duplicate reads from the same RNA molecule and used the GATK tool 'SplitNCigarReads' to reassign mapping qualities to 60 with the 'ReassignOneMappingQuality' read filter for alignments with a mapping quality of 255. Indels were detected and realigned with 'RealignerTargetCreator' and 'IndelRealigner'. The realigned reads were used for base recalibration with 'BaseRecalibrator' and 'PrintReads'. The base recalibration was performed in three sequential iterations in which recalibrated and filtered reads were used to train the next round of base recalibration, at which point there were no beneficial effects of additional base recalibration as verified by 'AnalyzeCovariates'. We next used the recalibrated reads from all three replicates of genotypic and sex-reversed males to call variants using 'HaplotypeCaller' with emission and calling confidence thresholds of 20. We applied 'genotypeGVCFs' to the variant calls from the two types of males for joint genotyping, and then we filtered the variants using 'VariantFiltration' with a cluster window size of 35 bp, cluster size

of 3 SNPs, FS > 20, and QD < 2. The final variant calls were used to identify heterozygous variants within genes using the coordinates from the genome sequencing project, annotation release 102 (Scott et al. 2014). We measured relative heterozygosity within each gene in genotypic (III^M/III) and sex-reversed (III/III) males as the number of heterozygous variants in genotypic males for a given gene (h_G) divided by the total number heterozygous variants in both genotypic and sex-reversed males (h_{SR}), times one hundred: $100h_G/(h_G + h_{SR})$.

For the variant calling from Nanopore long reads, the base called reads were indexed using fast5 files with the 'index' module of Nanopolish version 0.11.1 (Quick et al. 2016), and they were aligned with Minimap2 version 2.17 (Li 2018) to house fly genome assembly v2.0.2 (Scott et al. 2014). The aligned and raw reads were used to call variants using the "variants" module of Nanopolish version 0.11.1 with the "--ploidy 2" parameter (Quick et al. 2016). We used a python script 'nanopolish_makerange.py' provided in the package to split the genome into 50 kb segments because it was recommended to use the script for large datasets with genome size more than 50 kb.

Allele-specific expression

Diploid species can have two alleles at a locus, one of which was inherited maternally and the other paternally. The maternal and paternal alleles can be expressed unequally in the diploid, which is called allele-specific expression (ASE). We investigated if there is elevated ASE on the third chromosome in males carrying one III^M proto-Y and one proto-X chromosome compared to sex-reversed males with two proto-X chromosomes. To do this, we implemented the IDP-ASE tool at the gene level with house fly genome annotation release 102 (Scott et al. 2014), following the developers' recommended analysis steps (Deonovic et al. 2016). We first prepared information on the number and locations of variants within each gene (SNPs and small indels), as well as read counts at each variant location (see above). The IDP-ASE software was supplied with raw reads and aligned reads created by RNA-seq (Son et al. 2019) and Nanopore sequencing, and variant calls created only by GATK because Nanopore sequencing reads of each library (genotypic male and sex-reversed male) had less than 10× coverage across the house fly genome (i.e., too low for reliable variant calling).

The prepared data from each gene was next run in an MCMC (Markov chain Monte Carlo) sampling simulation to estimate the haplotype within each gene with a Metropolis-Hastings sampler (Bansal et al. 2008). Next, the software estimates the proportion of each estimated haplotype that contributes to the total expression of the gene (ρ) from each iteration using slice sampling (Neal and others 2003). A value of ρ=0.5 indicates equal expression between two alleles, whereas ρ<0.5 or ρ>0.5 indicates ASE. The MCMC sampling was run with a 1000 iteration burn-in followed by at least 500 iterations where data were recorded. The actual number of iterations was automatically adjusted by the software during the simulation to produce the best simulation output for quantifying ASE within a gene. The IDP-ASE simulation generated a distribution of ρ for each gene across all post-burn-in iterations, and then it calculated the proportion of iterations with ρ > 0.5. This proportion was used to estimate the extent of ASE for each gene. For example, if all iterations for a gene have ρ > 0.5, then the proportion is 1 and the gene has strong evidence for ASE of one allele. Similarly, if all iterations for a gene have ρ < 0.5, then the proportion is 0 and the gene has strong evidence for ASE of the other allele. In

contrast, if half of the iterations have $\rho > 0.5$ and the other half have $\rho < 0.5$, then the proportion is 0.5 and there is not any evidence for ASE. To classify whether genes have ASE or not, we considered only genes with total RNA-seq read counts above 10.

We used the output of IDP-ASE to compare expression of the III^M (proto-Y) and III (proto-X) alleles in genotypic males. IDP-ASE only quantifies ASE within bi-allelic loci, so we only included genes with heterozygous sites within transcripts in genotypic (III^M/III) or sex-reversed (III/III) males. In addition, we removed heterozygous variants with the same genotype in genotypic and sex-reversed males because they do not allow us discriminate between the proto-Y and proto-X alleles. Removing these variants may have also sped up the simulation times, but this was not rigorously investigated. To discriminate between the III^M and III alleles, we used haplotypes estimated during IDP-ASE runs and genotypes inferred from GATK for genotypic (III^M/III) and sex-reversed (III/III) males. For example, using genotypes called using GATK from the RNA-seq data, we first identified sites with heterozygous alleles in genotypic males and homozygous alleles in sex-reversed males. Next, we inferred the allele in common between genotypic and sex-reversed as the III allele, and the other allele that is unique to genotypic males as the III^M allele. Lastly, we matched those sites to the haplotypes estimated by IDP-ASE to quantify ASE within each genotype.

Acknowledgements and funding information

This work was supported by Grant-in-Aid of Research from the National Academy of Sciences, administered by Sigma Xi, The Scientific Research Society (grant number G2018100198487895 to J.H.S. and R.P.M.) and by the National Science Foundation (grant numbers OISE-1444220 and DEB-1845686 to R.P.M.). The Oxford Nanopore long reads data used in the study are available from the National Center for Biotechnology Information Sequence Read Archive under BioProject accession PRJNA620357 (BioSample accessions SAMN14518459 for sex-reversed males and SAMN14518460 for genotypic males).

Figure Legends

Figure 1. Elevated heterozygosity on the third and X chromosomes in genotypic (III^M/III) males relative to sex-reversed (III/III) males. The boxplots show the distributions of the percentages of heterozygous variants within genes on each chromosome in the genotypic males relative to the sex-reversed males. Values more than 50% indicate the increased heterozygosity in genotypic (III^M/III) males, and less than 50% is increased heterozygosity in sex-reversed males. The median across all autosomes is represented by a dashed line.

Figure 2. Histograms of ASE for third chromosome genes in (A) genotypic (III^M/III) and (B) sex-reversed (III/III) males. If a gene is expressed equally between the X and Y alleles, the proportion of focal haplotypes is 0.5; if a gene has ASE, the proportion is greater or less than 0.5.

Figure 3. (A) Proportions of genes with ASE in genotypic (III^M) and sex-reversed (SR) males on each chromosome. (B) Proportions of genes with ASE in genotypic males and non-ASE in sex-reversed males on the third chromosome and all other chromosomes (left two bars). Proportions of genes with non-ASE in genotypic and ASE in sex-reversed males on the third chromosome and all other chromosomes (right two bars).

Figure 4. (A) Diagnostic variable sites for allele-specific expression (ASE) of the *Md-HEATR2* gene based on haplotypes estimated in IDP-ASE; fragments per million mapped reads (FPM). (B) Variable sites that differ between genotypic (III^M) males and sex-reversed males across 1,273 base pairs upstream of *Md-HEATR2* using Oxford Nanopore long reads only. The variable sites are marked as triangles with their coordinate as positions in scaffold. (C) Transcription factor (TF) binding motifs predicted within 1,273 base pairs upstream of *Md-HEATR2*.

References

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

Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, Hahn MW, Kitano J, Mayrose I, Ming R, et al. 2014. Sex determination: why so many ways of doing it? PLoS Biol. 12(7):e1001899.

Bansal V, Halpern AL, Axelrod N, Bafna V. 2008. An MCMC algorithm for haplotype assembly from whole-genome sequence data. Genome Res. 18(8):1336–1346.

Bergero R, Qiu S, Forrest A, Borthwick H, Charlesworth D. 2013. Expansion of the pseudo-autosomal region and ongoing recombination suppression in the Silene latifolia sex chromosomes. Genetics. doi:10.1534/genetics.113.150755.

Beukeboom LW, Perrin N. 2014. The evolution of sex determination. Oxford University Press, USA.

Boyes JW, Corey MJ, Paterson HE. 1964. Somatic chromosomes of higher diptera: IX. Karyotypes of some muscid species. Can J Zool. 42(6):1025–1036.

Bull JJ. 1983. Evolution of sex determining mechanisms. The Benjamin/Cummings Publishing Company, Inc.

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(24):13239–13244.

Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. Philos Trans R Soc London Ser B Biol Sci. 355(1403):1563–1572.

Charlesworth D. 2018. Does sexual dimorphism in plants promote sex chromosome evolution? Environ Exp Bot. doi:10.1016/j.envexpbot.2017.11.005.

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

Chintapalli VR, Wang J, Dow JAT. 2007. Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat Genet. doi:10.1038/ng2049.

Darolti I, Wright AE, Sandkam BA, Morris J, Bloch NI, Farré M, Fuller RC, Bourne GR, Larkin DM, Breden F, et al. 2019. Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. Proc Natl Acad Sci U S A. doi:10.1073/pnas.1905298116.

Deonovic B, Wang Y, Weirather J, Wang X-J, Au KF. 2016. IDP-ASE: haplotyping and

quantifying allele-specific expression at the gene and gene isoform level by hybrid sequencing. Nucleic Acids Res. 45(5):e32--e32.

Diggle CP, Moore DJ, Mali G, zur Lage P, Ait-Lounis A, Schmidts M, Shoemark A, Garcia Munoz A, Halachev MR, Gautier P, et al. 2014. HEATR2 Plays a Conserved Role in Assembly of the Ciliary Motile Apparatus. PLoS Genet. doi:10.1371/journal.pgen.1004577.

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(1):15–21.

van Doorn GS, Kirkpatrick M. 2007. Turnover of sex chromosomes induced by sexual conflict. Nature. 449(7164):909.

van Doorn GS, Kirkpatrick M. 2010. Transitions between male and female heterogamety caused by sex-antagonistic selection. Genetics. 186(2):629–645.

Ghosh D. 1999. Object oriented transcription factors database (ooTFD). Nucleic Acids Res. doi:10.1093/nar/27.1.315.

Gu L, Walters JR. 2017. Evolution of sex chromosome dosage compensation in animals: A beautiful theory, undermined by facts and bedeviled by details. Genome Biol Evol. doi:10.1093/gbe/evx154.

Hamm RL, Meisel RP, Scott JG. 2015. The evolving puzzle of autosomal versus Y-linked male determination in Musca domestica. G3 Genes, Genomes, Genet. 5(3):371–384.

Hediger M, Henggeler C, Meier N, Perez R, Saccone G, Bopp D. 2010. Molecular characterization of the key switch F provides a basis for understanding the rapid divergence of the sex-determining pathway in the housefly. Genetics. 184(1):155–170.

Hediger M, Minet AD, Niessen M, Schmidt R, Hilfiker-Kleiner D, Çakir Ş, Nöthiger R, Dübendorfer A. 1998. The male-determining activity on the Y chromosome of the housefly (Musca domestica L.) Consists of separable elements. Genetics.

Kaiser VB, Zhou Q, Bachtrog D. 2011. Nonrandom gene loss from the drosophila miranda neo-Y chromosome. Genome Biol Evol. doi:10.1093/gbe/evr103.

Koerich LB, Wang X, Clark AG, Carvalho AB. 2008. Low conservation of gene content in the Drosophila Y chromosome. Nature. doi:10.1038/nature07463.

Krasovec M, Kazama Y, Ishii K, Abe T, Filatov DA. 2019. Immediate Dosage Compensation Is Triggered by the Deletion of Y-Linked Genes in Silene latifolia. Curr Biol. doi:10.1016/j.cub.2019.05.060.

Kurek R, Reugels AM, Glaätzer KH, Bünemann H. 1998. The Y chromosomal fertility factor Threads in Drosophila hydei harbors a functional gene encoding an axonemal dynein $β$ heavy

chain protein. Genetics. 149(3):1363–1376.

Li H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics. doi:10.1093/bioinformatics/bty191.

Mahajan S, Bachtrog D. 2017. Convergent evolution of y chromosome gene content in flies. Nat Commun. doi:10.1038/s41467-017-00653-x.

Mahajan S, Wei KHC, Nalley MJ, Gibilisco L, Bachtrog D. 2018. De novo assembly of a young Drosophila Y chromosome using single-molecule sequencing and chromatin conformation capture. PLoS Biol. doi:10.1371/journal.pbio.2006348.

Mank JE. 2013. Sex chromosome dosage compensation: Definitely not for everyone. Trends Genet. doi:10.1016/j.tig.2013.07.005.

Mank JE. 2017. The transcriptional architecture of phenotypic dimorphism. Nat Ecol Evol. doi:10.1038/s41559-016-0006.

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(9):1297–1303.

Meisel RP, Gonzales CA, Luu H. 2017. The house fly Y Chromosome is young and minimally differentiated from its ancient X Chromosome partner. Genome Res. 27(8):1417–1426.
Neal RM, others. 2003. Slice sampling. Ann Stat. 31(3):705–767.

Orzack SH, Sohn JJ, Kallman KD, Levin SA, Johnston R. 1980. Maintenance of the three sex chromosome polymorphism in the platyfish, Xiphophorus maculatus. Evolution (N Y). 34(4):663–672.

Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. Nat Rev Genet. doi:10.1038/nrg3376.

Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA, Koundouno R, Dudas G, Mikhail A, et al. 2016. Real-time, portable genome sequencing for Ebola surveillance. Nature. doi:10.1038/nature16996.

Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. Evolution (N Y). 38(4):735–742.

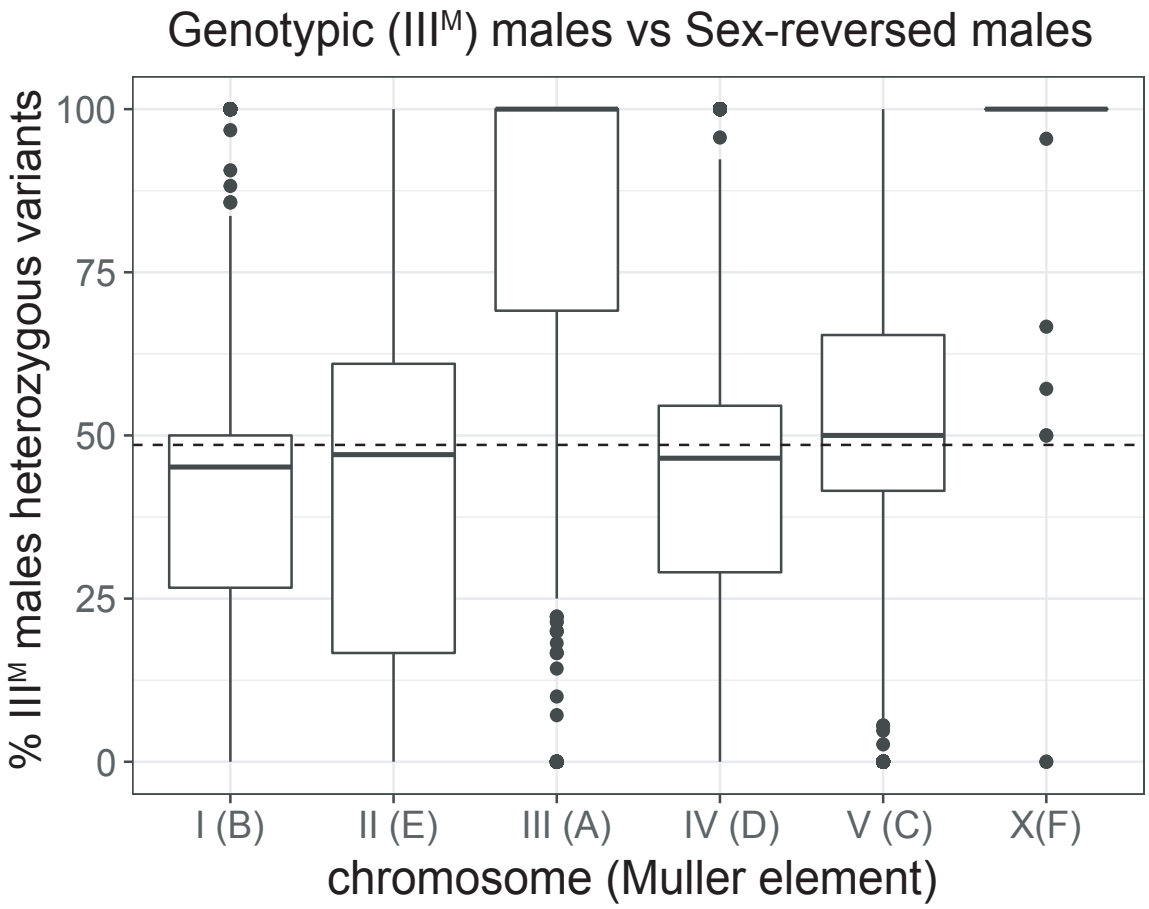
Rice WR. 1987. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. Evolution (N Y). 41(4):911–914.

Rice WR. 1996. Evolution of the Y sex chromosome in animals. Bioscience. 46(5):331–343.
Roberts RB, Ser JR, Kocher TD. 2009. Sexual conflict resolved by invasion of a novel sex

643 determiner in Lake Malawi cichlid fishes. Science (80-). 326(5955):998–1001.
644
645 Scott JG, Warren WC, Beukeboom LW, Bopp D, Clark AG, Giers SD, Hediger M, Jones AK,
646 Kasai S, Leichter CA, et al. 2014. Genome of the house fly, Musca domestica L., a global vector
647 of diseases with adaptations to a septic environment. Genome Biol. 15(10):466.
648
649 Sharma A, Heinze SD, Wu Y, Kohlbrenner T, Morilla I, Brunner C, Wimmer EA, van de Zande
650 L, Robinson MD, Beukeboom LW, et al. 2017. Male sex in houseflies is determined by Mdmd, a
651 paralog of the generic splice factor gene CWC22. Science (80-). 356(6338):642–645.
652
653 Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S,
654 Pyntikova T, Ali J, Bieri T, et al. 2003. The male-specific region of the human Y chromosome is
655 a mosaic of discrete sequence classes. Nature. 423(6942):825.
656
657 Son JH, Kohlbrenner T, Heinze S, Beukeboom LW, Bopp D, Meisel RP. 2019. Minimal effects
658 of proto-Y chromosomes on house fly gene expression in spite of evidence that selection
659 maintains stable polygenic sex determination. Genetics. doi:10.1534/genetics.119.302441.
660
661 Vicoso B. 2019. Molecular and evolutionary dynamics of animal sex-chromosome turnover. Nat
662 Ecol Evol. doi:10.1038/s41559-019-1050-8.
663
664 Zhou Q, Bachtrog D. 2012. Chromosome-wide gene silencing initiates y degeneration in
665 drosophila. Curr Biol. doi:10.1016/j.cub.2012.01.057.
666
667 Zhou Q, Zhang J, Bachtrog D, An N, Huang Q, Jarvis ED, Gilbert MTP, Zhang G. 2014.
668 Complex evolutionary trajectories of sex chromosomes across bird taxa. Science (80-).
669 346(6215):1246338.
670

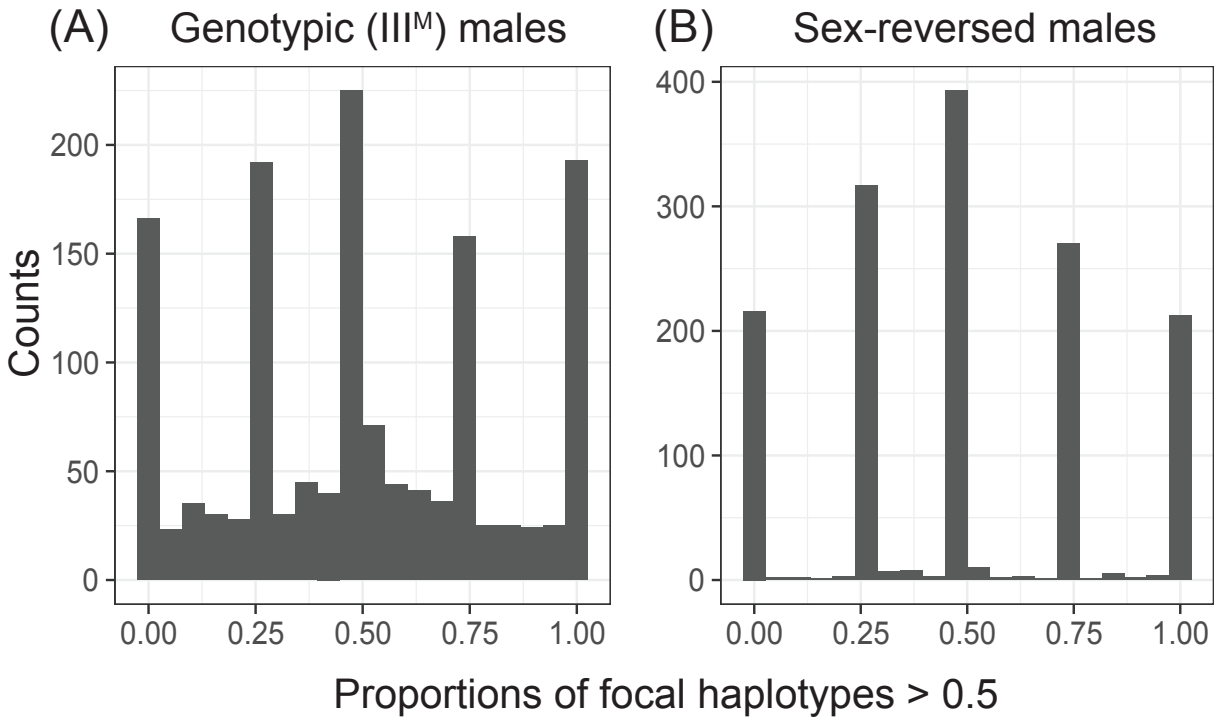
671 **Figure 1**

672



673

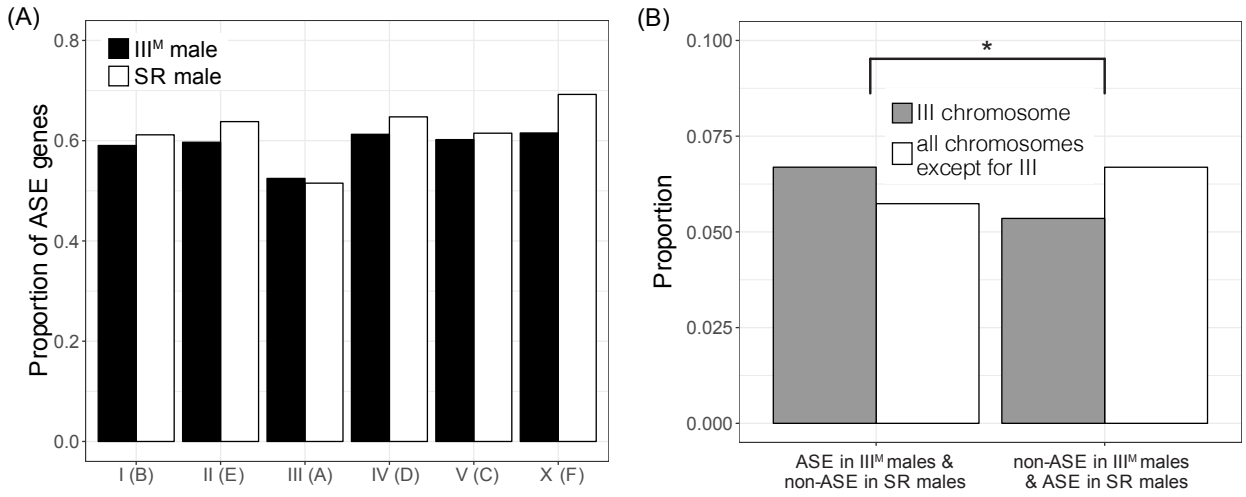
674 **Figure 2**



675

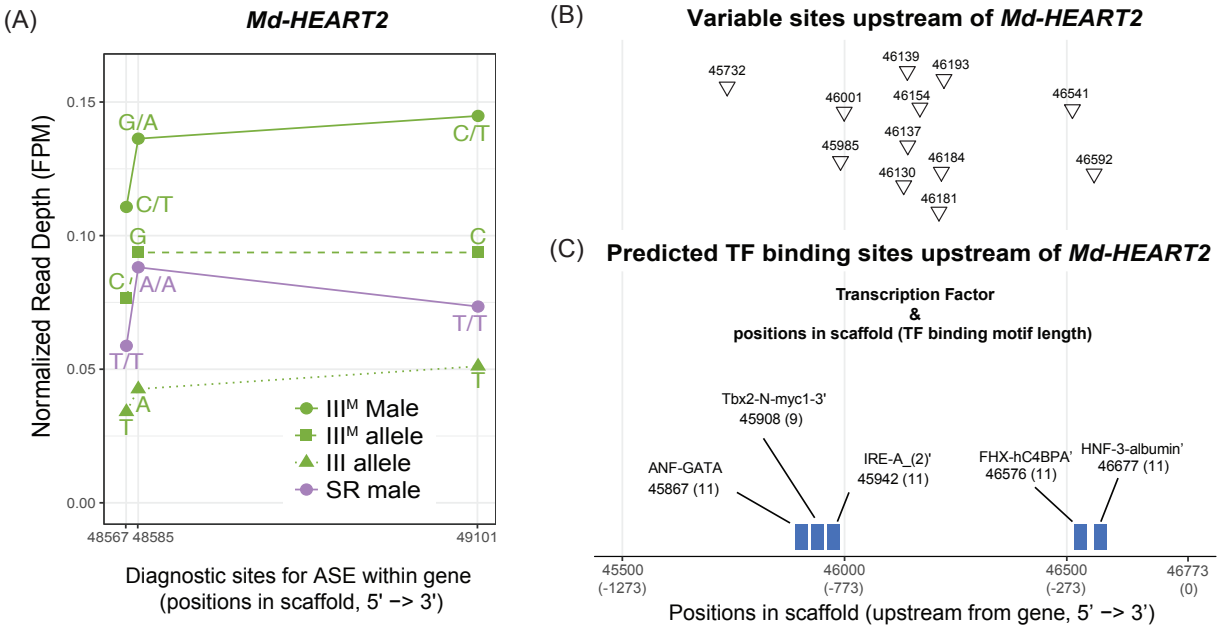
676 **Figure 3**

677



678

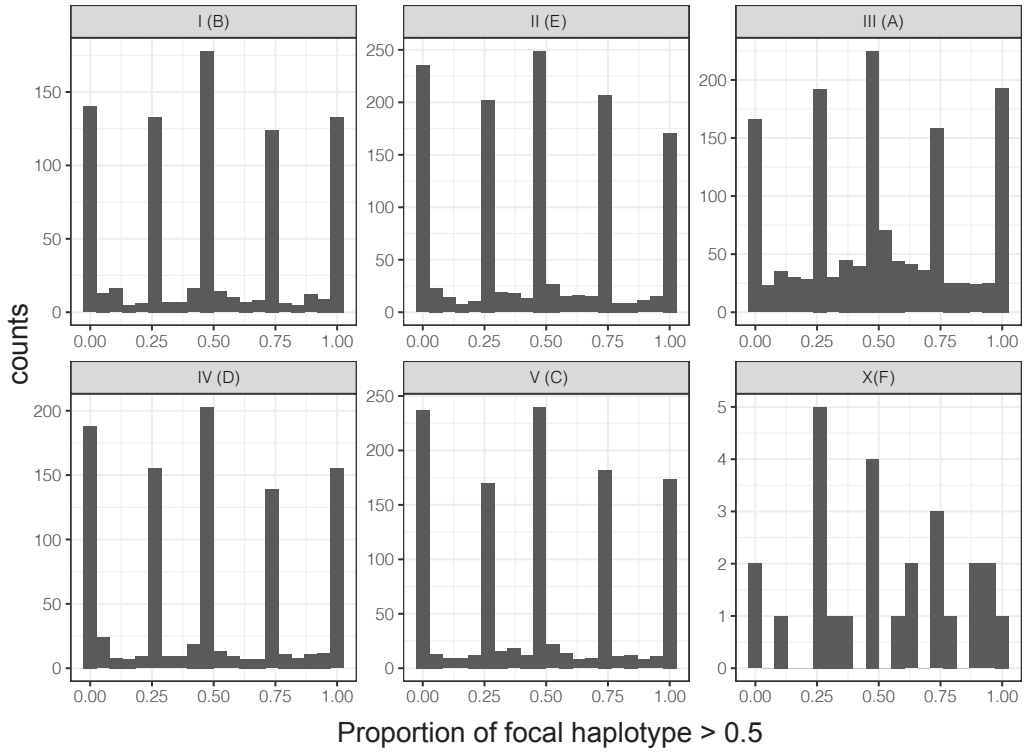
Figure 4



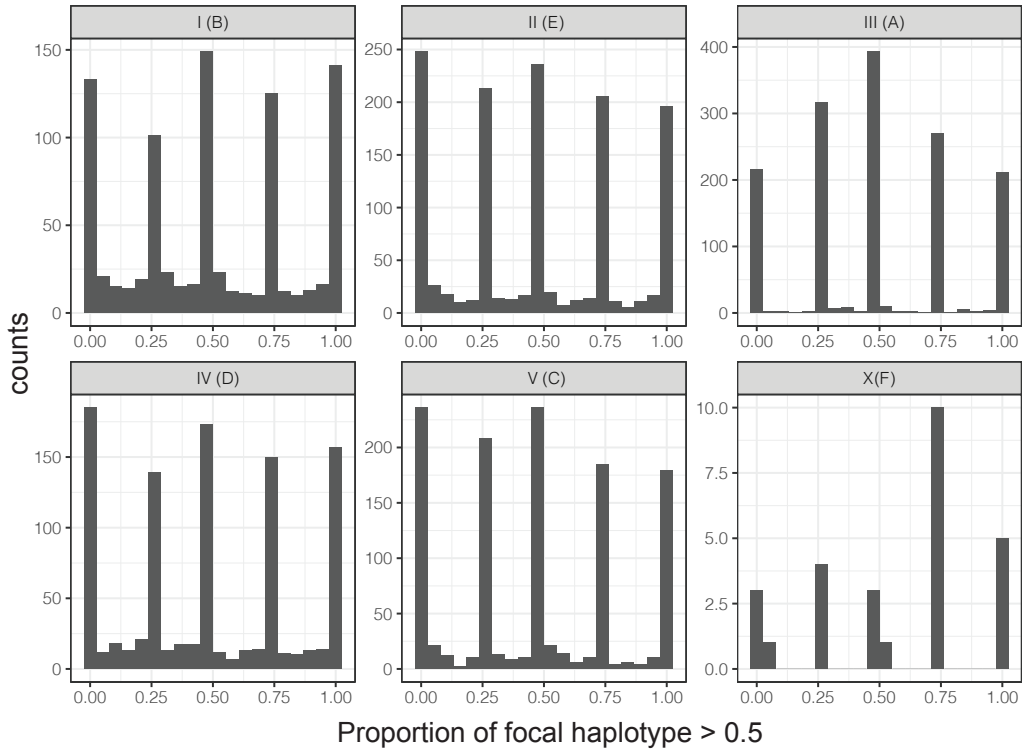
682

Supplementary Materials

(A)



(B)



683

Supplementary Figure 1. Histograms of allele-specific expression (ASE) in genes for all chromosome are shown in the genotypic (III^M/III) males and the sex-reversed (III/III) males. Muller element nomenclature (from *Drosophila*) for each chromosome is given in parentheses. If a gene is expressed equally between two alleles, the proportion of focal haplotype is 0.5; otherwise, the proportion is greater or less than 0.5.

689 Supplementary Table 1. Chromosomal distribution of allele-specific expression (ASE) and no
690 allele-specific expression (non-ASE) in genotypic (III^M) males and sex-reversed (SR) males.

Chromosome (Muller element)	# genes with ASE in III^M males	# genes with non-ASE in III^M males	# genes with ASE in SR males	# genes with non-ASE in SR males	Odds ratio	95% CI of odds ratio
III(A)	456	413	438	412	1.038574	0.8555646 - 1.2606663
genome except III(A)	**1635**	**1089**	**1711**	**1010**	**0.886281**	**0.7933539 - 0.9900421**
I(B)	320	222	334	212	0.915002	0.7123784 - 1.1750187
II(E)	465	314	513	291	0.840134	0.6821523 - 1.0344483
IV(D)	394	249	395	215	0.861373	0.6798614 - 1.0908786
V(C)	448	296	460	288	0.947652	0.7654588 - 1.1730543
X(F)	8	8	9	4	0.720539	0.102255 - 4.754933

691

692 Supplementary Table 2. Counts of genes with ASE on each chromosome in genotypic (III^M)
693 males and sex-reversed (SR) males. The total number of genes (# genes) in each chromosome
694 group (second column), # genes with ASE in genotypic males and no ASE in sex-reversed males
695 (third column), and # genes with no ASE in genotypic males and ASE in sex-reversed males
696 (fourth column) are shown. Bold indicates statistical significance ($P < 0.05$)

		ASE in III^M males and non-ASE in SR males	non-ASE in III^M males and ASE in SR males	Fisher's Exact Test	
Chromosome (Muller element)	# genes	# genes	# genes	Odds ratio compared with III(A)	95% CI
III(A)	1420	95	76		
genome except III(A)	4201	**241**	**281**	**1.45665**	**1.014895 - 2.095777**
I(B)	824	51	59	1.444139	0.8688985 - 2.4078676
II(E)	1236	**68**	**86**	**1.578592**	**0.9961489 - 2.5100267**
IV(D)	966	**55**	**76**	**1.724145**	**1.063163 - 2.809137**
V(C)	1149	67	77	1.434875	0.8982881 - 2.2981210
X(F)	26	0	0	0	0 - Infinity

697

Supplementary Table 3. Counts of genes with allele-specific expression (ASE) based on division of ASE measurements into five sections, following the rules described in the Methods. ASE proportions are sorted in the order of extreme (1st and 5th), moderate (2nd and 4th), and no (3rd) ASE. Only extreme ASE was used in the comparisons with no ASE.

Chr (ME)	Sections for ASE proportions	# of genes in genotypic (III^M/III) Males			# of genes in sex-reversed (III/III) Males		
I (B)	1st (extreme ASE)	167	320	627	167	334	667
	5th (extreme ASE)	153			167		
	2nd (moderate ASE)	158	307		167	333	
	4th (moderate ASE)	149			166		
	3rd (no ASE)	222	222	222	212	212	212
II (E)	1st (extreme ASE)	270	465	972	291	513	1013
	5th (extreme ASE)	195			222		
	2nd (moderate ASE)	251	507		258	500	
	4th (moderate ASE)	256			242		
	3rd (no ASE)	314	314	314	291	291	291
III (A)	1st (extreme ASE)	215	456	1043	220	438	1050
	5th (extreme ASE)	241			218		
	2nd (moderate ASE)	312	587		333	612	
	4th (moderate ASE)	275			279		
	3rd (no ASE)	413	413	413	412	412	412
IV (D)	1st (extreme ASE)	217	394	754	214	395	794
	5th (extreme ASE)	177			181		
	2nd (moderate ASE)	189	360		201	399	
	4th (moderate ASE)	171			198		
	3rd (no ASE)	249	249	249	215	215	215
V (C)	1st (extreme ASE)	258	448	891	268	460	908
	5th (extreme ASE)	190			192		
	2nd (moderate ASE)	220	443		239	448	
	4th (moderate ASE)	223			209		
	3rd (no ASE)	296	296	296	288	288	288
X(F)	1st (extreme ASE)	3	8	21	4	9	23
	5th (extreme ASE)	5			5		
	2nd (moderate ASE)	7	13		4	14	
	4th (moderate ASE)	6			10		
	3rd (no ASE)	5	5	5	4	4	4

Supplementary Data 1. A VCF file called with the RNA-seq reads

Supplementary Data 2. A VCF file called with the Oxford Nanopore reads

Supplementary Data 3. IDP-ASE (allele-specific expression) output for the genotypic males

Supplementary Data 4. IDP-ASE (allele-specific expression) output for the sex-reversed males