Supplementary Figures and Tables

for

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

Richard P. Meisel[1]*, Christopher A. Gonzales[1,2], and Hoang Luu[1]

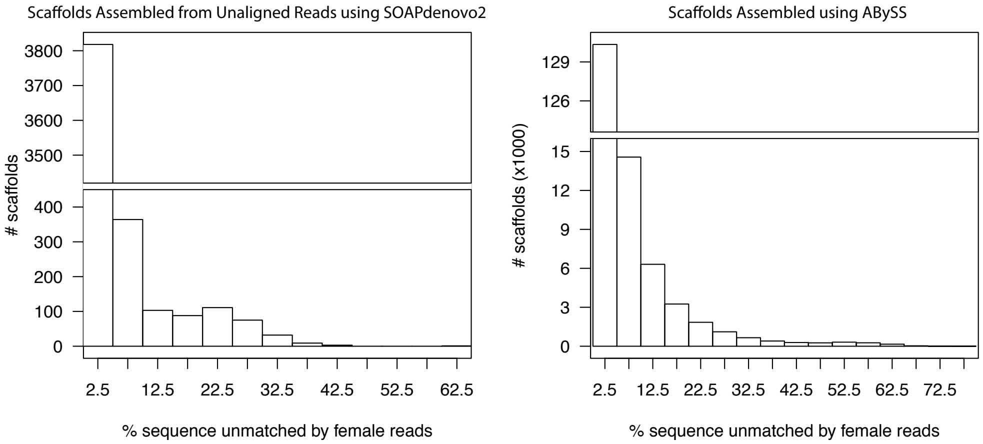
1. Department of Biology and Biochemistry, University of Houston
2. Current address: Department of Pharmacology, University of Texas Southwestern Medical Center

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

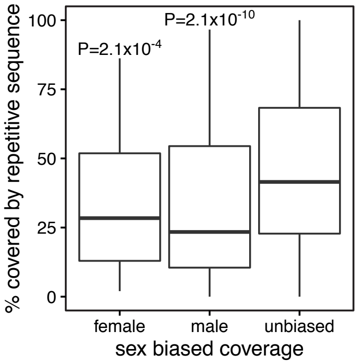
Running title: House fly sex chromosomes

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

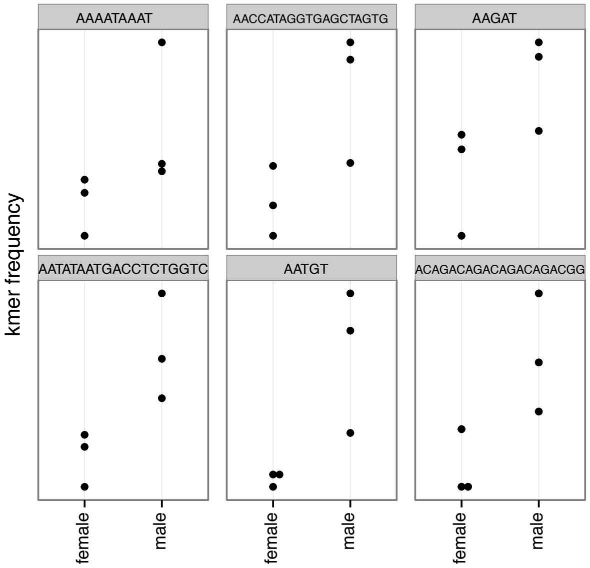
June 7, 2017



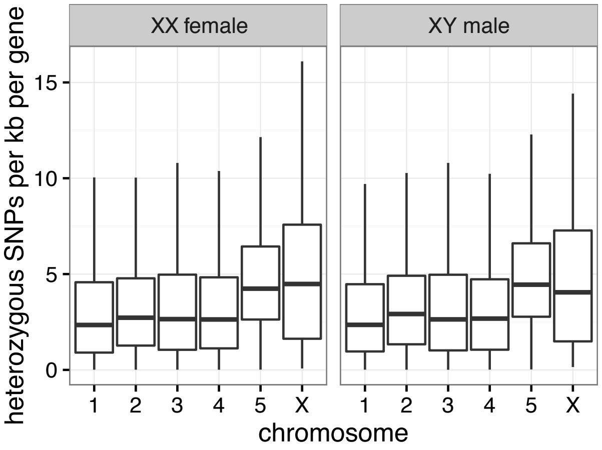
Supplemental Fig S1: Histograms of female read mapping coverage to scaffolds from male genomes assembled using SOAPdenovo2 with reads that did not align to the female reference genome (left) or using ABySS (right).



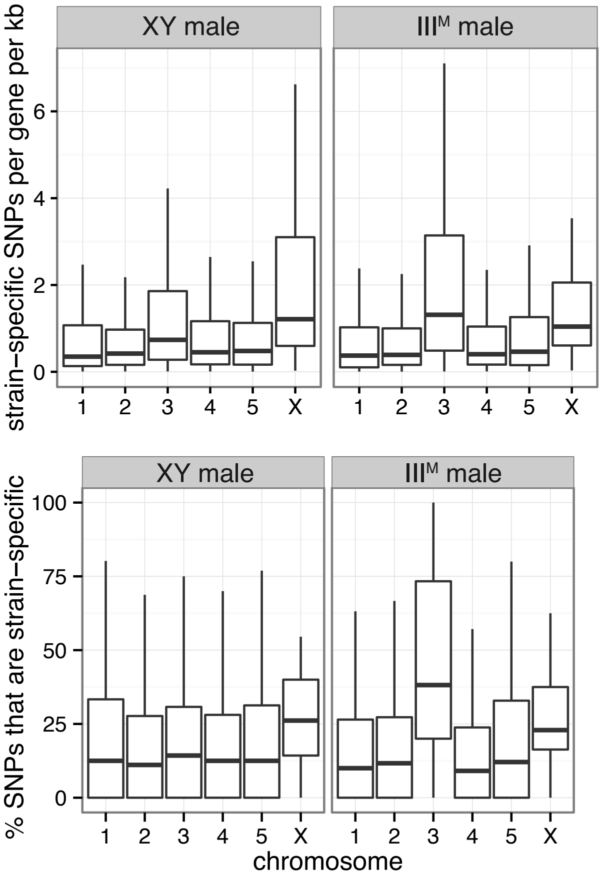
Supplemental Fig S2: Boxplots show the distribution of the percent of 1 kb windows that contain predicted repetitive sequence. Three different types of 1 kb windows are plotted: those with female-biased read mapping coverage ($\log_2 \frac{M}{F} < -1$), those with male-biased coverage ($\log_2 \frac{M}{F} > 1$), and those with insignificant differences in coverage (unbiased). P values comparing the female- and male-biased windows with the unbiased windows from a Mann-Whitney test are shown.



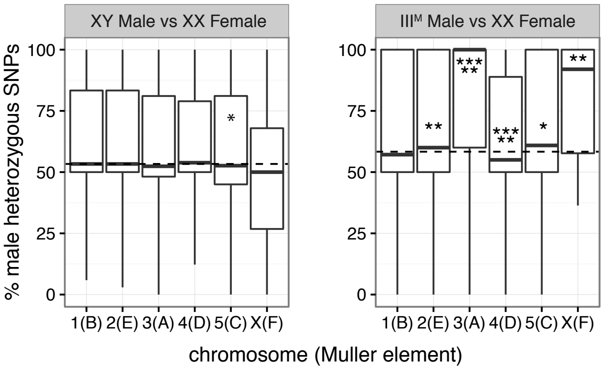
Supplemental Fig S3: The frequency of the six k-mers over-represented in males is plotted for each of the 3 female and 3 male libraries.



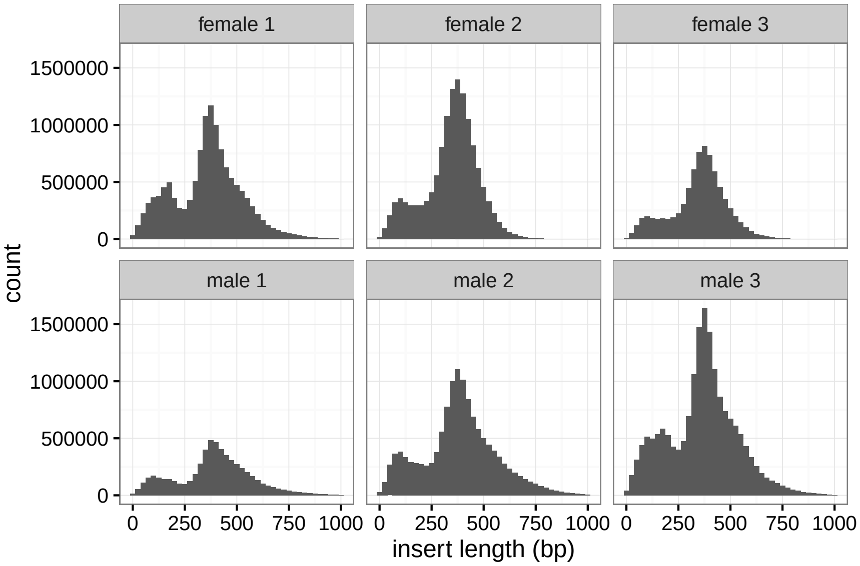
Supplemental Fig S4: Boxplots show the distributions of heterozygous SNPs per kb per gene for each chromosome from the re-sequencing of aabys females and males.



Supplemental Fig S5: Boxplots show the distributions of strain-specific SNPs per gene per kb (top) and the percent of SNPs that are strain-specific per gene (bottom) for each chromosome from RNA-seq data collected in XY and III^M males.



Supplemental Fig S6: Box plots show the distributions of the percent of heterozygous SNPs within genes on each chromosome in males relative to females from strains with XY males (left) or III$^{\mathrm{M}}$ males (right). The median across all autosomes is indicated by a dashed line. Asterisks indicate significant differences between a chromosome and all other autosomes in a Mann-Whitney test ($*P < 0.05$, $**P < 0.005$, $*****P < 0.000005$).



Supplemental Fig S7: Distributions of insert sizes for the 6 male and female aabys libraries.

strain	library	mapped	mapped_single	mapped_distant	unmapped
aabys	female1-26111291	13049678	887676	6292401	501177
aabys	female2-26123107	13712814	809792	7125338	427100
aabys	female3-26124109	7972753	550477	4716107	250508
aabys	male1-26127107	6081087	656412	4333290	221249
aabys	male2-26119223	13559671	1177151	8930649	514483
aabys	male3-26113283	18323724	1275507	9500131	563536
A3	1_A3_female_S10	21641219	2157816	5795725	1116445
A3	2_A3_female_S7	22594865	2070878	6009122	1099204
A3	3_A3_female_S6	28633153	2966799	8121340	1412310
A3	4_A3_male_S9	22766977	2407270	6228309	1206439
A3	5_A3_male_S5	21817355	2136473	5675312	1140114
A3	6_A3_male_S4	19437196	1902804	4873879	1077971
LPR	7_LPR_female_S3	16238580	2325002	5586155	1062667
LPR	8_LPR_female_S12	16779395	2388949	5921294	1174171
LPR	9_LPR_female_S2	19773472	3284957	7356944	1317193
LPR	10_LPR_male_S8	15172543	2379298	5457376	1070268
LPR	11_LPR_male_S11	14438841	2420096	5374716	1041780
LPR	12_LPR_male_S1	14150680	2252137	5148394	1023886

Supplemental Table S1: Read mapping statistics for aabys, A3, and LPR male and female sequencing libraries. The number of paired reads that map to the same scaffold (mapped), where only one of two reads in a pair mapped (mapped_single), where two reads map to different scaffolds (mapped_distant), and where both reads failed to map (unmapped) are shown.

Assembler	Reads	N50	Scaffolds/Contigs	Total Length
SOAPdenovo2	All	2491	231066	522258915
SOAPdenovo2	MaleUnaligned	1403	4604	6769258
ABySS	All	1936	159804	300356116

Supplemental Table S2: Assembly statistics for male genomes are shown for the three approaches: SOAPdenovo2 with all reads, SOAPdenovo2 with only reads that did not align to the reference genome, and ABySS. The statistics are scaffold/contig N50, the number of scaffolds/contigs in the assembly, and the total length of the assembled genome. Only scaffolds/contigs that are at least 1 kb long were included in the statistics.