Transcriptome Differences between Alternative Sex Determining Genotypes in the House

Fly, *Musca domestica*

Richard P. Meisel[1]*, Jeffrey G. Scott[2], Andrew G. Clark[3]

1. Department of Biology and Biochemistry, University of Houston

2. Department of Entomology, Cornell University

3. Department of Molecular Biology and Genetics, Cornell University

*Author for correspondence: Richard P. Meisel, Department of Biology and

Biochemistry, University of Houston, Houston, TX, USA, 77204-5001,

phone: 713-743-3607, fax: 713-743-2636, rpmeisel@uh.edu

Abstract

Sex determination evolves rapidly, often because of turnover of the genes at the top of the pathway. The house fly, *Musca domestica*, has a multifactorial sex determination system, allowing us to identify the selective forces responsible for the evolutionary turnover of sex determination in action. There is a male determining factor, M, on the Y chromosome (Y^M), which is probably the ancestral state. An M factor on the third chromosome (III^M) has reached high frequencies in multiple populations across the world, but the evolutionary forces responsible for the invasion of III^M are not resolved. To test if the III^M chromosome invaded because of sex-specific selection pressures, we used mRNA sequencing to determine if isogenic males that differ only in the presence of the Y^M or III^M chromosome have different gene expression profiles. We find that more genes are differentially expressed between Y^M and III^M males in testis than head, and that genes with male-biased expression are most likely to be differentially expressed between Y^M and III^M males. We additionally find that III^M males have a "masculinized" gene expression profile, suggesting that the III^M chromosome has accumulated an excess of male-beneficial alleles because of its male-limited transmission. These results are consistent with the hypothesis that sex-specific selection acts on alleles linked to the male-determining locus driving evolutionary turnover in the sex determination pathway.

Key words: sex determination, gene expression, RNA-Seq, sex-specific selection, sex chromosomes

1. Introduction

Sex determination (SD) is an essential developmental process responsible for sexually dimorphic phenotypes. It is therefore paradoxical that SD pathways are poorly conserved, with master SD (MSD) genes at the top of the pathway differing between closely related species and even variable within species (Bull, 1983; Wilkins, 1995; Pomiankowski et al., 2004; Beukeboom and Perrin, 2014). The hypotheses to explain the rapid evolution of SD pathways fall into three categories. First, SD evolution may be selectively neutral if MSD turnover is the result of mutational input without phenotypic or fitness consequences (van Doorn, 2014). Second, frequency dependent (sex-ratio) selection could favor a new MSD variant if one sex is below its equilibrium frequency (Eshel, 1975; Bull and Charnov, 1977; Bulmer and Bull, 1982; Werren and Beukeboom, 1998). Third, a new MSD locus can invade a population if the new MSD variant itself or genetically linked alleles confer a fitness benefit (Charlesworth and Charlesworth, 1980; Rice, 1986; Charlesworth, 1991; van Doorn and Kirkpatrick, 2007, 2010). Those fitness effects could be beneficial to both sexes (natural selection), increase the reproductive success of the sex determined by the new MSD variant (sexual selection), or be beneficial to the sex determined by the MSD variant and deleterious to the other sex (sexually antagonistic selection). Sexually antagonistic selection is predicted to be an especially important driver of MSD turnover because linkage to an MSD locus allows the sexually antagonistic allele to be inherited in a sex-limited manner, thereby resolving the inter-sexual conflict (Charlesworth and Charlesworth, 1980; van Doorn and Kirkpatrick, 2007; Roberts et al., 2009; van Doorn and Kirkpatrick, 2010).

The house fly, *Musca domestica*, is an ideal model for testing hypotheses about the evolution of SD because it has a multifactorial SD system, with male- and female-determining loci segregating in natural populations (Dübendorfer et al., 2002; Hamm et al., 2015). Most relevant to the work presented here is the the fact that the male-determining factor, M, can be located on the Y chromosome (Y^M), any of the five autosomes (A^M), and even the X chromosome (Hamm et al., 2015). It is unknown whether these M-factors are the same gene in different locations or different genes that have independently assumed the role of an MSD locus (Bopp, 2010). Y^M is a common arrangement (Hamm et al., 2015), and it is thought

3

to be the ancestral state because it is the genotype found in close relatives of the house fly (Boyes et al., 1964; Boyes and Van Brink, 1965; Dübendorfer et al., 2002). M on the third chromosome (III^M) is also common, but it is not clear what was responsible for the invasion of the III^M chromosome (Hamm et al., 2015). Note that when the M factor arrived on chromosome III, this entire chromosome essentially assumed Y-like properties, including male-biased transmission and reduced recombination (Hamm et al., 2015). However, the III^M chromosome is not a degenerated Y chromosome because III^M homozygotes are viable, fertile, and found in natural populations (Hamm et al., 2015). Identifying the selective forces responsible for the invasion of III^M will be a powerful test of the hypotheses to explain SD evolution.

Strong linkage to A^M is expected for alleles on the same autosome because recombination is low or non-existent in house fly males (Hiroyoshi, 1961; Hamm et al., 2015), but see Feldmeyer et al. (2010). It is possible that A^M chromosomes invaded house fly populations because of selection on phenotypic effects of either the autosomal M loci themselves or alleles linked to M-factors (Franco et al., 1982; Tomita and Wada, 1989; Kozielska et al., 2006; Feldmeyer et al., 2008). M variants are known to have subtle phenotypic effects, which include differential splicing and expression of SD pathway genes between Y^M and A^M males (Schmidt et al., 1997; Hediger et al., 2004; Siegenthaler et al., 2009). In addition, A^M chromosomes form stable latitudinal clines on multiple continents (Franco et al., 1982; Tomita and Wada, 1989; Hamm et al., 2005; Kozielska et al., 2008), and seasonality in temperature is somewhat predictive of their distribution (Feldmeyer et al., 2008). Furthermore, in laboratory experiments, III^M males out-competed Y^M males for female mates; the III^M chromosome increased in frequency over generations in population cages; and III^M males had higher rates of emergence from pupae than Y^M males (Hamm et al., 2009). The most specific phenotype that has been linked to A^M is insecticide resistance (Kerr, 1960, 1961, 1970; Denholm et al., 1983; Kence and Kence, 1992), but insecticide resistance alone cannot entirely explain the invasion of A^M chromosomes (Shono and Scott, 1990; Hamm et al., 2005). These results all support the hypothesis that natural, sexual, or sexually antagonistic selection on M variants or linked alleles drove the invasion of A^M chromosomes.

To test whether sex-specific selection pressures could be responsible for the invasion of the

III^M chromosome, we used high-throughput mRNA sequencing (mRNA-Seq) to compare gene expression profiles between nearly isogenic Y^M and III^M males that only differ in their M-bearing chromosome. These contrasts are essentially a comparison between flies with the ancestral Y chromosome (Y^M) and individuals with a recently evolved "neo-Y" (III^M). The gene expression differences we detected were the result of both differentiation of *cis* regulatory regions between the III^M and "standard" third chromosome and *trans* effects of the III^M and/or Y^M chromosome(s) on expression throughout the genome. We found that genes responsible for male phenotypes are more likely to be differentially expressed between Y^M and III^M males, suggesting that Y^M and III^M males have phenotypic differences that would be differentially affected by male-specific selection pressures. We also found that III^M males have a "masculinized" gene expression profile. These results support the hypothesis that sexual or sexually antagonistic selection drives evolutionary turnover at the top of SD pathways.

2. Materials and Methods

2.1. Strains

We compared gene expression between two nearly isogenic house fly strains that differ only in the chromosome carrying M. The first, Cornell susceptible (CS), is an inbred, lab adapted strain with XX males that are heterozygous for a III^M chromosome and a standard third chromosome that lacks an M factor (X/X; III^M/III^{CS}) (Scott et al., 1996; Hamm et al., 2005) (Figure 1A). CS females are XX and homozygous for the standard third chromosome (X/X; III^{CS}/III^{CS}). We created a strain with Y^M males that has the X chromosome and all standard autosomes from the CS strain. To do so, we used a backcrossing approach to move the Y chromosome from the genome strain (aabys) onto the CS background (Figure 1B), creating the strain CS-aabys-Y (CSaY). CSaY males are XY and homozygous for the standard CS third chromosome (X/Y; III^{CS}/III^{CS}). The aabys strain has a recessive phenotypic marker on each of the five autosomes (Wagoner, 1967; Tomita and Wada, 1989). To confirm that the aabys autosomes had been purged from the CSaY genome, we crossed CSaY flies to aabys and observed only wild-type progeny. CS and

5

CSaY males are nearly isogenic, differing only in that CS males are XX and heterozygous for the III^M and standard III^{CS} chromosomes, and CSaY males are XY and homozygous for the standard III^{CS} chromosome (Figure 1). Females are genetically identical between strains.

We are confident that the strains are isogenic except for the M-bearing chromosome because there is very little evidence for recombination in male house flies with an XY genotype (Hiroyoshi, 1961; Hamm et al., 2015). However, if there were minimal recombination between the CS and aabys chromosomes in our crossing scheme, the majority of autosomal alleles in the CSaY strain would still have originated from the CS genotype, with very little contribution from aabys autosomes.

2.2. Samples and mRNA-Seq

CS and CSaY flies were kept at 25°C with a 12:12 hour light:dark cycle. Larvae were reared in media made with 1.8 L water, 500 g calf manna (Manna Pro, St. Louis, MO), 120 g bird and reptile litter wood chips (Northeastern Products, Warrensburg, NY), 60 g dry active baker's yeast (MP Biomedical Solon, OH), and 1.21 kg wheat bran (Cargill Animal Nutrition, Minneapolis, MN), as described previously (Hamm et al., 2009).

We sampled two types of tissue from CS and CSaY males and females: head and gonad. All dissections were performed on living, non-anesthetized 4–6 day old unmated adult flies. Heads were separated from males and females, homogenized in TRIzol reagent (Life Technologies) using a motorized grinder, and RNA was extracted on QIAGEN RNeasy columns following the manufacturer's instructions including a genomic DNA (gDNA) elimination step. Testes were dissected from males, and ovaries were dissected from females in Ringer's solution (182 mM KCl, 46 mM NaCl, 3 mM CaCl$_2$, 10 mM Tris-Cl in ddH$_2$O). Ovary and testis samples were dissolved in TRIzol and RNA was extracted on QIAGEN RNeasy columns with gDNA elimination. Three biological replicates of CS (III^M) male heads, CSaY (Y^M) male heads, CS testes, and CSaY testes were collected; one sample was collected for each of the four female dissections (CS head, CSaY head, CS ovary, and CSaY ovary).

Barcoded mRNA-Seq libraries were prepared using the Illumina TruSeq kit following the manufacturer's instructions. Briefly, mRNA was purified using oligo-dT magnetic beads,

6

cDNA was synthesized using random hexamer primers, and sequencing libraries were constructed using the cDNA. The samples were run on 2 lanes of an Illumina HiSeq2500 at the Cornell Medical School Genomics Resources Core Facility. One lane had the eight head samples, and the other lane had the eight gonad (testis and ovary) samples. We generated 101 base pair single-end reads, and the sequencing reads were processed using Casava 1.8.2.

2.3. mRNA-Seq data analysis

Illumina mRNA-Seq reads were aligned to house fly genome assembly v2.0.2 and annotation release 100 (Scott et al., 2014) using TopHat2 v2.0.8b (Kim et al., 2013) and Bowtie v2.1.0.0 (Langmead et al., 2009) with the default parameters. We tested for differential expression between males and females and between Y^M and III^M males using the Cuffdiff program in the Cufflinks v2.2.1 package (Trapnell et al., 2013) with the default parameters, including geometric normalization. We used a false discovery rate (FDR) corrected P-value of 0.05 to identify genes that are differentially expressed (Benjamini and Hochberg, 1995). Genes were considered not differentially expressed if Cuffdiff returned an "OK" value for the test status (at least 10 reads aligned to the transcript, and data were sufficient for testing for differential expression) but the expression levels were not significantly different. Genes without an "OK" value were not included in downstream analyses. We repeated this analysis by also requiring a two-fold difference in expression to call genes as differentially expressed. In comparisons between male and female expression levels, we treated all 6 male samples as biological replicates and did the same for both female samples. We also repeated the analysis using only 2 replicates of each sample to control for sample-size effects.

We used expression level estimates from Cuffdiff2 (Fragments Per Kilobase of transcript per Million mapped reads, FPKM) to calculate correlations of expression levels between our experimental samples (Figure S1). Only genes with an "OK" value for test status in Cuffdiff were included. The correlations between testis and ovary expression are lowest, which is expected because they are dramatically different tissues. The correlations between male and female head samples are substantially higher than between testis and ovary, but still lower than the correlations within sexes. The two ovary samples are more highly correlated than any of the pairwise comparisons between CS and CSaY testis samples (Figure S1), most

7

likely because CS and CSaY females are genetically identical (Figure 1). All data analysis was performed in the R statistical programming environment (R Core Team, 2015).

2.4.   Chromosomal assignments of house fly genes

The house fly genomic scaffolds have not formally been assigned to chromosomes, but homologies have been inferred between house fly chromosomes and the five major chromosome arms of *Drosophila*, also known as Muller elements A–E (Foster et al., 1981; Weller and Foster, 1993). Additionally, the house fly X chromosome is most likely homologous to the *Drosophila* dot chromosome (Muller element F, or *Drosophila melanogaster* chromosome 4) (Vicoso and Bachtrog, 2013, 2015). We therefore assigned house fly genes that are conserved as one-to-one orthologs with *D. melanogaster* (Scott et al., 2014) to house fly chromosomes based on the Muller element mapping of the *D. melanogaster* orthologs. For our analysis of gene families that are differentially expressed between Y^M and III^M males, we assigned house fly scaffolds to chromosomes based on the Muller element mapping of the majority of *D. melanogaster* orthologs on each scaffold.

2.5.   Gene ontology analysis

We used the predicted *D. melanogaster* orthologs (Scott et al., 2014) to infer the functions of house fly genes. Gene ontology (GO) annotations of house fly genes were determined using Blast2GO (Conesa et al., 2005; Götz et al., 2008) as described previously (Scott et al., 2014). We then used Blast2GO to identify GO classes that are enriched amongst differentially expressed genes relative to non-differentially expressed genes using an FDR-corrected Fisher's exact test (Benjamini and Hochberg, 1995).

2.6.   qPCR validation of differentially expressed genes

We used quantitative PCR (qPCR) on cDNA to validate the differential expression of four genes between Y^M and III^M testes. Dissections of testes were performed as described above on three batches of Y^M and III^M males each (i.e., 3 biological replicates of each strain). RNA was extracted from testes using TRIzol homogenization followed by purification using Direct-zol RNA MiniPrep columns (Zymo Research). We synthesized cDNA using M-MLV reverse

transcriptase (Promega) with oligo-dT primers following the manufacturer's instructions. We designed PCR primers to amplify a 71–110 bp product at the 3' end of each transcript, with one primer on either side of the 3'-most annotated intron when possible (Table S1). We also designed primers to amplify one transcript that was not differentially expressed between Y^M and III^M males or between males and females in either gonad or head in our mRNA-Seq analysis (XM_005187313). We tested our primer pairs with PCR using testis cDNA as a template to validate that they amplify a single product.

We then performed qPCR on 3 technical replicates of 5 serial dilutions of 1/5 each using a 60°C annealing temperature on an Applied Biosystems StepOnePlus Real-Time PCR System with Power SYBR Green PCR Master Mix (Life Technologies) following the manufacturer's instructions. We assigned a threshold cycle (C_T) to the qPCR curves, and we validated that the primer pairs gave a linear relationship between C_T and $-log_{10}$ concentration. We next used the same reagents and equipment to perform qPCR using each primer pair on 3 technical replicates of the 3 biological replicates from both strains (18 samples total), with the samples interspersed on a 96-well plate to avoid batch effects. In addition to the 18 samples, each qPCR plate contained 3 technical replicates of a 5-step 1/5 serial dilution (15 samples). Those 33 samples were amplified by qPCR using primers for a gene that was detected as differentially expressed using mRNA-Seq and the 33 samples were also amplified with primers in our control gene (XM_005187313), for a total of 66 reactions on a single plate. From each plate, we constructed a standard curve for each primer pair using the relationship between C_T and $-log_{10}$ concentration from our serial dilutions, and we used the slope of these lines to estimate the initial concentration of our template cDNA in each sample. We divided the initial concentration for our experimental gene by the estimated concentration for the control gene to determine a relative concentration for the experimental gene for each of the technical replicates.

To test for differential expression between Y^M and III^M samples, we first constructed a linear model with replicates nested within strains predicting the relative concentration of the gene in the R statistical programming environment (R Core Team, 2015). We then used Tukey's Honest Significant Differences method to perform an analysis of variance to determine if there is a significant effect of strain on expression level.

3. Results

3.1. Genes on the house fly third chromosome are more likely to be differentially expressed between Y^M and III^M males than genes on other autosomes

We used mRNA-Seq to compare gene expression in heads and gonads of house fly males and females of a Y^M strain (CSaY) and a III^M strain (CS). Males of the III^M strain are XX and heterozygous for the III^M chromosome and a standard third chromosome without M (Figure 1A). Males of the Y^M strain are XY (with the same X as the III^M strain) and homozygous for the standard third chromosome found in the III^M strain (Figure 1B). The rest of the genome is isogenic, and females of the two strains are genetically identical (Figure 1).

We detected 873 and 1338 genes that are differentially expressed between Y^M and III^M males in heads or testes, respectively, at an FDR-corrected P-value of 0.05 (Table 1, Figure S2, Supplementary Data). Genes on the house fly third chromosome are more likely than genes on other autosomes to be differentially expressed between Y^M and III^M males (Figure 2A). Approximately 30% of the differentially expressed genes are predicted to be on the third chromosome, which is greater than the fraction assigned to any of the other four autosomes (14.8–20.6%). There is a slight, but significant, signal of higher expression from the third chromosome in Y^M males when compared to III^M males (Figure 2B), and a significant excess of third chromosome genes is up-regulated in Y^M males (Figure S3). However, a significant excess of third chromosome genes is also up-regulated in III^M males (Figure S3), suggesting that the differential expression of third chromosome genes between Y^M and III^M males is not the result of degeneration of the III^M chromosome.

X-linked genes also trend towards an excess that are differentially expressed between Y^M and III^M males (Figure 2A), but we do not have the power to detect statistically significant deviations from the expectation because only 56 X-linked genes are expressed in head and 52 X-linked genes are expressed in testis. Surprisingly, expression from the X chromosome is higher in Y^M (XY) than III^M (XX) male heads (Figure 2B), and a significant excess of X-linked genes is up-regulated in Y^M heads relative to III^M heads (Figure S3). These results demonstrate that differential expression of X-linked genes between Y^M and III^M males

is not the result of a dosage deficiency in hemizygous males. In addition, these patterns suggest that either a dosage compensation mechanism provides greater than two-fold up-regulation of the X chromosome in XY males or *trans* effects of the Y chromosome lead to up-regulation of X-linked expression.

Our chromosomal assignments of house fly genes are almost certainly less than perfect because of gene movement between chromosomes since the divergence between *D. melanogaster* the *M. domestica* lineages. However, errors in chromosomal assignments should obstruct the signal of elevated expression divergence on the third and X chromosomes, making our results conservative.

3.2. More differential expression between Y^M and III^M males in testis than in head, but a common set of genes co-regulated in both tissues

A higher fraction of genes is differentially expressed in testes between Y^M and III^M males than in heads (Table 1; $P < 10^{-16}$ in Fisher's exact test, FET), suggesting that genes involved in male fertility phenotypes are more affected by the M-bearing chromosome. When we restricted the analysis to only genes expressed in both heads and gonads, we still observe more genes differentially expressed in testes than heads between Y^M and III^M males (Table S2). When we used a two-fold cutoff in addition to an FDR-corrected $P < 0.05$ cutoff, the number of genes differentially expressed in head and testis between Y^M and III^M males goes down to 373 and 558, respectively. However, there is still a higher fraction of genes differentially expressed in testis than head ($P_{FET} < 10^{-5}$)

Genes that are differentially expressed between males and females are said to have "sex-biased" expression (Ellegren and Parsch, 2007). The fraction of genes differentially expressed between the testes of Y^M and III^M males is nearly as large as the fraction with sex-biased expression in head (Table 1). We have a different number of replicates for male samples (three Y^M and three III^M, for a total of six male replicates) than female samples (two), and so our power to detect differential expression may differ between the inter-strain (Y^M vs. III^M) and inter-sex comparisons. To control for sample size effects, we repeated our analysis using only two replicates of each male strain in the inter-strain comparison and two male replicates (one from each strain) in the inter-sex comparison. With only two replicates

11

of each sample we confirmed that more genes are differentially expressed between Y^M and III^M testes than heads, and we found that more genes are differentially expressed between Y^M and III^M males in either tissue than have sex-biased expression in head (Table S3). This result confirms that the inter-strain expression differences are of a similar or greater magnitude than the amount of sex-biased expression in head.

If the probability that a gene is differentially expressed between Y^M and III^M male heads is independent of the probability that the gene is differentially expressed in testes, we expect <1% of genes to be differentially expressed in both head and testis. We find that 176 genes (2.12%) are differentially expressed between Y^M and III^M males in both head and testis when using an FDR-corrected P-value to test for differential expression, which is significantly greater than the expectation ($P_{FET} < 10^{-25}$). In contrast, there is not a significant excess of genes with sex-biased expression in both head and gonad—we expect 9.41% of genes to have sex-biased expression in both head and gonad (Table S2), and we observe that 809 genes (9.27%) are sex-biased in both tissue samples ($P_{FET} = 0.655$). We obtain qualitatively similar results when using a two-fold cutoff in addition to an FDR-corrected P-value threshold of 0.05 to test for differential expression: there is a four-fold excess of genes that are differentially expressed in both head and testis between Y^M and III^M males ($P_{FET} < 10^{-26}$), and a <10% excess of genes that are differentially expressed in both head and gonad between males and females ($P_{FET} = 0.036$). We also get the same result when analyzing only two replicates of each sample: a significant excess of genes is differentially expressed in both head and testis between Y^M and III^M males ($P_{FET} < 10^{-4}$), but not between males and females in both head and gonad ($P_{FET} > 0.3$). These results suggest that there are many genes under common regulatory control by the M-bearing chromosome in both male head and testis, but there is not the same degree of sex-specific regulation in common between head and gonad.

3.3. Genes that are differentially expressed between Y^M and III^M males are more likely to have male-biased expression

Genes whose expression is significantly higher in males than females are said to have "male-biased" expression, and genes that are up-regulated in females have "female-biased" ex-

pression (Ellegren and Parsch, 2007). We found that genes with male-biased expression in head are more likely to be differentially expressed between Y^M and III^M male heads than genes with either female-biased or unbiased expression (Figure 3A). Similarly, genes that are up-regulated in testis relative to ovary ("testis-biased") are more likely to be differentially expressed between Y^M and III^M testes than genes with "ovary-biased" or unbiased expression in gonad (Figure 3B). We repeated this analysis using two replicates of each sample, and we consistently observe that genes with male-biased expression in head or gonad are more likely to be differentially expressed between Y^M and III^M males (Figures S4 and S5).

We also found that 14.8% of genes that are up-regulated in III^M male heads have male-biased expression, whereas $<2\%$ of genes that are up-regulated in Y^M male heads have male-biased expression ($P_{FET} < 10^{-10}$). We observe the same excess of male-biased genes up-regulated in III^M male heads when we only use two replicates of each strain and sex to test for differential expression ($P_{FET} < 0.05$ using most combinations of two replicates). This suggests that III^M male heads have a "masculinized" expression profile relative to Y^M heads.

3.4. Functional annotations of genes that are differentially expressed between Y^M and III^M males

We tested for GO categories that are over-represented amongst genes with sex-biased expression that are differentially expressed between Y^M and III^M males (Supplementary Data). We found that nearly half (49.7%) of genes that are differentially expressed between Y^M and III^M male heads are annotated with the functional category "catalytic activity", whereas only 43% of genes not differentially expressed have that GO annotation ($P_{FET} < 0.05$ corrected for multiple tests). Over 10% of the genes with the catalytic activity annotation that are up-regulated in III^M male head are predicted to encode proteins involved in a metabolic process, including metabolism of organic acids, amino acids, and lipids. Among those genes, 15 are annotated as cytochrome P450 (CYP450) genes, and four of those also have male-biased expression in head (Table S6). CYP450s collectively carry out a wide range of chemical reactions including metabolism of endogenous (e.g., steroid hormones)

and exogenous (e.g., xenobiotics) compounds (Scott, 2008). All 15 differentially expressed CYP450s are up-regulated in III^M males, and no CYP450 genes are up-regulated in Y^M males. Five of the CYP450s are on scaffolds that we assign to the third chromosome (Table S6), suggesting that *cis* regulatory sequences controlling the expression of CYP450s have diverged between III^M and the standard third chromosome. However, five of the CYP450s can be assigned to other autosomes (the remaining 5 cannot be assigned to a chromosome), demonstrating that divergence of *trans* factors between III^M and the standard third chromosome are also responsible for differential expression of CYP450s between Y^M and III^M males. The 15 CYP450s represent a range of different clans (2, 3 and 4) and families (4, 28, 304, 313, 438 and 3073) (Scott et al., 2014). However, an excess of CYP450s from clan 4 are up-regulated in III^M male head ($\chi^2 = 4.19$, $P = 0.041$), and thus over-expression of CYP450s is not random.

Genes that are annotated as encoding proteins located in extracellular regions are over-represented amongst genes with testis-biased expression (15.0% of genes with testis-biased expression; 9.9% of genes not differentially expressed between testis and ovary; $P_{FET} < 10^{-3}$ corrected for multiple tests) and amongst genes that are differentially expressed between Y^M and III^M testes (13.9% of differentially expressed genes; 8.1% of non-differentially expressed genes; $P_{FET} < 10^{-4}$ corrected for multiple tests). In addition, 3.1% of the genes differentially expressed between Y^M and III^M testes are predicted to encode carbohydrate binding proteins (compared to 1.4% of non-differentially expressed genes; $P_{FET} < 0.05$ corrected for multiple tests), and 7.2% of differentially expressed genes are predicted to encode structural molecules (compared to 3.7% of non-differentially expressed genes; $P_{FET} < 10^{-3}$ corrected for multiple tests). Three of those structural molecules are predicted to be β-tubulin proteins encoded by genes that are up-regulated in Y^M testes relative to III^M, and two of those genes also have testis-biased expression. We tested for differential expression of the two of the β-tubulin genes with testis-biased expression using qPCR (Figure S6; Supplementary Data). Only one of the two (XM_005187368) was up-regulated in Y^M testis when assayed with qPCR ($P < 10^{-4}$), whereas the other gene (XM_005175742) was not ($P = 0.653$) possibly because of high variance in the Y^M measurement (Figure S6). The *D. melanogaster* genome encodes a testis-specific β-tubulin paralog that is essential for spermatogenesis (Kemphues

et al., 1982; Hoyle and Raff, 1990), suggesting that the β-tubulin gene that is up-regulated in Y^M testis may be important for sperm development.

Four genes that are differentially expressed between Y^M and III^M testes are homologs of the *D. melanogaster* Y-linked fertility factors *kl-2*, *kl-3*, and *kl-5* (Goldstein et al., 1982; Gepner and Hays, 1993; Carvalho et al., 2000, 2001). These proteins encode components of the dynein heavy chain, which is necessary for flagellar activity of sperm. All four of the predicted dynein heavy chain genes that are differentially expressed between Y^M and III^M testes are autosomal in house fly. Three of these genes have testis-biased expression— two of those are up-regulated in III^M testis (XM_005175130, XM_005176585), while the third is up-regulated in Y^M testis (XM_005184828). The fourth gene (XM_005184771) is up-regulated in III^M testis, but it is not differentially expressed between testis and ovary. Using qPCR, we validated that XM_005184828 is up-regulated in Y^M testis ($P < 10^{-4}$) and XM_005176585 is up-regulated in III^M testis ($P < 0.01$) (Figure S6; Supplementary Data). Two additional genes encoding components of other dynein chains have testis-biased expression and are up-regulated in III^M testis relative to Y^M testis.

Finally, there are numerous predicted RNAs in the house fly genome annotation that have no identifiable homology to any known RNAs or proteins (Scott et al., 2014). We identified six of these uncharacterized RNAs that both have testis-biased expression and are differentially expressed between Y^M and III^M testes. XR_225504, XR_225520, and XR_225639 are up-regulated in Y^M testes, and XR_225442, XR_225497, and XR_225737 are up-regulated in III^M testes. These genes are annotated as encoding non-coding RNAs (ncRNAs), and we were unable to detect long open reading frames in the transcripts. It is possible that these ncRNAs are responsible for the regulation of gene expression in testis, and differential expression of these ncRNAs between Y^M and III^M testis could be responsible for the differential expression of other genes between Y^M and III^M males.

15

4. Discussion

4.1. Differential expression between Y^M and III^M males is driven by both cis and trans effects

We compared gene expression in head and testis between Y^M and III^M males. Y^M males are homozygous for a standard third chromosome that does not have M, whereas III^M males are heterozygous for a III^M chromosome and a standard third chromosome (Figure 1). Differences in the expression levels of autosomal genes between Y^M and III^M males could be the result of: 1) divergence of *cis*-regulatory sequences between the III^M and standard third chromosomes that affect the expression of genes on the third chromosome; 2) divergence of *trans*-factors between III^M and the standard third chromosome that differentially regulate gene expression throughout the genome; 3) downstream effects of the first two processes that lead to further differential expression.

The two strains also differ in the genotype of their sex chromosomes; Y^M males are XY, whereas III^M males are XX (Figure 1). The house fly Y chromosome is highly heterochromatic and does not harbor any known genes other than M (Boyes et al., 1964; Hediger et al., 1998; Dübendorfer et al., 2002). It is clear that the Y chromosome does not contain any genes necessary for male fertility or viability because XX males are fertile and viable. The X chromosome is also highly heterochromatic and probably homologous to the *Drosophila* dot chromosome (Hediger et al., 1998; Vicoso and Bachtrog, 2013, 2015). The heterochromatic *Drosophila* Y chromosome can affect the expression of autosomal genes (Lemos et al., 2008, 2010; Sackton et al., 2011; Zhou et al., 2012), suggesting that the house fly X and Y chromosomes could have *trans* regulatory effects on autosomal gene expression.

A higher fraction of third chromosome genes are differentially expressed between Y^M and III^M house fly males than genes on any other autosome (Figure 2A). Therefore, divergence of *cis*-regulatory sequences between III^M and the standard third chromosome are at least partially responsible for the expression differences between Y^M and III^M males. However, $\sim70\%$ of the genes differentially expressed between Y^M and III^M males map to one of the other four autosomes, suggesting that the majority of expression differences are the result of *trans* effects of the X, Y, and third chromosomes along with further downstream effects.

16

4.2. Reproductive and male phenotypes are more likely to be affected by M variation

Reproductive traits are more sexually dimorphic than non-reproductive traits, and reproductive traits also tend to evolve faster, possibly as a result of sexual selection (Eberhard, 1985). A similar faster evolution of gene expression in reproductive tissues has been observed across many taxa (Khaitovich et al., 2005; Zhang et al., 2007; Brawand et al., 2011), and increased variation within species for sex-biased gene expression often accompanies elevated expression divergence (Meiklejohn et al., 2003; Ayroles et al., 2009). Consistent with these patterns, more genes are differentially expressed between Y^M and III^M males in testis than head (Table 1, Figure S2). Somatic and germline SD in house fly are under the same genetic control (Hilfiker-Kleiner et al., 1994), so the exaggerated differences in expression between Y^M and III^M testes relative to heads cannot be attributed to differences in the SD pathway between gonad and head. We also find that genes with male-biased expression are more likely to be differentially expressed between Y^M and III^M males (Figure 3). Genes with male-biased expression are more likely to perform sex-specific functions (Connallon and Clark, 2011), suggesting that genes that are differentially expressed between Y^M and III^M males disproportionately affect male phenotypes.

4.3. Evaluating the role of sex-specific selection in MSD turnover

Many models of SD evolution predict that a new MSD locus will invade a population if it is genetically linked to an allele with a beneficial, sexually selected, or sexually antagonistic fitness effect (Charlesworth and Charlesworth, 1980; Rice, 1987; Charlesworth, 1991, 1996; Rice, 1996; van Doorn and Kirkpatrick, 2007, 2010). Alternatively, evolutionary turnover of MSD loci could be the result of neutral drift in a highly labile system (van Doorn, 2014).

Our results are consistent with a model in which the III^M chromosome invaded because it harbors alleles with male-specific beneficial effects. First, the expression of genes that are likely to perform male-specific functions—especially in male fertility—are more likely to be affected by the III^M chromosome (Table 1; Figure 3), and those male-specific phenotypic differences could have been targets of sex-specific selection pressures. However, as mentioned above, the expression of male-biased genes is more variable than other genes even in species

17

without multifactorial SD systems (Meiklejohn et al., 2003; Ayroles et al., 2009). Additional experiments in which a non-M-bearing chromosome is placed on a common genetic background are therefore necessary to further test the hypothesis that the M-bearing chromosome disproportionately affects male-biased gene expression.

We also found that III^M heads have a masculinized expression profile relative to Y^M heads, suggesting that the male-limited transmission of the III^M chromosome favored the accumulation of alleles with male-beneficial fitness effects (Rice, 1984). Previous work found that III^M males outperformed Y^M males in multiple laboratory fitness assays (Hamm et al., 2009), providing additional support for the accumulation of male-beneficial alleles on the III^M chromosome. However, despite the apparent selective advantage of the III^M chromosome, it surprisingly does not appear to be expanding rapidly (Hamm et al., 2015), suggesting that the fitness benefits of III^M may be environment-specific (Feldmeyer et al., 2008).

Our data do not allow us to distinguish between two possible orders of events in the invasion of the III^M chromosome. In the first scenario, male-beneficial alleles on the third chromosome could have driven the initial invasion of III^M (van Doorn and Kirkpatrick, 2007). In the second scenario, beneficial alleles could have accumulated on the III^M chromosome after it acquired an M-locus because male-limited inheritance promotes the fixation of male-beneficial alleles (Rice, 1984, 1987). These scenarios are not mutually exclusive. Regardless of the sequence of events, we have provided evidence that the house fly multifactorial male-determining system is associated with phenotypic differences that likely have male-specific fitness effects, which could explain the invasion of the III^M chromosome under sexual or sexually antagonistic selection.

5. Acknowledgements

6. Tables

Tissue	comparison	# diff		# non-diff	# genes	freq diff
		Y^M	III^M			
male head	Y^M *vs.* III^M	320	553	8568	9441	0.092
testis	Y^M *vs.* III^M	782	556	8937	10275	0.130
		male	female			
head	male *vs.* female	205	1087	8387	9679	0.133
gonad	testis *vs.* ovary	3426	4369	2791	10586	0.736

Table 1: Differential expression between strains and sexes. Counts of the number of genes that are differentially expressed (# diff), tested and non-differentially expressed (#non-diff), and total genes tested (# genes), as well as the frequency of genes that are differentially expressed (freq diff). For the # diff columns, the number of genes up-regulated in each of the two samples being compared are presented.
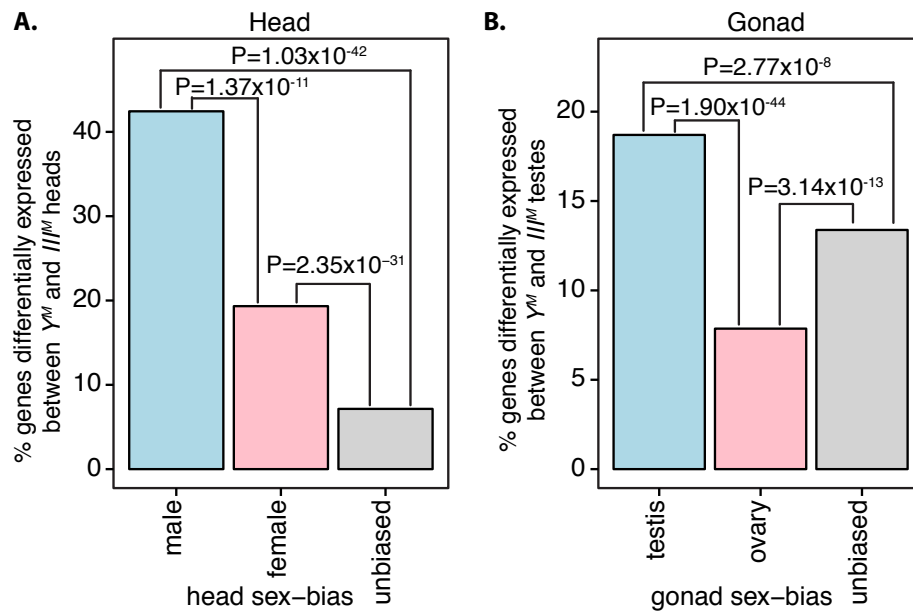
7. Figures
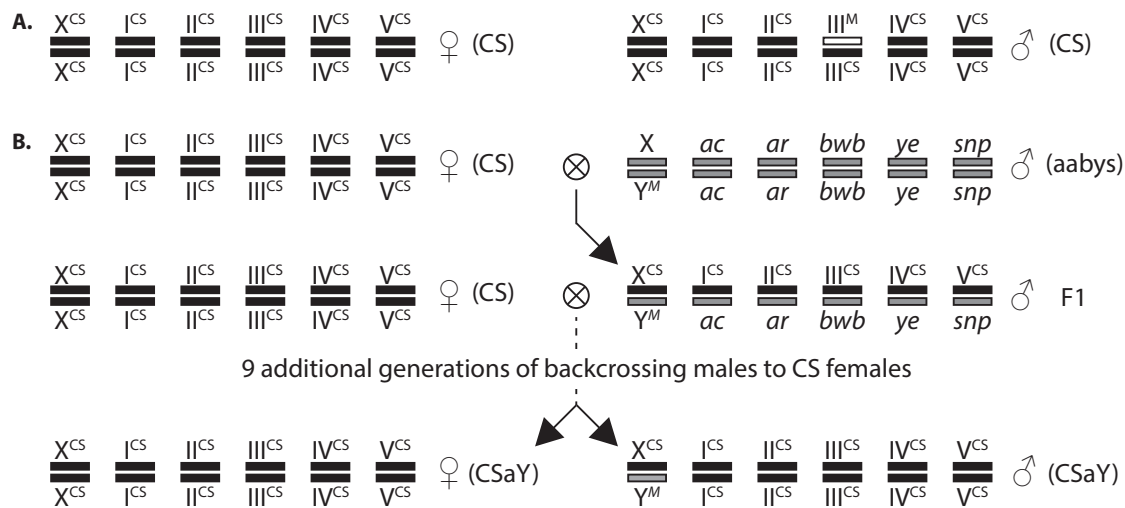
Figure 1: Genotypes of flies. **A.** Genotypes of CS males and females. **B.** Crossing scheme used to generate the CSaY (Y^M) strain. Each pair of parallel rectangles represent homologous chromosomes; there is one pair of sex chromosomes (X and Y) and five autosomes. Chromosomes of CS origin are black and indicated by "CS", except for the III^M chromosome which is white. Chromosomes of aabys origin are gray, and the aabys strain has a recessive phenotypic marker on each autosome. Females from the CS strain were crossed to aabys males, and the male progeny were backcrossed to CS females for 10 generations to create the CSaY strain. Because there is no recombination in XY males, CS and CSaY are isogenic except that the CSaY males have a Y chromosome and CS males have a III^M chromosome.
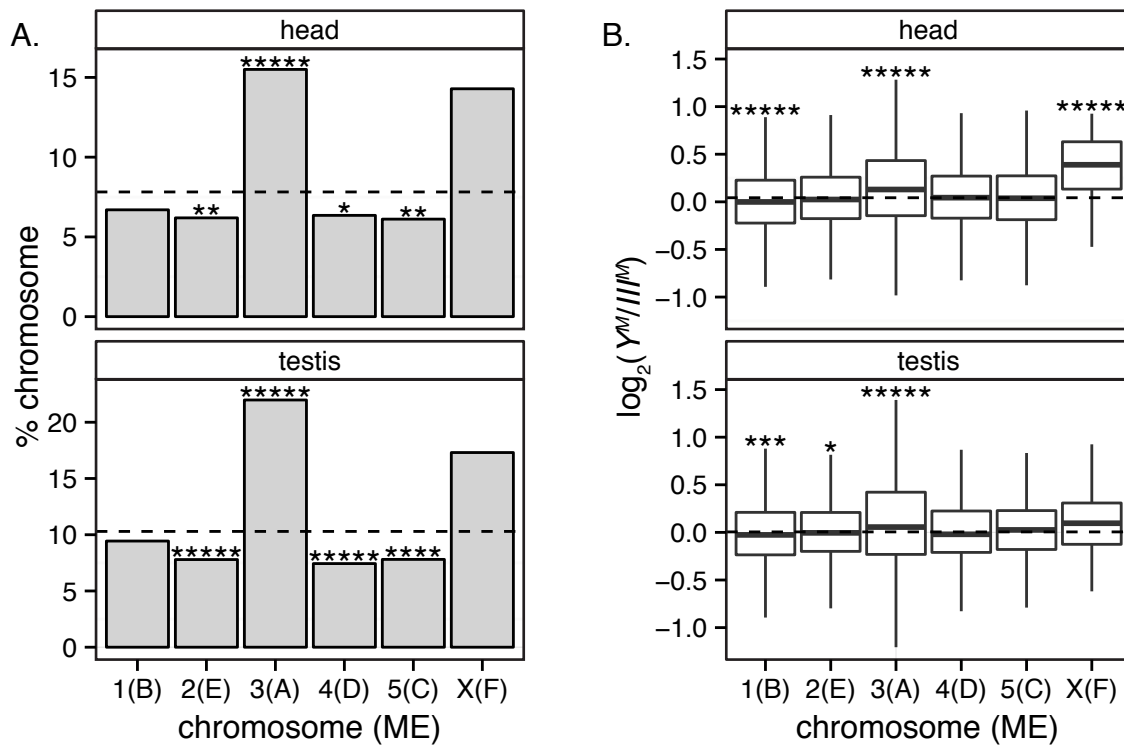
Figure 2: Chromosomal mapping and differential expression between Y^M and III^M males. A. Bar graphs indicate the percent of genes on each chromosome (*Drosophila* Muller element in parentheses) that are differentially expressed between Y^M and III^M male heads (top) or testes (bottom). The dashed line indicates the percentage of genes that are differentially expressed across all chromosomes. Asterisks indicate P-values from Fisher's exact test comparing the number of differentially expressed genes with the number not differentially expressed on a chromosome and summed across all other chromosomes (*$P < 0.05$, **$P < 0.005$, ****$P < 0.00005$, *****$P < 0.000005$). **B.** Box plots show the relative expression levels of genes in Y^M and III^M males on each chromosome. Expression level was measured in head (top) and testis (bottom). The dashed line indicates the average log_2 expression ratio across all genes. Asterisks indicate P-values from a Mann-Whitney test comparing the expression ratio between genes on one chromosome versus the rest of the genome (*$P < 0.05$, **$P < 0.005$, ***$P < 0.0005$, ****$P < 0.00005$, *****$P < 0.000005$).

22

Figure 3: Sex-biased expression of genes differentially expressed between Y^M and III^M males.
Bar graphs indicate the percentage of genes with male-biased (blue), female-biased (pink), or unbiased (gray) expression in either (**A**) head or (**B**) gonad that are differentially expressed between Y^M and III^M males in either (**A**) head or (**B**) testis. P-values are for Fisher's exact test between groups.

Supplementary Tables

gene	gene family	forward	reverse	product (bp)
XM_005187313	control	GTTGTGTCGCAAATATGGCTTG	ACCACTCATACGCTGCAAAAG	106
XM_005184828	dynein	ACGTTCTAATGCCTGTGGTAC	ACCGGACACGTATAAAGTTTCG	79
XM_005176585	dynein	TCACCATGGTCTCCCAAGTTG	ATGCCCTCCATCCAAATTCG	71
XM_005187368	β-tubulin	CCGAAAGCAACATGAAACGAC	CGAATTCTGCATCTTCGTCAGC	78
XM_005175742	β-tubulin	TTTTGGATGTGGTGCGCAAG	AATCAGCAAGGTACCCATGC	110

Table S1: qPCR primers to test for differential expression.

25

Tissue	comparison	# diff		# non-diff	# genes	freq diff
		Y^M	III^M			
male head	Y^M *vs.* III^M	260	426	7628	8314	0.083
testis	Y^M *vs.* III^M	587	386	7341	8314	0.117
		male	female			
head	male *vs.* female	143	931	7522	8596	0.125
gonad	testis *vs.* ovary	2386	3989	2221	8596	0.742

Table S2: Differential expression between strains and sexes using the same genes in gonad and head tissue samples. Same as Table 1 except only genes that are expressed in both head and gonad are counted.

Tissue	comparison	# diff		# non-diff	# genes	freq diff
		Y^M	III^M			
male head	Y^M *vs.* III^M	321	316	8829	9465	0.067
testis	Y^M *vs.* III^M	602	482	9285	10365	0.104
		male	female			
head	male *vs.* female	232	170	9891	10293	0.039
gonad	testis *vs.* ovary	3549	3308	3616	10473	0.655

Table S3: Differential expression between strains and sexes using only two replicates. Same as Table 1 except only two biological replicates of male samples were used in the analysis to control for sample-size effects. In the Y^M *vs.* III^M comparisons, we present the average of the number of genes differentially expressed between 1) the two Y^M strains that are best correlated and the two III^M strains that are best correlated; 2) the two Y^M strains that are worst correlated and the two III^M strains that are worst correlated; 3) a third pair of Y^M strains and a third pair of III^M strains. Correlations can be seen in Figure S1. In the male *vs.* female and testis *vs.* ovary comparisons, we present the average of the number of genes differentially expressed between 1) the first two Y^M and III^M replicates and the second two female replicates; 2) the second two Y^M and III^M replicates and the second two female replicates; 3) the third two Y^M and III^M replicates and the third two female replicates.

Y^M *vs.* III^M head:	differentially expressed		not diff. expressed		
sex-bias (head)	# genes	freq	# genes	freq	P_{FET}
male-biased	87	0.101	118	0.014	1.03×10^{-42}
female-biased	196	0.228	818	0.097	2.35×10^{-31}
sex-biased	283	0.329	936	0.111	3.21×10^{-57}
unbiased	577	0.671	7498	0.889	

Table S4: Sex-biased expression in head of genes that are differentially expressed between Y^M and III^M male heads. Sex-biased expression is based on comparisons of male and female heads. The number of genes (# genes) that are differentially expressed between strains and non-differentially expressed is reported for each sex-bias category. We also report the frequency of genes within the differential expression class that are in each sex-bias category (freq). A *P*-value is given for Fisher's exact test comparing differentially and non-differentially expressed genes in each sex-bias category with genes that have unbiased expression (P_{FET}).

28

Y^M vs. III^M testis:	differentially expressed		not diff. expressed		
sex-bias (gonad)	# genes	freq	# genes	freq	P-value
testis	629	0.481	2735	0.309	2.77×10^{-8}
ovary	325	0.248	3810	0.431	3.14×10^{-13}
sex-biased	954	0.729	6545	0.740	0.381
unbiased	355	0.271	2298	0.260	

Table S5: Sex-biased expression in gonad of genes that are differentially expressed between Y^M and III^M testes. Sex-biased expression is based on comparisons of testis and ovary. The number of genes (# genes) that are differentially expressed between strains and non-differentially expressed is reported for each sex-bias category. We also report the frequency of genes within the differential expression class that are in each sex-bias category (freq). A P-value is given for Fisher's exact test comparing differentially and non-differentially expressed genes in each sex-bias category with genes that have unbiased expression (P_{FET}).

| name	gene ID	protein ID	Clan	male head RPKM III^M	male head RPKM Y^M	q value	house fly sex-bias whole fly	house fly sex-bias gonad	house fly sex-bias head	chr (ME)
CYP4D56	LOC101897841	XP_005177308	4	15.16	9.15	0.02033	no	no	no	3 (A)
CYP4E12	LOC101897587	XP_005186334	4	2.03	0.29	0.04781	NA	NA	NA	5 (C)
CYP4G2	LOC101887882	XP_005176351	4	5.49	0.29	0.00138	female	male	female	3 (A)
CYP4G13	LOC101887550	XP_005176349	4	9.33	0.44	0.00138	no	male	no	3 (A)
CYP4G99	LOC101890640	XP_005177793	4	48.07	31.12	0.01919	female	male	male	3 (A)
CYP4P10	LOC101891224	XP_005180953	4	147.11	98.44	0.02389	no	male	female	5 (C)
CYP28G6	LOC101890714	XP_005184312	3	76.29	31.07	0.00138	no	NA	male	NA
CYP28J1	LOC101891587	XP_005184316	3	37.12	19.34	0.00249	no	no	no	NA
CYP28*	LOC101900938	XP_005192026	3	46.07	17.13	0.00138	NA	male	no	NA
CYP304A1	LOC101890089	XP_005183115	2	2.15	0.54	0.01327	female	male	no	2 (E)
CYP313D1	LOC101887770	XP_005175098	4	14.57	6.30	0.03271	NA	NA	no	NA
CYP313D1	LOC101890728	XP_005189882	4	505.51	274.52	0.00138	no	no	male	2 (E)
CYP313D2	LOC101893162	XP_005188336	4	34.67	11.79	0.00138	female	no	male	2 (E)
CYP438A4	LOC101900906	XP_005183433	3	216.27	113.87	0.00138	no	NA	female	3 (A)
CYP3073B1	LOC101896469	XP_005187068	4	120.8	60.90	0.00138	female	no	female	NA

Table S6: CYP450 genes that are differentially expressed between III^M and Y^M male heads. The asterisk (*) indicates a gene with an incomplete sequence in the assembly. The q value is an FDR-corrected P value comparing expression in III^M and Y^M heads. Sex-biased expression (male, female, and no) is given for whole fly measurements (Scott et al., 2014), gonad, and head. Genes for which expression level was too low to test for sex-biased expression are given as NA. House fly chromosome arm assignments (and Muller element homologies) are provided for genes that are on scaffolds where the majority of 1:1 *D. melanogaster* homologs are on a single Muller element (NA is listed for genes whose scaffolds could not be assigned to a chromosome arm).

Supplementary Figures
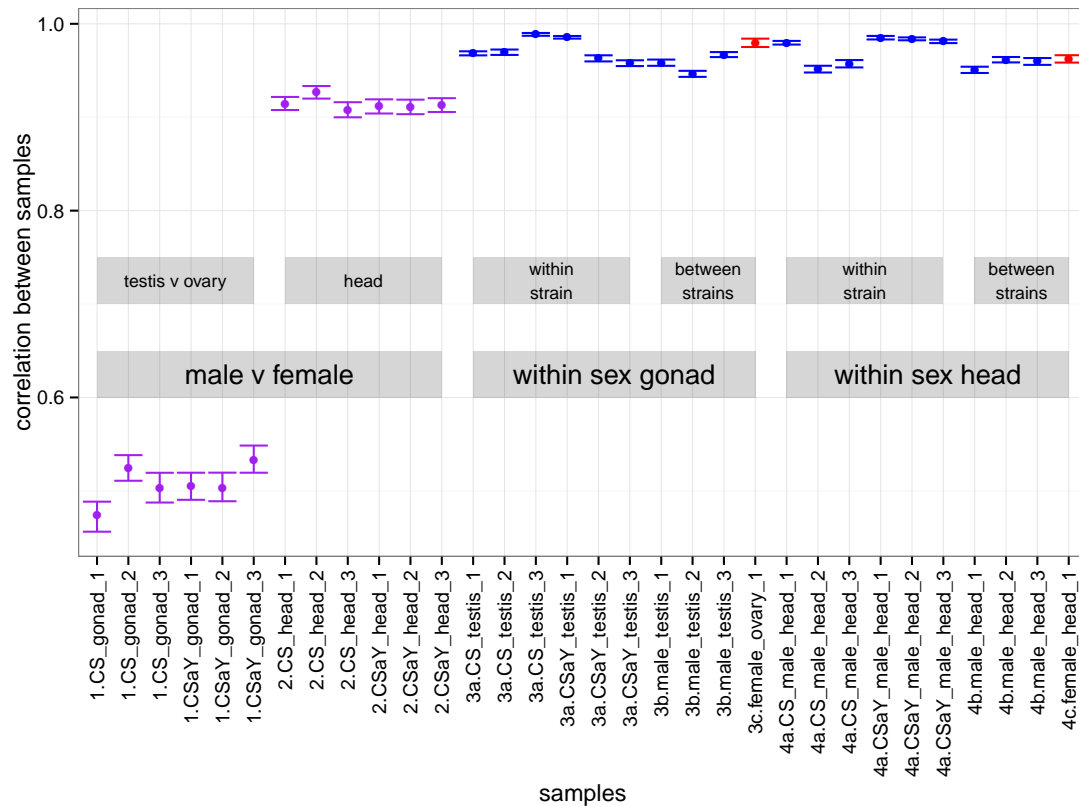
Figure S1: Correlations of expression levels of genes between samples. Point estimates of Spearman's correlation coefficient (ρ) of gene expression between samples are plotted, along with the 95% confidence intervals (CIs) of the correlations. CIs were estimated by bootstrapping the data 1000 times. Comparisons between male and female samples are colored purple, comparisons between male samples are colored blue, and comparisons between female samples are colored red.
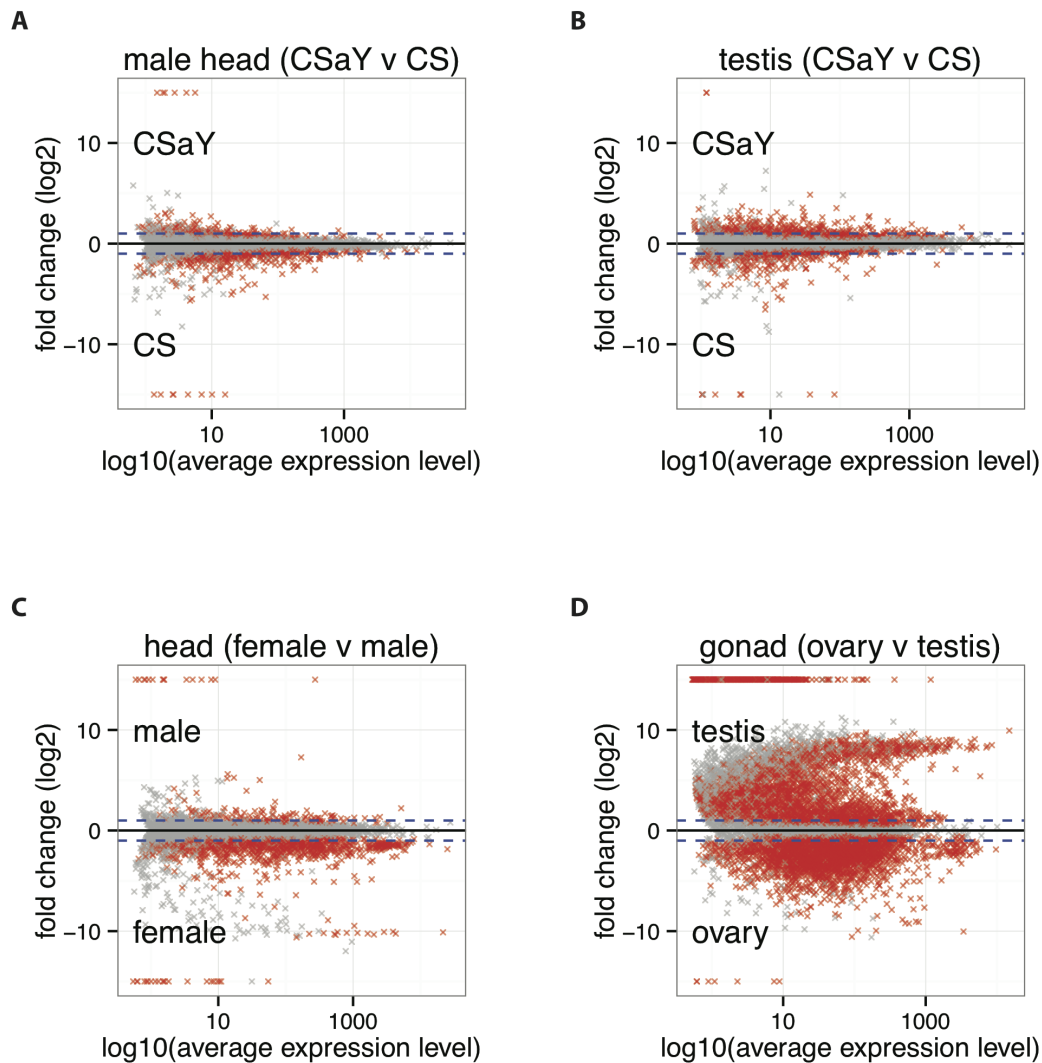
Figure S2: MA plots for comparisons of expression between (A) male heads from the CSaY (Y^M) and CS (III^M) strains, (B) testes from the CSaY and CS strains, (C) heads from females and males, and (D) testes and ovaries. The fold change (log_2 expression ratio) is plotted against log_{10} of the average expression in the two samples being compared. Gray points represent genes whose expression does not significantly differ between strains or sexes (darker gray indicates a higher density of points), and red points represent genes whose expression does differ significantly. Points at the top or bottom of the plot (log_2 fold change -15 or 15) are genes with an expression value of zero in one strain or sex. Only genes that were tested for significantly differential expression are plotted. The dashed blue lines are the 2-fold cutoffs.
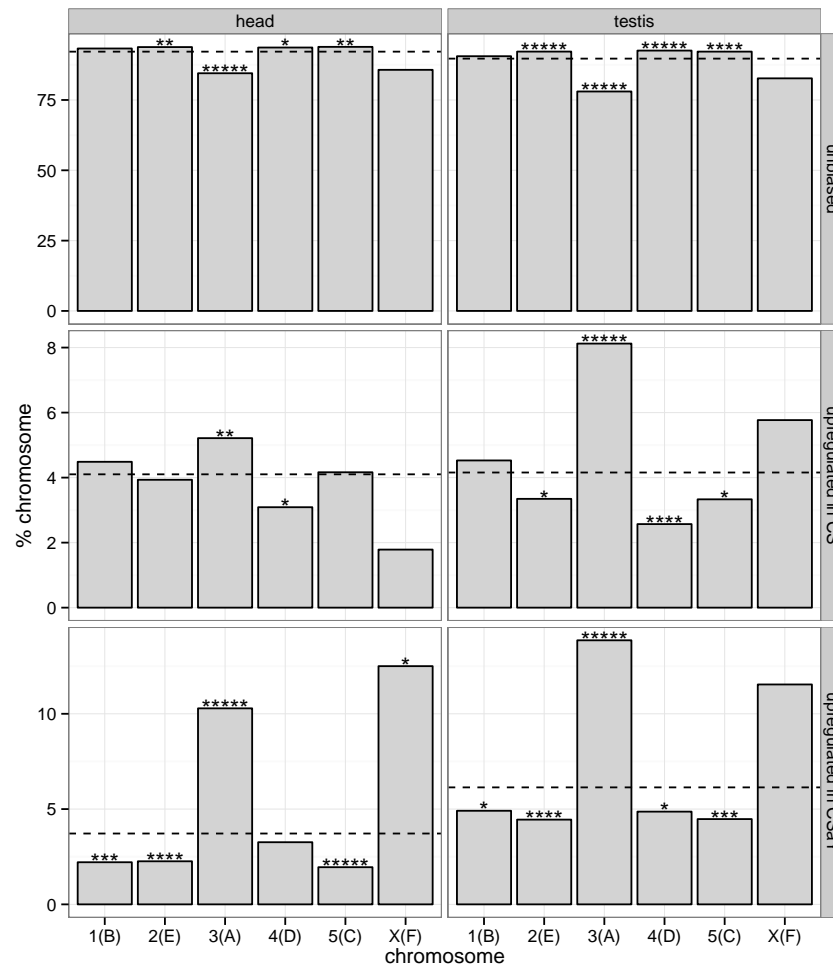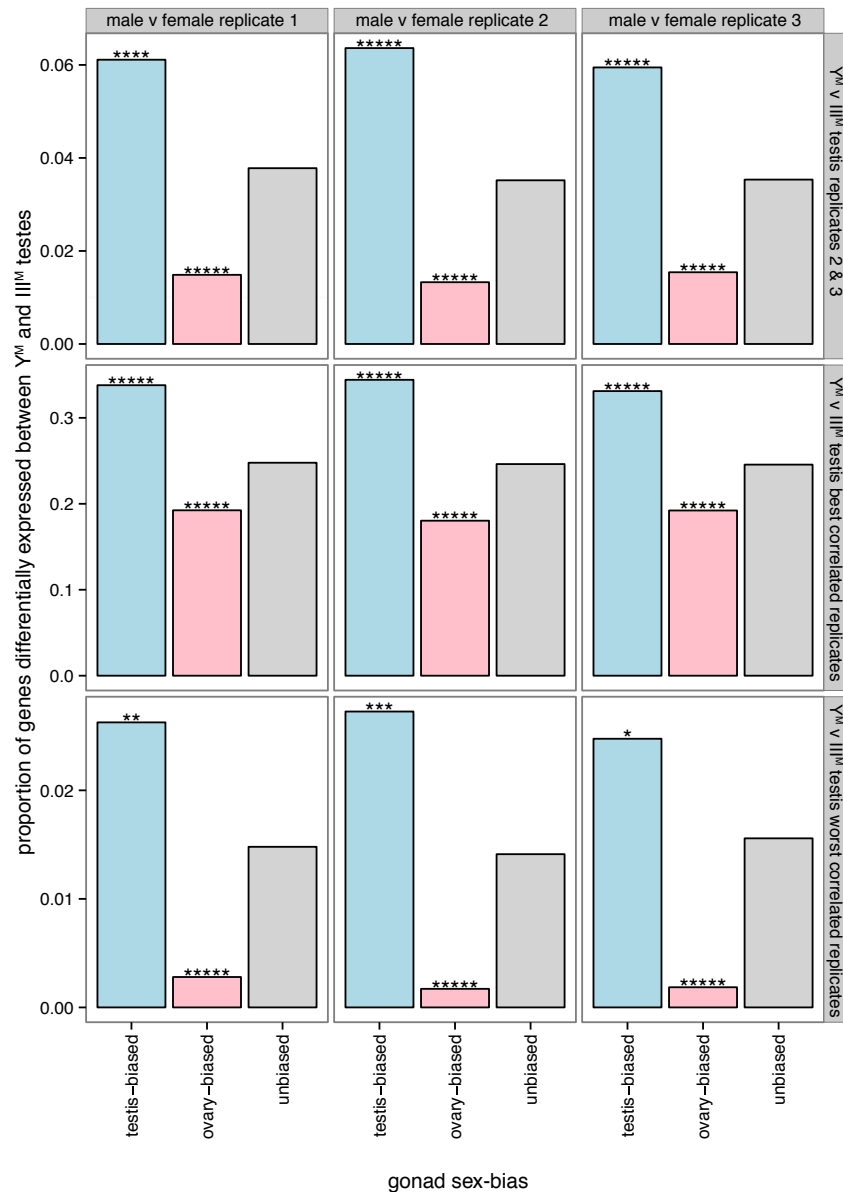
Figure S3: Chromosomal locations of genes that are differentially expressed between Y^M and III^M males. Same as Figure 2 except that differentially expressed genes are separated into those that are up-regulated in III^M (CS) males (middle), or up-regulated in Y^M (CSaY) males (bottom).
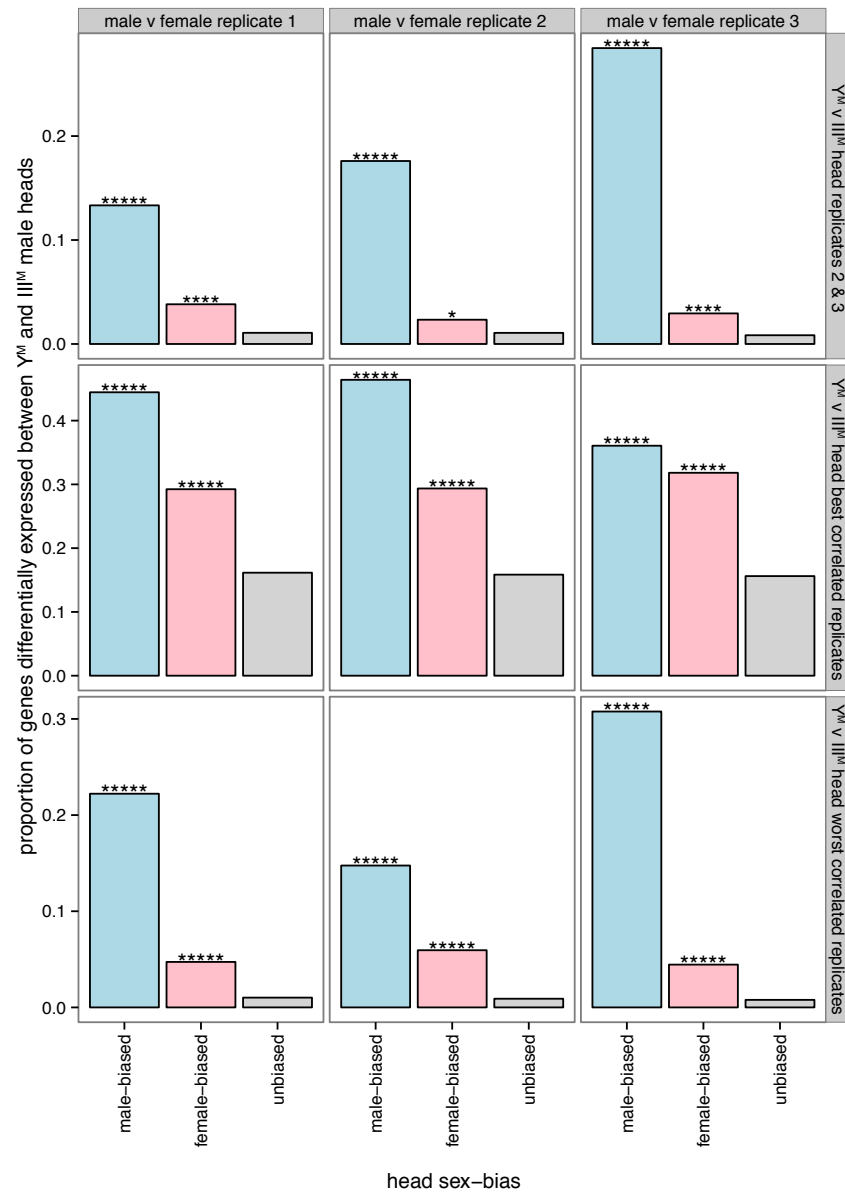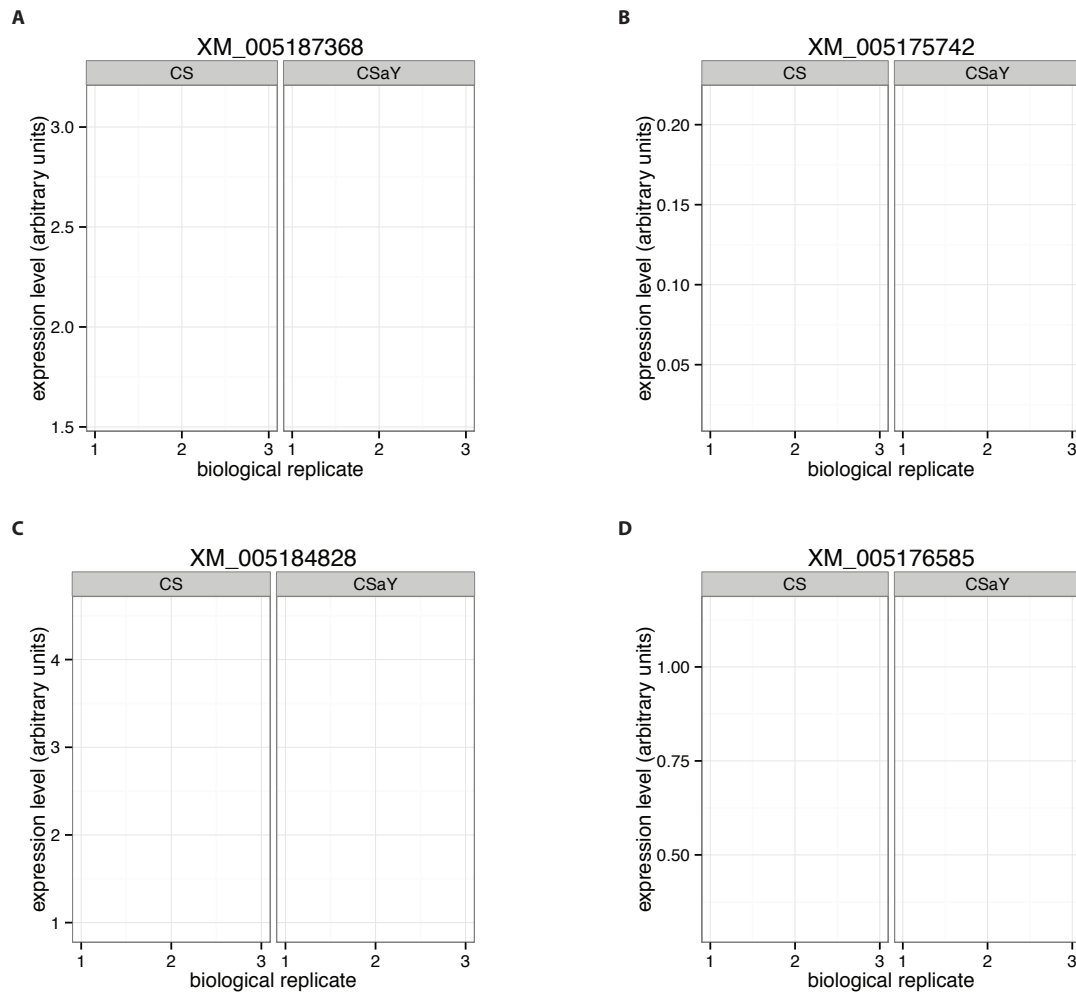
Figure S4: Sex-biased expression of genes differentially expressed between Y^M and III^M heads. Same as Figure 3 except only two replicates of each strain were used to test for differential expression between Y^M and III^M male heads and between male and female heads. Three different subsets of the data were used for the male-female comparison (listed along the top of the figure). Three different subsets were used in the Y^M *vs.* III^M comparisons: 1) a random pair of Y^M strains and a pair of III^M strains; 2) the two Y^M strains that are best correlated and the two III^M strains that are best correlated; 3) the two Y^M strains that are worst correlated and the two III^M strains that are worst correlated. Asterisks indicate significant differences between the proportion of genes with sex-biased and unbiased expression determined using Fisher's exact test (* $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$, **** $P < 0.00005$, ***** $P < 0.000005$).

35

Figure S5: Sex-biased expression of genes differentially expressed between Y^M and III^M testis. Same as Figure 3 except only two replicates of each strain were used to test for differential expression between Y^M and III^M testes and between testis and ovary. Three different subsets of the data were used for the male-female comparison (listed along the top of the figure). Three different subsets were used in the Y^M vs. III^M comparisons: 1) a random pair of Y^M strains and a pair of III^M strains; 2) the two Y^M strains that are best correlated and the two III^M strains that are best correlated; 3) the two Y^M strains that are worst correlated and the two III^M strains that are worst correlated. Asterisks indicate significant differences between the proportion of genes with sex-biased and unbiased expression determined using Fisher's exact test (* $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$, **** $P < 0.00005$, ***** $P < 0.000005$).

36

A

B

C

D

Figure S6: Estimates of expression level from the qPCR assay of four genes are plotted. Three technical replicates of three biological replicates were measured. Results of the statistical analyses are presented in the main body of the manuscript.

Supplementary Data

mRNA-Seq results

Column information:

1. **gene** - NCBI gene symbol

2. **protein_id** - NCBI protein identifier

3. **locus** - Coordinates in genomic scaffold

4. **orthocat** - Orthology category:

   - sco: single-copy ortholog
   - mto1: multiple house fly genes orthologous to 1 *D. melanogaster* gene
   - 1tom: 1 house fly gene orthologous to many *D. melanogaster* genes
   - mtom: multiple house fly genes homologous to multiply *D. melanogaster* genes
   - lr: lineage-restricted house fly gene without any identified *D. melanogaster* orthologs

5. **og** - Orthology group (gene family)

6. **dmel** - FlyBase identifier of *D. melanogaster* orthologs

7. **Dmel_ME** - Muller element location of *D. melanogaster* ortholog (only if sco)

8. **Musca_chr** - House fly chromosome homolog of Muller element

9. **scaff_ME** - Muller element assignment of house fly scaffold (based on majority rule of genes on scaffold)

10. **Musca_scaff** - House fly chromosome based on scaff_ME

11. **CS_MaleHead_FPKM** - Expression level in CS (III^M) male head

12. **CSaY_MaleHead_FPKM** - Expression level in CSaY (Y^M) male head

38

13. **MaleHead_TestStat** - Test statistic comparing CS and CSaY male head expression levels

14. **MaleHead_qval** - FDR-corrected P-value of test statistic comparing CS and CSaY male head expression levels

15. **CS_testis_FPKM** - Expression level in CS (III^M) testis

16. **CSaY_testis_FPKM** - Expression level in CSaY (Y^M) testis

17. **testis_TestStat** - Test statistic comparing CS and CSaY testis expression levels

18. **testis_qval** - FDR-corrected P-value of test statistic comparing CS and CSaY testis expression levels

19. **CS_FemaleHead_FPKM** - Expression level in CS female head

20. **CSaY_FemaleHead_FPKM** - Expression level in CSaY female head

21. **FemaleHead_TestStat** - Test statistic comparing CS and CSaY female head expression levels

22. **FemaleHead_qval** - FDR-corrected P-value of test statistic comparing CS and CSaY female head expression levels

23. **CS_ovary_FPKM** - Expression level in CS ovary

24. **CSaY_ovary_FPKM** - Expression level in CSaY ovary

25. **ovary_TestStat** - Test statistic comparing CS and CSaY ovary expression levels

26. **ovary_qval** - FDR-corrected P-value of test statistic comparing CS and CSaY ovary expression levels

27. **female_head_FPKM** - Expression level in female head

28. **male_head_FPKM** - Expression level in CS male head

29. **head_TestStat** - Test statistic comparing female and male head expression levels

39

30. **head_qval** - FDR-corrected P-value of test statistic comparing female and male head expression levels

31. **ovary_FPKM** - Expression level in ovary

32. **testis_FPKM** - Expression level in testis

33. **gonad_TestStat** - Test statistic comparing ovary and testis expression levels

34. **gonad_qval** - FDR-corrected P-value of test statistic comparing ovary and testis expression levels

Gene ontology tests

File information (test vs reference comparisons):

1. head_female-biased_GO.txt - GO categories that are enriched in genes with female-biased expression in head (foreground) vs genes with non-sex-biased expression in head

2. head_male-biased_GO.txt - GO categories that are enriched in genes with male-biased expression in head vs genes with non-sex-biased expression in head

3. male_head_diff_GO.txt - GO categories that are enriched in genes that are differentially expressed between CS and CSaY male heads vs genes that are not differentially expressed between CS and CsaY male heads

4. male_testis_diff_GO.txt - GO categories that are enriched in genes that are differentially expressed between CS and CSaY testes vs genes that are not differentially expressed between CS and CsaY testes

5. ovary-biased_unbiased_GO.txt - GO categories that are enriched in genes with ovary-biased expression vs genes with non-sex-biased expression in gonad

6. testis-biased_unbiased_GO.txt - GO categories that are enriched in genes with testis-biased expression vs genes with non-sex-biased expression in gonad

Column information:

1. GO-ID - Gene ontology ID

2. Term - GO term

3. Category - Cellular Component (C), Biological Process (P), or Molecular Function (F)

4. FDR - False discovery corrected P-value for enrichment

5. P-value - Fisher's exact test comparing #Test, #Ref, #notAnnotTest, #notAnnotRef

6. #Test - Count of genes with ontology annotation that are differentially expressed

7. #Ref - Count of genes with ontology annotation that are not differentially expressed

8. #notAnnotTest - Count of genes without ontology annotation that are differentially expressed

9. #notAnnotRef - Count of genes without ontology annotation that are not differentially expressed

10. Over/Under - Term is enriched (over) or depleted (under) in test data set

qPCR results

Raw qPCR data are presented for each of the four genes assayed (one file per gene). Each table has three columns: 1) gene name (XM_005187313 is the control gene, and the other gene is being tested for differential expression); 2) replicate (whether the data are from one of the serial dilutions used to calculate the standard curve or a biological replicate of the two strains being tested); 3) CT value of the technical replicate.

References

Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RRH, Mackay TFC. 2009. Systems genetics of complex traits in *Drosophila melanogaster*. Nat Genet 41:299–307.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy, Statist Soc Ser B 57:289–300.

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

Bopp D. 2010. About females and males: continuity and discontinuity in flies. J Genet 89:315–323.

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

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

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, Albert FW, Zeller U, Khaitovich P, Grützner F, Bergmann S, Nielsen R, Pääbo S, Kaessmann H. 2011. The evolution of gene expression levels in mammalian organs. Nature 478:343–348.

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

Bull JJ, Charnov EL. 1977. Changes in the heterogametic mechanism of sex determination. Heredity 39:1–14.

Bulmer MG, Bull JJ. 1982. Models of polygenic sex determination and sex ratio control. Evolution 36:13–26.

44

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 USA 98:13225–13230.

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 USA 97:13239–13244.

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

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

Charlesworth D, Charlesworth B. 1980. Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. Genet Res 35:205–214.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676.

Connallon T, Clark AG. 2011. Association between sex-biased gene expression and mutations with sex-specific phenotypic consequences in *Drosophila*. Genome Biol Evol 3:151–155.

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.

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.

Eberhard WG. 1985. Sexual selection and animal genitalia. Cambridge, MA: Harvard University Press.

Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. Nat Rev Genet 8:689–698.

45

Eshel I. 1975. Selection on sex-ratio and the evolution of sex-determination. Heredity 34:351–361.

Feldmeyer B, Kozielska M, Kuijper B, Weissing FJ, Beukeboom LW, Pen I. 2008. Climatic variation and the geographical distribution of sex-determining mechanisms in the housefly. Evol Ecol Res 10:797–809.

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.

Foster GG, Whitten MJ, Konovalov 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.

Franco MG, Rubini PG, Vecchi M. 1982. Sex-determinants and their distribution in various populations of *Musca domestica* L. of Western Europe. Genet Res 40:279–293.

Gepner J, Hays TS. 1993. A fertility region on the Y chromosome of *Drosophila melanogaster* encodes a dynein microtubule motor. Proc Natl Acad Sci USA 90:11132–11136.

Goldstein LS, Hardy RW, Lindsley DL. 1982. Structural genes on the Y chromosome of *Drosophila melanogaster*. Proc Natl Acad Sci USA 79:7405–7409.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucl Acids Res 36:3420–3435.

Hamm RL, Gao JR, Lin GGH, Scott JG. 2009. Selective advantage for III^M males over Y^M males in cage competition, mating competition, and pupal emergence in *Musca domestica* L. (Diptera: Muscidae). Environ Entomol 38:499–504.

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

46

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.

Hediger M, Burghardt G, Siegenthaler C, Buser N, Hilfiker-Kleiner D, Dübendorfer A, Bopp D. 2004. Sex determination in *Drosophila melanogaster* and *Musca domestica* converges at the level of the terminal regulator *doublesex*. Dev Genes Evol 214:29–42.

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

Hilfiker-Kleiner D, Dubendorfer A, Hilfiker A, Nothiger R. 1994. Genetic control of sex determination in the germ line and soma of the housefly, *Musca domestica*. Development 120:2531–2538.

Hiroyoshi T. 1961. The linkage map of the house fly, *Musca domestica* L. Genetics 46:1373–1380.

Hoyle HD, Raff EC. 1990. Two *Drosophila* beta tubulin isoforms are not functionally equivalent. J Cell Biol 111:1009–1026.

Kemphues KJ, Kaufman TC, Raff RA, Raff EC. 1982. The testis-specific β-tubulin subunit in *Drosophila melanogaster* has multiple functions in spermatogenesis. Cell 31:655–670.

Kence M, Kence A. 1992. Genetic consequences of linkage between malathion resistance and an autosomal male-determining factor in house fly (Diptera: Muscidae). J Econ Entomol 85:1566–1570.

Kerr RW. 1960. Sex-limited DDT-resistance in house-flies. Nature 185:868–868.

Kerr RW. 1961. Inheritance of DDT-resistance involving the Y-chromosome in the housefly (*Musca domestioa* L.). Aust J Biol Sci 14:605–619.

Kerr RW. 1970. Inheritance of DDT resistance in a laboratory colony of the housefly, *Musca domestica*. Aust J Biol Sci 23:377–400.

Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Paabo S. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science 309:1850–1854.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg S. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36.

Kozielska M, Feldmeyer B, Pen I, Weissing FJ, Beukeboom LW. 2008. Are autosomal sex-determining factors of the housefly (*Musca domestica*) spreading north? Genet Res 90:157–165.

Kozielska M, Pen I, Beukeboom LW, Weissing FJ. 2006. Sex ratio selection and multi-factorial sex determination in the housefly: a dynamic model. J Evol Biol 19:879–888.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25.

Lemos B, Araripe LO, Hartl DL. 2008. Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. Science 319:91–93.

Lemos B, Branco AT, Hartl DL. 2010. Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. Proc Natl Acad Sci USA 107:15826–15831.

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. 2003. Rapid evolution of male-biased gene expression in *Drosophila*. Proc Natl Acad Sci USA 100:9894–9899.

Pomiankowski A, Nöthiger R, Wilkins A. 2004. The evolution of the *Drosophila* sex-determination pathway. Genetics 166:1761–1773.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. Evolution 38:735–742.

Rice WR. 1986. On the instability of polygenic sex determination: the effect of sex-specific selection. Evolution 40:633–639.

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

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

Roberts RB, Ser JR, Kocher TD. 2009. Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. Science 326:998–1001.

Sackton TB, Montenegro H, Hartl DL, Lemos B. 2011. Interspecific Y chromosome introgressions disrupt testis-specific gene expression and male reproductive phenotypes in *Drosophila*. Proc Natl Acad Sci USA 108:17046–17051.

Schmidt R, Hediger M, Roth S, Nothiger R, Dubendorfer A. 1997. The Y-chromosomal and autosomal male-determining *M* factors of *Musca domestica* are equivalent. Genetics 147:271–280.

Scott J, Warren W, Beukeboom L, Bopp D, Clark A, Giers S, Hediger M, Jones A, Kasai S, Leichter C, Li M, Meisel R, Minx P, Murphy T, Nelson D, Reid W, Rinkevich F, Robertson H, Sackton T, Sattelle D, Thibaud-Nissen F, Tomlinson C, van de Zande L, Walden K, Wilson R, Liu N. 2014. Genome of the house fly, *Musca domestica* L., a global vector of diseases with adaptations to a septic environment. Genome Biol 15:466.

Scott JG. 2008. Insect cytochrome P450s: thinking beyond detoxification. In: Liu N, editor. Recent Advances in Insect Physiology, Toxicology and Molecular Biology. Kerala, India: Research Signpost. p. 117–124.

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

Shono T, Scott JG. 1990. Autosomal sex-associated pyrethroid resistance in a strain of house fly (Diptera: Muscidae) with a male-determining factor on chromosome three. J Econ Entomol 83:686–689.

Siegenthaler C, Maroy P, Hediger M, Dubendorfer A, Bopp D. 2009. Hormones and sex-specific transcription factors jointly control yolk protein synthesis in *Musca domestica*. Int J Evol Biol 2009:291236.

Tomita T, Wada Y. 1989*a*. Migration and linkage disequilibrium in local populations of the housefly (*Musca domestica*) in Japan. Jpn J Genet 64:383–389.

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

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotech 31:46–53.

van Doorn GS. 2014. Evolutionary transitions between sex-determining mechanisms: a review of theory. Sex Dev 8:7–19.

van Doorn GS, Kirkpatrick M. 2007. Turnover of sex chromosomes induced by sexual conflict. Nature 449:909–912.

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

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.

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

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

Werren JH, Beukeboom LW. 1998. Sex determination, sex ratios, and genetic conflict. Ann Rev Ecol Sys 29:233–261.

Wilkins AS. 1995. Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. Bioessays 17:71–77.

Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B. 2007. Constraint and turnover in sex-biased gene expression in the genus *Drosophila*. Nature 450:233–237.

Zhou J, Sackton TB, Martinsen L, Lemos B, Eickbush TH, Hartl DL. 2012. Y chromosome mediates ribosomal DNA silencing and modulates the chromatin state in *Drosophila*. Proc Natl Acad Sci USA 109:9941–9946.