Trans regulation of an odorant binding protein by a proto-Y chromosome affects male courtship in house fly

Pablo J Delclos[†,1,2], Kiran Adhikari[†,1,3], Alexander B Mai[1], Oluwatomi Hassan[1], Alexander A Oderhowho[1,4], Vyshnika Sriskantharajah[1,5], Tammie Trinh[1,6], Richard P Meisel[1]

[†]Co-first author/equal contribution

Affiliation: [1]Department of Biology & Biochemistry, University of Houston, Houston, TX, USA

[2]Current affiliation: Department of Natural Sciences, University of Houston-Downtown, Houston, TX, USA

[3]Current affiliation: Oxford Biomedica Solutions LLC, Bedford, MA

[4]Current affiliation: Department of Neurosurgery, Ochsner LSU Health Shreveport, Shreveport, LA, USA

[5]Current affiliation: University of Texas Health Science Center School of Biomedical Informatics, Houston, TX, USA

[6]Current affiliation: University of Texas School of Dentistry, Houston, TX, USA

ABSTRACT

Y chromosomes have male-limited inheritance, which favors the fixation of alleles that affect spermatogenesis, courtship, and other male-specific traits. Y-linked male-beneficial alleles can also have female-deleterious (sexually antagonistic) effects because they never experience direct selection in females. However, determining the mechanisms underlying these male-beneficial effects is challenging because it can require studying Y-linked alleles while they still segregate as polymorphism. We used a Y chromosome polymorphism in the house fly, *Musca domestica*, to address this challenge. Two common male-determining Y chromosomes (Y^M and III^M) segregate as stable polymorphisms in natural house fly populations, and they differentially affect multiple traits, including male courtship performance. We performed a meta-analysis of RNA-seq data and identified differentially expressed genes encoding odorant binding proteins (in the *Obp56h* family) as candidate causal agents in the courtship differences. The *Obp56h* genes are not found on either the Y^M or III^M chromosomes, suggesting that they must be regulated in *trans* by one of the house fly sex chromosomes. Using a network analysis and allele-specific expression measurements, we identified multiple genes on the house fly III^M chromosome that could serve as *trans* inhibitors of *Obp56h* gene expression. One of those genes is homologous to *D. melanogaster CG2120*, which encodes a transcription factor that binds both up- and down-stream of *Obp56h*. We found that up-regulation of *CG2120* in *D. melanogaster* nervous tissues reduces copulation latency, consistent with this transcription factor acting as a negative regulator of *Obp56h* expression. We propose the name *speed date* (*spdt*) for *CG2120*, with the house fly homolog named *Md-spdt*. The expression of *spdt* across *D. melanogaster* development and tissues suggests that evolution of higher expression in neurons may be constrained by pleiotropic or sexual antagonistic effects. We hypothesize that a *cis*-regulatory allele that increases expression of *Md-spdt* on the III^M chromosome exists because Y-linkage of this allele releases it from those constraints. This provides evidence for a molecular mechanism by which a Y-linked gene can evolve a male-beneficial function regardless of the negative effects on females.

INTRODUCTION

In species with genetic sex determination, a Y or W chromosome can have sex-limited inheritance (Bachtrog et al. 2014; Beukeboom and Perrin 2014). The male-limited inheritance of Y chromosomes is predicted to allow for the accumulation of alleles with male-specific beneficial effects (Rice 1996). These male-beneficial alleles can have female-deleterious (sexually antagonistic) effects because they are never exposed to direct selection in females (Charlesworth et al. 2005; Abbott et al. 2017).

One important way that sex-limited Y and W chromosomes appear to affect sex-specific traits is via *trans* regulation of genes elsewhere in the genome. For example, Y chromosome genotypes in *Drosophila melanogaster* have *trans* effects on gene expression throughout the genome, which modify a broad range of phenotypes, including immunity and chromatin (Lemos et al. 2008, 2010; Brown et al. 2020). *D. melanogaster* Y chromosome genotypes also have fitness consequences that depend on the genetic background, suggesting epistatic interactions between Y-linked alleles and the X or autosomes (Chippindale and Rice 2001). Similarly, in *Poecilia spp.*, Y-linked alleles may affect sexually selected male pigmentation patterns by acting as *trans* regulators of autosomal gene expression (Morris et al. 2020; Kawamoto et al. 2021; Sandkam et al. 2021), although the specific mechanisms of these effects are not well understood.

Even though we are aware of sex-specific phenotypic and fitness effects of Y and W chromosomes, the mechanisms underlying these effects are not as well understood. Notably, we have a limited understanding of how Y and W chromosomes act as *trans* regulators of sex-specific and sexually antagonistic traits genome-wide. There are also very few sexually antagonistic alleles that have been genetically or molecular characterized on young sex chromosomes (cf. Roberts et al. 2009), which limits our ability to make generalizations about the molecular mechanisms underlying sexually antagonistic selection on Y and W chromosomes (Mank et al. 2014). To address these shortcomings, we sought to identify the genetic mechanism underlying how a young Y chromosome affects male courtship behavior in the house fly, *Musca domestica*.

House fly has a multifactorial sex determination system, in which multiple young proto-Y and proto-W chromosomes segregate as polymorphisms in natural populations (Hamm et al. 2015). The two most common proto-Y chromosomes (III^M and Y^M) are distributed along latitudinal clines on multiple continents (Franco et al. 1982; Tomita and Wada 1989; Hamm et al. 2005), and they affect thermal traits in ways that are consistent with their geographic distributions (Delclos et al. 2021). This polymorphism has remained stable in natural populations since at least the mid-20th century, suggesting that selection maintains multifactorial sex determination (Kozielska et al. 2008; Meisel et al. 2016).

82 Here, we focus on the effect of the house fly proto-Y chromosomes on male courtship
83 performance. Males carrying the III^M chromosome (III^M males) outcompete males carrying the
84 Y^M chromosome (Y^M males) for female mates (Hamm et al. 2009). In light of this and other
85 aforementioned phenotypic differences between Y^M and III^M males, it is remarkable that the III^M
86 and Y^M chromosomes carry nearly all of the same genes as their homologous proto-X
87 chromosomes (III and X), and only a small number of allelic differences have been identified
88 (Meisel et al. 2017; Son and Meisel 2021). These similarities between the proto-Y and proto-X
89 chromosomes has led to the hypothesis that the phenotypic effects of the proto-Y chromosomes
90 may be mediated by *trans* effects of Y^M and III^M alleles on the expression of genes elsewhere in
91 the genome (Adhikari et al. 2021). We aimed to identify the *trans* regulatory allele(s) on the Y^M
92 or III^M chromosomes that affect differences in courtship performance between the genotypes.

93 **METHODS**

94 ***RNA-seq differential gene expression analysis***

95 We analyzed published RNA-seq data from *M. domestica* male heads (NCBI Gene Expression
96 Omnibus accessions GSE67065, GSE126685, and GSE126689, shown in Supplementary Table
97 S1). The RNA-seq data include nine Y^M and fifteen III^M samples (Meisel et al. 2015; Son et al.
98 2019). We assigned RNA-Seq reads to house fly transcripts from genome assembly v2.0.2 and
99 annotation release 102 (Scott et al. 2014) using kallisto in single-end read mode (Bray et al.
100 2016). All RNA-seq reads were single-end, and we set the average fragment length to 300 bp and
101 standard deviation to 30 bp for all samples. Our expression analysis was performed at the level
102 of transcripts (as opposed to genes), which should not affect our conclusions because all of our
103 focal genes each produce one annotated transcript.

104 We tested for differentially expressed (DE) transcripts between males with a Y^M chromosome
105 and males with a III^M chromosome using a combination of DESeq2 (Love et al. 2014), sva (Leek
106 et al. 2012), and limma (Ritchie et al. 2015). We only included transcripts that passed an initial
107 threshold filter of 0.5 counts per million in at least 4 samples. Read counts for each of those
108 transcripts were normalized by variance stabilizing transformation in DESeq2. To remove batch
109 effects across data sets, we used the sva package to identify and estimate surrogate variables that
110 adjust for latent sources of variation (e.g., batch effects). To identify DE transcripts between Y^M
111 and III^M males, we used the lmFit() function in limma to fit a linear model comprised of male
112 type (Y^M vs III^M) and our surrogate variables as fixed effects, and read counts as the response
113 variable. We then computed contrasts between male types and calculated test statistics using the
114 eBayes() function. Transcripts below a false discovery rate (FDR) adjusted p value (p_{ADJ}) of 0.05
115 were categorized as DE (Benjamini and Hochberg 1995).

116 ***Weighted gene co-expression network analysis***

117 We used weighted gene co-expression network analysis (WGCNA) to identify modules of house
118 fly transcripts whose expression correlates with male type (Y^M or III^M) on normalized read count
119 data that were adjusted for batch effects in sva (Langfelder and Horvath 2008). For all pairs of
120 transcripts with variable expression across samples, we calculated Pearson's correlation
121 coefficient across all samples. We created an unsigned correlation matrix and adjusted the soft-
122 threshold value (β) to which among-transcript covariances are exponentially raised to ensure a
123 scale-free topology (this resulted in $\beta = 7$), thereby creating a weighted network of gene
124 expression. An unsigned matrix allows us to identify connected transcripts whose expression is
125 either positively or negatively correlated. Within this topological overlapping network (Li and
126 Horvath 2007), transcripts were hierarchically clustered, and modules were identified based on
127 the degree of similarity among transcripts. We used a merging threshold of 0.2, with a minimum
128 module size of 30 and a mean connectivity threshold of greater than or equal to 0.7. We used the
129 default parameters of WGCNA for the rest of the analyses. We then correlated module eigengene
130 values for a given module across samples via Pearson's correlation and identified modules
131 differentially regulated between male types at FDR-adjusted $p < 0.05$.

132 To visualize WGCNA genetic covariance results among modules significantly associated with
133 male type, we exported final co-expression networks to Cytoscape (Shannon et al. 2003). We
134 attached information on $\log_2$ fold-change in expression between III^M and Y^M males, as well as
135 chromosomal location, to the network as metadata so this information could be visualized. To
136 identify transcripts that may have more central functions within and across our significant
137 modules, we ranked transcripts in descending order based on intramodular connectivity
138 (kWithin), calculated in WGCNA. Hub genes identified by intramodular connectivity are
139 generally functionally important transcripts within a module (Langfelder et al. 2013).

140 We further analyzed among-transcript connections involving a family of odorant-binding protein
141 genes (*Obp56h*). Specifically, to identify transcripts that may regulate or be regulated by genes
142 within the family, we calculated a "connection score" C_i for every transcript i as follows:

143 $$C_i = \sum_{j=1}^{n} a_{ij} |F_j|,$$

144 where a_{ij} represents the adjacency (the Pearson correlation coefficient raised to the soft-
145 threshold power β) between transcript i and *Obp56h* gene j, and F_j represents the $\log_2$ fold-
146 change in expression between III^M and Y^M males for *Obp56h* gene j. This weighted product
147 ensured that connections with *Obp56h* genes that are more differentially expressed between male
148 types were prioritized in calculating a transcript's connection score. Transcripts were then ranked
149 by C_i to identify candidate genes that may be strongly tied to *Obp56h* expression. Transcripts
150 with the 100 highest *Obp56h* connection scores were classified as "central genes". We tested for
151 chromosomal enrichment among these central genes using Fisher's exact tests (comparing the

152 number of central and non-central genes on a focal chromosome with the number of central and
153 non-central genes on all other chromosomes) to determine whether the expression of *Obp56h*
154 genes (which are all located on the *M. domestica* chromosome V) might be involved in *trans*
155 regulation with genes located on the III^M proto-Y chromosome.

156 ***Gene ontology enrichment analysis***

157 To identify gene ontology (GO) classes and molecular pathways that are enriched among DE
158 transcripts, across co-expression modules identified in WGCNA, or among central genes co-
159 expressed with *Obp56h* genes, we used the BiNGO plug-in within Cytoscape (Maere et al.
160 2005). We identified *D. melanogaster* orthologs for each house fly gene within a given gene list
161 via NCBI blastx best hits (with default parameters) and used the *D. melanogaster* gene name as
162 input (Adhikari et al. 2021). We identified GO terms that are significantly enriched in BiNGO
163 for biological processes, cellular components, and molecular function.

164 ***Allele-specific expression analysis***

165 We tested for differential expression of house fly chromosome III genes between the allele on
166 the III^M chromosome and the allele on the standard third (III) chromosome in III^M males. To do
167 so, we followed methods as in previous studies (Meisel et al. 2017; Son and Meisel 2021), which
168 used the Genome Analysis Toolkit (GATK) best practices workflow for single nucleotide
169 polymorphism (SNP) calling to identify sequence variants in our RNA-Seq data (McKenna et al.
170 2010). We focused our analysis on libraries that were sequenced from head tissue of male house
171 flies that comprise a CS genetic background (Meisel et al. 2015; Son et al. 2019; Adhikari et al.
172 2021). We used STAR (Dobin et al. 2013) to align reads from a total of 30 head libraries (15 III^M
173 and 15 Y^M libraries) to the house fly reference genome (Musca_domestica-2.0.2). We then
174 followed the same methods and applied the same parameters as we have done previously to
175 identify SNPs and genotype individual strains (Meisel et al. 2017; Son and Meisel 2021). We
176 performed separate joint genotyping for each house fly strain within a given experiment (a total
177 of 4 III^M and 4 Y^M strain-by-experimental batch combinations).

178 We use the following approach to differentiate between III^M and standard chromosome III
179 alleles. We first identified SNPs in the exonic regions of the top "hub" genes within a WGCNA
180 module that mapped to house fly chromosome III. We selected SNPs in those genes that are
181 heterozygous in III^M males and homozygous in Y^M males. We used the genotype of these SNPs
182 in Y^M males (which possess two standard third chromosome alleles) to determine the standard
183 chromosome III allele. The allele not found in Y^M genotypes was assigned to the III^M
184 chromosome. We also identified positions where III^M males appear monoallelic for an allele not
185 found in Y^M males. These positions that exhibit a complete bias for a III^M allele are suggestive of
186 monoallelic expression of the III^M allele (i.e., no expression from the III allele).

187 We tested for differences in expression of III^M and standard chromosome III alleles by following
188 best practices for comparing allele-specific expression (Castel et al. 2015). First, for each strain-
189 by-experimental batch combination, we calculated the normalized read depth at each variable
190 site as the number of mapped reads at that site divided by the total number of mapped reads
191 throughout the genome. At each variable site, we used Wilcoxon rank sum tests to make three
192 different pairwise comparisons per site. First, we compared normalized read depths between III^M
193 and III alleles in III^M males (III^M-III). Second, we compared the read depths of the III^M allele in
194 III^M males with the normalized read depth of both III alleles in Y^M males (III^M-Y^M). Third, we
195 compared the read depths of the III allele in III^M males with the normalized read depth of both III
196 alleles in Y^M males (III-Y^M). We set a threshold of significance at $p < 0.05$ for all comparisons.

197 **Drosophila melanogaster *RNA-seq and microarray data analysis***

198 We analyzed RNA-seq results reported in a previous study (Shorter et al. 2016) to determine
199 how knockdown of *Obp56h* affects gene expression in *D. melanogaster*. Shorter et al. (2016)
200 identified DE genes between *Obp56h* knockdown and control samples. This analysis was done
201 separately in males and females, and in separate tissue samples within a given sex (head or the
202 remaining body). We conducted GO enrichment analysis, as described above, on the list of DE
203 genes in *D. melanogaster* male head tissue upon *Obp56h* knockdown.

204 We tested if an excess of DE genes (between *Obp56h* knockdown and controls) are found on the
205 *D. melanogaster* X chromosome, which is homologous to house fly chromosome III (Foster et
206 al. 1981; Weller and Foster 1993). This chromosome is known as Muller element A across flies
207 (Meisel and Scott 2018; Schaeffer 2018). *Obp56h* is located on *D. melanogaster* chromosome
208 2R (Muller element C), which is homologous to house fly chromosome V. We used Fisher's
209 exact tests (comparing the number of X and non-X chromosome genes that are DE in a given
210 tissue within a given sex with the number of X and non-X chromosome genes that are not DE) to
211 determine whether *Obp56h* knockdown in *D. melanogaster* results in the disproportionate
212 differential expression of X chromosome genes in male heads, male bodies, female heads, or
213 female bodies.

214 We also tested if the same genes are DE between III^M vs Y^M house flies and *Obp56h* knockdown
215 vs control *D. melanogaster*. Using NCBI blastx best hits, we identified 20 *M. domestica*
216 transcripts that are orthologous to *D. melanogaster* genes that are DE upon knockdown of
217 *Obp56h* (11 matches to upregulated *D. melanogaster* genes, and 9 matches to downregulated *D.*
218 *melanogaster* genes). This list represents 40% of the 50 *D. melanogaster* DE genes. We
219 compared the mean $\log_2$ fold-changes between Y^M and III^M house fly males for those 20 genes to
220 10,000 random subsets of $\log_2$ fold-change values taken from our data (10,000 subsamples
221 without replacement of 11 genes to test for an excess of positive $\log_2$ fold-change values, and
222 10,000 subsamples of 9 genes to test for an excess of negative $\log_2$ fold-change values; see
223 Additional Files for R script). We assessed significance by calculating the proportion of

224 replicated subsamples that generated a mean $\log_2$ fold-change value more extreme than our
225 observed mean.

226 We additionally obtained gene expression measurements from microarray and RNA-seq data
227 collected from various larval, pupal, and adult *D. melanogaster* tissue samples (Supplementary
228 Table S2). Microarray data were sampled from FlyAtlas expression measurements of larval
229 central nervous system, adult head, adult eye, and adult brain (Chintapalli et al. 2007). RNA-seq
230 data were sampled from larval, pupal, or adult antennae (Shiao et al. 2013; Menuz et al. 2014;
231 Pan et al. 2017; Mohapatra and Menuz 2019) and adult proboscis (May et al. 2019). We only
232 included samples from wild-type flies, and we used the published expression estimates (e.g.,
233 microarray signal intensity, transcripts per million, reads per kilobase per million mapped reads)
234 measured at either the gene or transcript level. When data from multiple replicates were
235 available, we calculated the mean expression level across all replicates for each gene or
236 transcript. The range and distributions of expression levels varied across tissue samples because
237 the data were collected using different methodologies. To compare across tissue samples, we
238 calculated a normalized expression level for each gene or transcript in each tissue sample by
239 dividing by the mean value across all genes or transcripts in that tissue sample. Any gene or
240 transcript with an expression value of 0, detected in <4 replicates of the microarray data, or that
241 failed to pass threshold in an RNA-seq data set (i.e., status not OK) was excluded. We further
242 extracted expression measurements from genes identified as hubs within the WGCNA co-
243 expression network (see above). If a hub gene had multiple annotated transcripts, and expression
244 was measured for transcripts, we calculated the gene expression level as the mean across all
245 annotated transcripts.

246 We obtained measurements of the expression level of *CG2120* from FlyAtlas microarray and
247 modENCODE RNA-seq data sampled at multiple developmental stages and across tissues
248 (Chintapalli et al. 2007; Graveley et al. 2011; Brown et al. 2014). The FlyAtlas microarray data
249 are the average signal intensity across all probes for the gene, and the modENCODE data are
250 reported as reads per kilobase per million mapped reads (RPKM), both of which were obtained
251 from FlyBase (Gelbart and Emmert 2013).

252 ***Competitive courtship assays***

253 We performed competitive courtship experiments in which two different house fly males were
254 combined with a single female, and we recorded the "winning" male (i.e., the one who mated
255 with the female), similar to what was done previously (Hamm et al. 2009). In these experiments,
256 we used the same two house fly strains as in Hamm et al. (2009): a III^M strain called CS and a
257 Y^M strain called IsoCS. These two strains have a common genetic background (CS), and only
258 differ in which proto-Y chromosome they carry. Both strains are represented in the RNA-seq
259 data we analyzed (Meisel et al. 2015; Son et al. 2019), and IsoCS was also included in a previous
260 RNA-seq study comparing the effects of proto-Y chromosome and temperature on gene

261 expression (Adhikari et al. 2021). Our experiment differed from previous work because we
262 reared larvae from each strain at either 18°C and 29°C, whereas Hamm et al. (2009) worked with
263 flies raised at 28°C. We used the same larval wheat bran diet as done previously, and we fed
264 adults an *ad libitum* supply of water and an *ad libitum* 1:1 mixture of dry-milk:sugar. This is also
265 the same diet and rearing protocol used for the flies in the RNA-seq datasets that we analyzed
266 (Meisel et al. 2015; Son et al. 2019). Male flies were isolated from females within ~1 hour of
267 eclosion, and each sex was kept separately to ensure that flies had not mated prior to the
268 experiment.

269 We carried out two distinct competitive courtship experiments: 1) inter-strain competition
270 between males with different genotypes (i.e., Y^M vs III^M) that were reared at the same
271 temperature (363 successful mating trials out of 490 total attempts across 27 experimental
272 batches); and 2) intra-strain competition between males with the same genotype that were reared
273 at different temperatures (104 successful mating trials out of 129 total attempts across 7 batches).
274 When we competed flies with different genotypes raised at the same temperature, all males were
275 aged 4-6 d post pupal emergence. When we compared flies with the same genotype raised at
276 different temperatures, 29°C males were aged 4-5 d post emergence and 18°C males were aged
277 6-7 d post emergence. We aged flies from the colder temperature for more days than flies from
278 the warmer temperature because developmental rate is positively correlated with developmental
279 temperature in flies (Atkinson 1996). The ages we selected ensure that all males were
280 physiologically capable of mating, while also sampling flies at similar physiological ages across
281 experiments. Aging calculations are reported in Supporting Information 1.

282 The two males in each experiment were labeled using red and blue luminous powder (BioQuip)
283 by shaking the flies in an 8 oz paper cup. The color assigned to males was switched in each
284 successive batch (i.e., blue Y^M and red III^M in one batch, and then red Y^M and blue III^M in the
285 next batch). In addition, we included the genotype or developmental temperature of the blue-
286 colored male as a fixed effect in our statistical analysis (see below), which provides an additional
287 control for color.

288 For each replicate of the competitive courtship assay, we placed the two different males in a 32
289 oz transparent plastic container, along with a single virgin female. Each plastic container also
290 contained a 1:1 mixture of dry milk:sugar in a 1 oz paper cup and water in a glass scintillation
291 vial plugged with a cotton roll. Virgin females from the LPR strain (Scott et al. 1996) raised at
292 25°C were used for all combinations of males. The LPR strain has a different genetic background
293 than the males used in the assay, minimizing any effects of co-adaptation between females and a
294 particular subset of males. All flies were transferred into the mating containers using an aspirator
295 and without anesthesia. All matings were performed in a 25°C incubator because copulation
296 latency is too long for experimentally tractable measurement at lower temperatures. The color
297 (i.e., genotype) of the first male to mate was recorded, as well as the time to mate.

298 We used the glmer() function in the lme4 package in R (Bates et al. 2015) to test for the effects
299 of genotype and temperature on male mating success. First, to test the effect of genotype, we
300 constructed a logistic regression model as follows:

301 $$W \sim G_B + T + G_B \times T + b,$$

302 with developmental temperature (T), genotype of the blue male (G_B), and their interaction as
303 fixed effects. Experimental batch (b) was modeled as a random effect, with the winning male
304 (W: CS or IsoCS) as a response variable. We then assessed significance of fixed effects (type II
305 sum of squares) using the Anova() function in the car package in R (Fox et al. 2013). To test for
306 the effect of temperature on mating success, we similarly constructed a logistic regression model
307 as follows:

308 $$W \sim G + T_B + G \times T_B + b,$$

309 with genotype (G), developmental temperature of the blue male (T_B), and their interaction as
310 fixed effects, experimental batch (b) as a random effect, and the winning male (W: 18°C or
311 29°C) as a response variable. We then assessed significance of fixed effects (type II sum of
312 squares) using the Anova() function in the car package in R (Fox et al. 2013).

313 ***Single-choice courtship assays***

314 We performed experiments to measure copulation latency, or the amount of time elapsed before
315 mating, according to male type (Y^M or III^M). In these experiments, we used the same IsoCS (Y^M)
316 and CS (III^M) strains as above and in Hamm et al. (2009). We also tested one other strain from
317 each genotype. CSrab (III^M) was created by backcrossing the III^M chromosome from the rspin
318 strain isolated in New York onto the CS background (Shono and Scott 2003; Son et al. 2019).
319 CSaY (Y^M) was created by backcrossing the Y^M chromosome from the aabys genome reference
320 strain onto the CS background (Scott et al. 2014; Meisel et al. 2015). Virgin females used in the
321 assays were all from the LPR strain (Scott et al. 1996), which has a different genetic background
322 than all males tested. In addition, we also assayed LPR males to determine how copulation
323 latencies of III^M and Y^M males compare to those of males from the same genetic background as
324 the females.

325 We first attempted to test flies reared at the same temperatures as in our competitive courtship
326 assays (18°C and 29°C), as well as at an intermediate developmental temperature (22°C).
327 However, we did not generate enough flies at 18°C, and so we only have data for flies raised at
328 22°C and 29°C. Our results demonstrate that 22°C is a sufficiently low temperature to detect
329 effects of both genotype and developmental temperature on courtship success (see below). All
330 larvae from each male strain were reared in 32 oz plastic containers on the same wheat bran diet
331 described above (Hamm et al. 2009). Upon emergence, unmated male and female progeny were
332 separated and fed water and a 1:1 mixture of dry milk:sugar *ad libitum* until assays were

333 conducted. Assays of males raised at 22°C were conducted 10-11 days after eclosion, while those
334 of males raised at 29°C were conducted 6-7 days after eclosion. This ensures that males were
335 assayed at similar physiological ages. Females were all raised at 25°C, and unmated females
336 were aged 8-9 days after eclosion (see Supporting Information 1 for all accumulated degree day
337 calculations).

338 We followed a similar protocol as in a previous experiment testing copulation latency in
339 *D. melanogaster* (Shorter et al. 2016). Briefly, five males from a single strain were aspirated
340 without anesthesia into an 8 oz container covered with a fine mesh cloth secured by rubber band.
341 Five LPR females were similarly transferred into the container, marking the start of the courtship
342 assay. The house flies were then observed every 10 minutes over the course of four hours.
343 Copulation latency was determined in two ways. First, we measured the amount of time elapsed
344 between the start of an assay and each mating within a container, defined as a male remaining
345 attached to a female for at least 1 minute (Hamm et al. 2009). Male house flies typically remain
346 attached to females for >60 minutes (Bryant 1980), making it unlikely, although possible, for us
347 to miss matings within 10 minute intervals. Individuals who did not mate were excluded from
348 this analysis. Second, we used a binary variable noting whether each male mated during the 4
349 hour assay. Although we were unable to distinguish between individual males in this assay, we
350 did not observe any males mate more than once within 4 hours in a pilot study conducted
351 between one male and five females, suggesting that observed matings were by different males.
352 All trials were conducted at 22–23°C.

353 To determine the effects of male type on the amount of time taken to mate, we used the glmer()
354 function in the lme4 package in R (Bates et al. 2015) to create a mixed effects model as follows:

355 $$L \sim G + T + G \times T + b + s,$$

356 including male genotype (G), developmental temperature (T), and their interaction as fixed
357 effects, batch (b) and strain (s) as random effects, and our response variable as copulation latency
358 (L) in minutes. For the binary measure of copulation latency, we used a binomial logistic
359 regression of the same model, with whether a fly mated as our dependent variable. We then
360 assessed significance of fixed effects (type II sum of squares) using the Anova() function in the
361 car package in R (Fox et al. 2013). Pairwise comparisons between male types (III^M, Y^M, and
362 LPR) were conducted using Z-tests of proportions.

363 Lastly, we conducted a survival analysis, with unmated males treated as right-censored
364 observations since information about their mating latency was incomplete. Specifically, we used
365 the Surv() function in the survival package in R (Fox and Carvalho 2012) to convert our
366 copulation latency measurements (in minutes) to a right-censored survival object. We then
367 created a Cox proportional hazards regression model using the same parameters as above, with
368 copulation latency (in minutes) as our response variable. Due to limitations of the survival

369 package, we included batch and strain as fixed effects in our model. We assessed significance of
370 fixed effects using analysis of deviance for a Cox model using the anova() function in the
371 survival package, and we report hazard ratios (HRs) and their 95% confidence intervals for
372 copulation latency effect sizes (Burke and Holwell 2021).

373 ***Assaying copulation latency in* CG2120 *upregulated* D. melanogaster**

374 We used CRISPR/dCas9 transcriptional activation (CRISPRa) to increase the expression of
375 *CG2120* in *D. melanogaster*. We used a strain that expresses single guide RNA (sgRNA)
376 targeting *CG2120* (BDSC ID 79962), from the Transgenic RNAi Project CRISPR
377 Overexpression (TRiP-OE) VPR collection (Ewen-Campen et al. 2017). We crossed males with
378 the sgRNA to females carrying a transgene encoding a deactivated Cas9 protein (dCas9)
379 expressed under one of two Gal4 drivers (Gal4>dCas9). One Gal4>dCas9 strain expresses Gal4
380 under regulation of the *apolpp* promoter (BDSC ID 67043), which expresses in the fat body
381 (Brankatschk and Eaton 2010; Van De Bor et al. 2015). The other Gal4>dCas9 strain regulates
382 Gal4 under the *elav* promoter (BDSC ID 67038), which expresses in neurons (Sink et al. 2001;
383 Zhang et al. 2002). We collected male progeny from the cross between the sgRNA strain and
384 each of the Gal4>dCas9 strains to assay copulation latency. Control males were generated by
385 substituting the *CG2120* sgRNA strain with a strain carrying an sgRNA that targets the QUAS
386 sequence from *Neurospora crassa* (BDSC stock 67539). All flies were raised at 25°C on a
387 medium consisting of cornmeal, yeast, sugar, and agar.

388 We measured copulation latency in *CG2120* upregulated males and control males. In each
389 experiment, we used a single unmated male from one of the crosses described above. That male
390 was combined with a single unmated female from either the CantonS or OregonR strain in a vial
391 containing our standard medium. The flies were observed for up to 2.5 h at 25°C, and the time at
392 which they began to copulate was recorded as our measurement of copulation latency. Fly pairs
393 that did not copulate were excluded from the analysis. Experiments were performed in four
394 batches for the *apolpp* driver and five batches for the *elav* driver.

395 We compared the fit of linear models in order to test for differences in copulation latency
396 between *CG2120* upregulated males and control males. We first constructed a model that
397 captured all variables in our experiment,

398 $$C \sim M + F + M \times F + b,$$

399 which included the fixed effects of male genotype (M, which can be control or *CG2120*
400 upregulated), female strain (F, for CantonS or OregonR), and the interaction between male
401 genotype and female strain, as well as the random effect of experimental block (b). We used the
402 lmer() function in the lme4 package in R (Bates et al. 2015) to determine how each variable
403 affects copulation latency (C). We compared that full model with one excluding the interaction
404 term with a χ^2 test in the anova() function in R (R Core Team 2019) in order to test if including

405 the interaction term offers a significantly better fit. If the interaction term did not significantly
406 improve the fit of the model, we compared the model without the interaction term with one
407 without any effect of the female strain (i.e., F excluded from the model). If there was not a
408 significant difference in fit between the models with and without F, we modeled female strain as
409 a random effect. To test for an effect of *CG2120* upregulation on copulation latency, we
410 compared the fit of the model with male genotype as the only fixed effect and two random
411 effects (f and b) with a model that only had the random effects (i.e,no fixed effects) using the
412 anova() function.

413 **RESULTS**

414 ***Differential expression of odorant binding protein genes between III^M and Y^M males***

415 Our first goal was to analyze RNA-seq data in order to identify DE genes between III^M and Y^M
416 male heads that could be responsible for previously observed differences in competitive
417 courtship assays between the genotypes (Hamm et al. 2009). We confirmed that the gene
418 expression profiles of III^M and Y^M male heads are minimally differentiated (Meisel et al. 2015;
419 Son et al. 2019). There are only 40 DE genes between heads of III^M and Y^M adult males (21
420 upregulated in Y^M males, 19 upregulated in III^M, Supplementary Table S3). Gene ontology
421 analysis revealed no significant biological process, molecular function, or cellular component
422 terms enriched within the list of DE genes.

423 Within the list of DE genes, we identified one gene (*LOC105261916*) encoding an odorant
424 binding protein (Obp) upregulated in Y^M males. House fly Obp genes can be grouped into
425 families corresponding to their *D. melanogaster* orthologs (Scott et al. 2014). The DE Obp gene
426 in our analysis is orthologous to *Obp56h*. The *Obp56h* family, as well as other Obp families, was
427 greatly expanded within Muscidae (house fly and close relatives, including stable fly and horn
428 fly) compared to *D. melanogaster* (Scott et al. 2014; Olafson and Saski 2020; Olafson et al.
429 2021). In addition to *LOC105261916*, seven of the remaining eight house fly *Obp56h* genes for
430 which we obtained RNA-seq count data showed similar trends of greater expression in Y^M than
431 III^M males, with three of these showing significant DE ($p<0.05$) before an FDR correction
432 (Fig. 1). All but one of the *Obp56h* genes has higher expression in Y^M than III^M males (8/9,
433 regardless of significance), which is significantly greater than the fraction of other genes with
434 higher expression in Y^M males, regardless of significance, in the rest of the genome (Fisher's
435 exact test, $p = 0.019$).

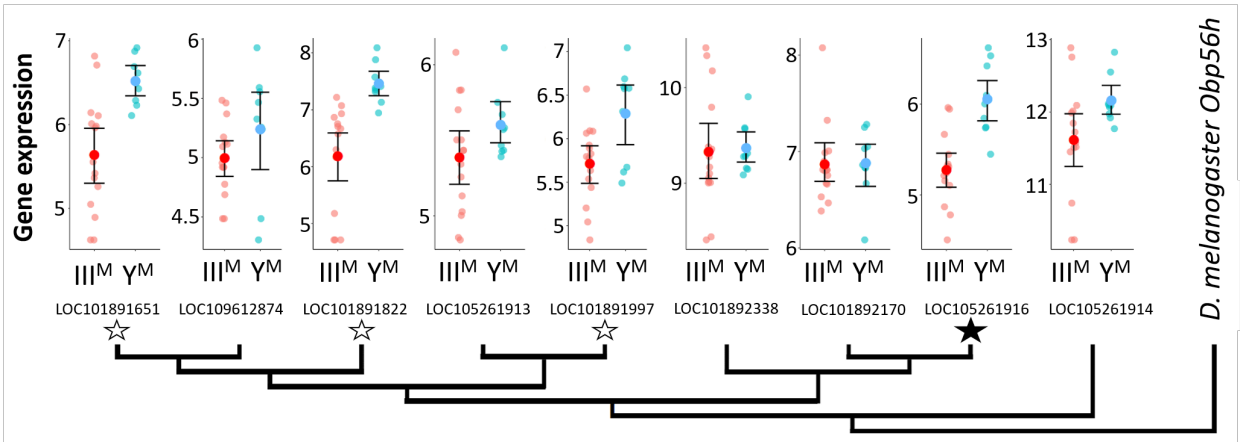


Figure 1 - Neighbor-joining phylogenetic tree of the *Obp56h* gene family within *M. domestica* and *D. melanogaster* based on protein sequences constructed in MEGA X (Kumar et al. 2018). Amino acid sequences were aligned by MUSCLE (Edgar 2004). *M. domestica Obp56h* genes are identified based on gene IDs. The bootstrap consensus tree was inferred from 10,000 replicates. Branch lengths are scaled according to the number of amino acid substitutions per site. The phylogeny was arbitrarily rooted at *D. melanogaster Obp56h.* Graphs at the branch tips show batch-adjusted expression levels for each *M. domestica Obp56h* gene from each replicate (small circles). Large circles show the average across all replicates, with error bars denoting the standard error (unfilled stars: $p < 0.05$ before FDR correction for multiple comparisons; filled star: $p < 0.05$ after correction).

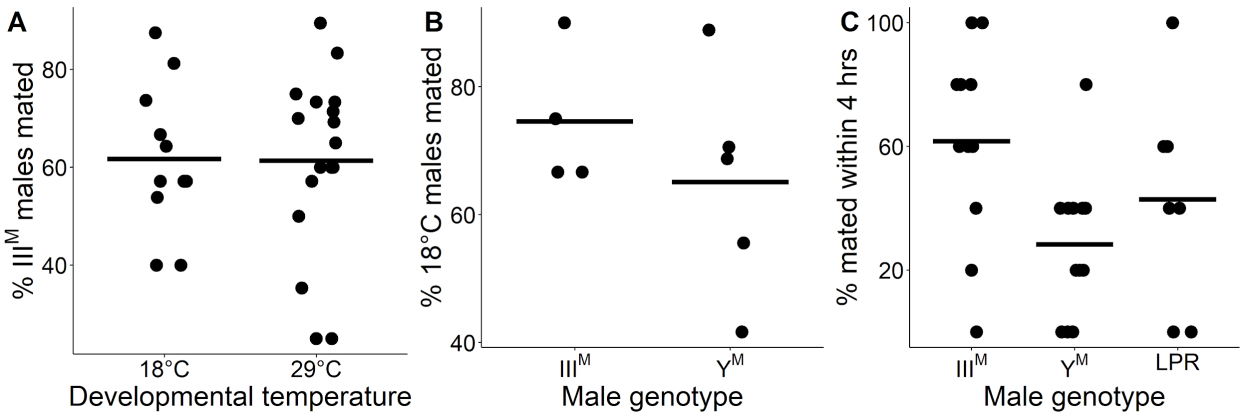
III^M confers a courtship advantage by reducing copulation latency

Knockdown of *Obp56h* in *D. melanogaster* decreases male copulation latency, or the time it takes for a male to begin to mate with a female after they are first exposed to one another (Shorter et al. 2016). The *Obp56h* gene family is generally expressed higher in Y^M males relative to III^M males (Fig. 1). A previous study identified a competitive advantage of III^M over Y^M male house flies in successfully engaging a female in mating (Hamm et al. 2009), consistent with shorter copulation latency in III^M males because of lower expression of *Obp56h* genes. We confirmed this result by performing competitive courtship assays in which we allowed males carrying III^M or Y^M to compete for a female of an unrelated strain. Consistent with the previous results, we observed that III^M males are more successful at mating than Y^M males when raised at 29°C (Fig. 2A), a temperature similar to the developmental temperature used by Hamm et al. (2009).

Two *Obp56h* genes are only upregulated in Y^M males at 29°C, but not at 18°C (Fig. S1), raising the possibility that the effect of III^M/Y^M genotype on male courtship success may be temperature-dependent. In addition, III^M males have greater heat tolerance than Y^M males (Delclos et al. 2021), further suggesting that the benefits of the III^M chromosome may be limited to warm temperatures. We therefore tested if the differences in courtship performance between Y^M and III^M males are sensitive to temperature. We found that III^M males were more successful

464 at mating than Y^M males regardless of developmental temperature (ANOVA, $X^2 = 20.7$, $p = 6.53$
465 x 10^{-6}, Fig. 2A). We then tested whether males reared at different developmental temperatures,
466 but with the same genotype, have a difference in courtship success. We found that males reared
467 at 18°C outcompete males reared at 29°C, regardless of genotype (ANOVA, $X^2 = 13.1$, $p = 2.93$
468 x 10^{-4}, Fig. 2B). This demonstrates that there is an effect of developmental temperature on
469 courtship success, but it appears to be independent of the genotype effect. More generally, our
470 results suggest that the III^M male courtship advantage is robust to developmental temperature,
471 which differs from the prediction based on temperature-dependent *Obp56h* expression (Fig. S1).

472



473
474 **Figure 2** - III^M chromosome and developmental temperature affect male courtship success. A) Outcomes
475 of competitive courtship assays between III^M and Y^M males reared at 18°C or 29°C. Data points represent
476 experimental batches. Horizontal lines denote the median across all batches. B) Outcomes of competitive
477 courtship assays conducted between males reared at 29°C and 18°C. Trials were conducted between
478 males of the same proto-Y chromosome genotype (III^M or Y^M). Each data point represents ten replicate
479 trials within a single batch. C) Outcomes of single-choice courtship assays in males reared at 22°C. Data
480 points refer to the percentage of males (five males within one replicate) that mated with females within 4
481 hours within each experimental trial. Horizontal lines denote means within male groups. All females used
482 were from the LPR strain.
483
484 We next tested if the effects of developmental temperature and genotype on competitive mating
485 advantage could be caused by differences in copulation latency. To measure copulation latency,
486 we combined five males from a single strain raised at a single temperature with five females
487 from the unrelated strain used in our competitive courtship assays. Developmental temperature
488 had a significant effect on copulation latency (ANOVA, $X^2 = 15.3$, $p = 9.40$ x 10^{-5}), with males
489 reared at 22°C mating faster than those reared at 29°C (Fig. S3A). This result is consistent with
490 increased courtship success of males raised at 18°C relative to those raised at 29°C in our
491 competitive assays (Fig. 2B).

492 We conducted a survival analysis to test for differences in copulation latency between Y^M and
493 III^M. To those ends, we fit a Cox proportional hazards regression model, treating unmated males

494 as censored data (Burke and Holwell 2021), and we observed significant effects of both
495 developmental temperature (ANOVA, $X^2 = 22.5$, $p = 2.16 \times 10^{-6}$) and male genotype ($X^2 = 18.0$,
496 $p = 1.22 \times 10^{-4}$) on copulation latency. However, there was no significant interaction effect
497 between genotype and developmental temperature ($X^2 = 2.63$, $p = 0.277$). Consistent with our
498 hypothesis that III^M males have a reduced copulation latency relative to Y^M males, at 22°C, III^M
499 males mated 3.5 times faster than Y^M males (Cox model: HR = 3.50, CI = 0.95 to 12.9, Fig.
500 S2A). We do not observe this trend at 29°C (Fig. S2B), possibly due to very few successful
501 matings at the warmer developmental temperature.

502 To further address the limitations of censored data, we treated copulation latency as a binary
503 variable by calculating the proportion of the five males per trial that mated within the 4 hour
504 assay. We observed significant effects of male genotype (ANOVA, $X^2 = 10.2$, $p = 6.18 \times 10^{-3}$)
505 and developmental temperature (ANOVA, $X^2 = 11.0$, $p = 9.04 \times 10^{-4}$) on the proportion of males
506 that mated. The effect of developmental temperature was largely a result of very few matings for
507 males that developed at 29°C relative to 22°C (Fig. S3B), consistent with our finding that males
508 that developed at a lower temperature have higher courtship success (Fig. 2B). In the 22°C
509 treatment, a significantly greater proportion of III^M males mated within 4 hours than Y^M (61.7%
510 v. 28.3%; Z-test of proportions, $p = 1.21 \times 10^{-4}$; Fig. 2C). This is further evidence that III^M males
511 have reduced copulation latency, which is consistent with their previously documented
512 competitive advantage (Hamm et al. 2009), the competitive advantage that we observe (Fig. 2A),
513 and the reduced expression of *Obp56h* genes (Fig. 1).

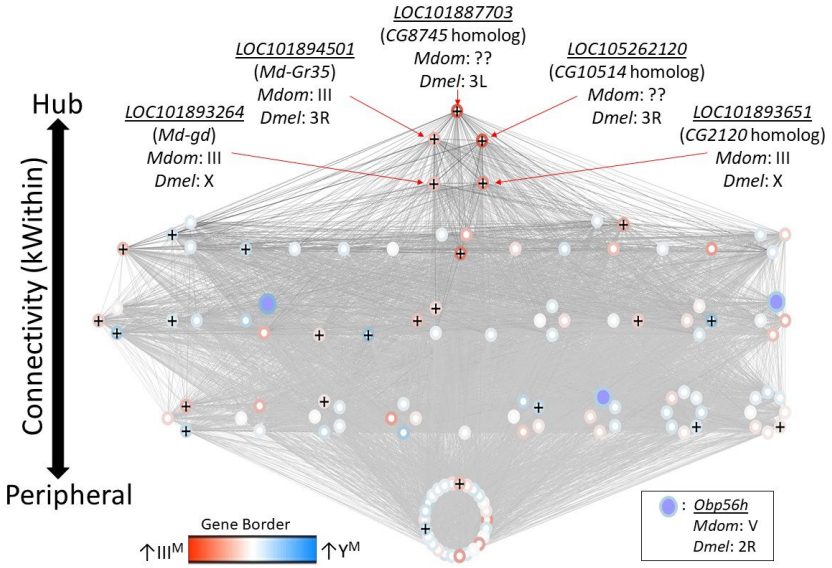
514 The Y^M and III^M males we and others used in courtship experiments all share the CS genetic
515 background that comes from a III^M strain (Hamm et al. 2009). This raises the possibility that III^M
516 males perform better because they have a proto-Y chromosome that is co-adapted to its genetic
517 background. To test this hypothesis, we measured copulation latency in Y^M males from the same
518 strain (LPR) as the females in our experiments. We observed a greater proportion of III^M males
519 mating within 4 hours when compared to the LPR Y^M males (Z-test of proportions, $p = 0.038$),
520 although the copulation latency in LPR males was highly variable (Fig. 2C). Therefore, the
521 reduced copulation latency conferred by the III^M chromosome overwhelms any potential effects
522 of coadaptation of the proto-Y chromosome to male genetic background or male-female co-
523 adaptation within strains. The reduced copulation latency of III^M males is only detectable when
524 house flies develop at 22°C, suggesting that it is either temperature-dependent or we lack the
525 resolution to detect it when males develop at warmer temperatures (because they take too long to
526 mate).

527 ***House fly chromosome III genes and* Drosophila *X chromosome genes have correlated***
528 ***expression with* Obp56h *genes***

529 The *Obp56h* gene cluster is found on house fly chromosome V, suggesting that it is regulated in
530 *trans* by genes on the Y^M or III^M chromosome. Chromosome V is unlikely to differ between the

531 Y^M and III^M males in our experiments—the majority of males compared in the RNA-seq data
532 and mating experiments have a common genetic background (including chromosome V) and
533 differ only in whether they carry III^M or Y^M. Removing samples with a different genetic
534 background did not affect the general difference in *Obp56h* expression between III^M and Y^M
535 males (see Supporting Information 2 for a summary of these results). We therefore aimed to
536 identify genes on the Y^M or III^M chromosome that could be responsible for the differential
537 expression of *Obp56h* genes between Y^M and III^M males.

538 We first identified 27 co-expression modules across Y^M and III^M house fly male heads. We focus
539 on one of these modules (containing 122 genes, Supplementary Table S4) because it is
540 differentially expressed between Y^M and III^M males ($p_{ADJ} = 0.001$), and it contains three *Obp56h*
541 genes that are DE between III^M and Y^M males (*LOC105261916, LOC101891822,* and
542 *LOC101891651*) (Fig. 3). GO analysis revealed significant enrichment ($p_{ADJ} < 0.05$) of 15
543 biological process terms including those related to immune system processes (GO:0032501),
544 responses to stress (GO:0006950), and response to external stimuli (GO:0009605) within this
545 module (Supplementary Table S5). We used the WGCNA measure of intramodular connectivity,
546 kWithin, to identify hub genes within the module that likely have important roles in the
547 regulation of gene expression. The top five hub genes are (with *D. melanogaster* orthologs in
548 parentheses): *LOC101887703* (*CG8745*), *LOC105262120* (*CG10514*), *LOC101894501* (*Md-*
549 *Gr35*), *LOC101893264* (*gd*), and *LOC101893651* (*CG2120*) (Fig. 3, S4).

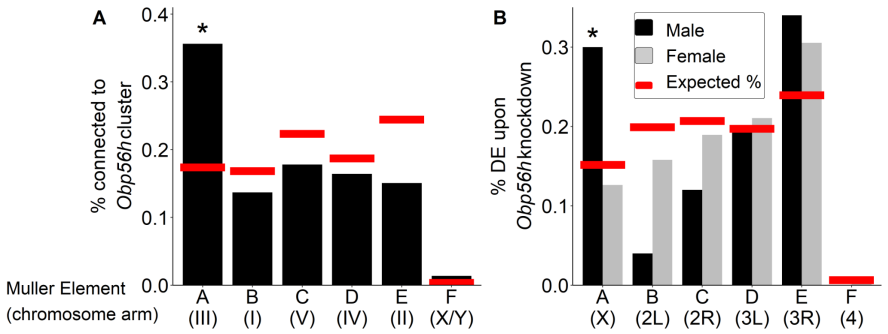


550
551 **Figure 3** - Network visualization of the co-expression module that is differentially regulated between III^M
552 and Y^M males. Each circle within the module is a gene, and *Obp56h* genes are indicated with purple fill.
553 Lines represent edge connections between genes. Genes labeled with "+" are within the top 100 most
554 strongly connected to *Obp56h* genes. Genes are ordered from top to bottom according to intramodular
555 connectivity (kWithin), with genes of higher connectivity (i.e., hub genes) on top, and peripheral genes on
556 the bottom. Borders around genes denote $\log_2$ fold-change in expression between Y^M and $IIIIM$ male

557 heads, with darker blue borders denoting upregulation in Y^M, and darker red borders denoting
558 upregulation in III^M. Chromosomal locations in house fly (*Mdom*) and *D. melanogaster* (*Dmel*) are shown
559 for the 5 hub genes and *Obp56h*.

560 The genes present in the focal co-expression module provide multiple lines of evidence that
561 *Obp56h* gene expression is regulated by *trans* factors that map to chromosome III. First, the
562 module is enriched for house fly chromosome III genes (31 chromosome III genes versus 38
563 genes assigned to other chromosomes, Fisher's exact test $p < 1 \times 10^{-5}$, with 53 genes not assigned
564 to a chromosome) and for DE genes between Y^M and III^M males (16 DE genes in this module
565 versus 24 DE genes assigned to other modules, Fisher's exact test $p < 1 \times 10^{-5}$). In addition, the
566 co-expression module is enriched for *Obp56h* genes relative to other Obp genes—three *Obp56h*
567 genes and no other Obp genes were assigned to this module (Fisher's exact test, $p = 5.1 \times 10^{-3}$,
568 Fig. 3). Furthermore, there is a significant enrichment of chromosome III genes within the 100
569 genes whose expression covaries most with *Obp56h* gene expression (corresponding to the top
570 0.55% covarying genes; Supplementary Table S6); of the 100 genes with the highest *Obp56h*
571 connection scores, 26 are on chromosome III (Fisher's exact test $p = 2.0 \times 10^{-4}$, Fig. 4A). This
572 enrichment is robust to varying the threshold used to classify a gene as in the top covarying;
573 considering genes with the top 1%, 5%, or 10% covarying expression also resulted in significant
574 enrichment of chromosome III genes (Fisher's exact test, all $p < 0.05$). These results support the
575 hypothesis that *trans* regulatory variants that differ between III^M and the standard chromosome
576 III are responsible for the differential expression of *Obp56h* genes between III^M and Y^M house
577 fly males. We cannot perform the same analysis for the effect of Y^M because only 51 genes have
578 been assigned to the house fly X/Y^M chromosome (Meisel and Scott 2018), limiting our power to
579 detect an excess of genes.

580



581

582 **Figure 4** - Percent of genes on each chromosome within (A) the top 100 genes with the strongest
583 correlated co-expression to the *Obp56h* family in house fly, and (B) genes differentially expressed (DE)
584 between *Obp56h* knockdown and control *D. melanogaster* (black bars: males, grey bars: females).
585 Asterisks indicate a significant difference between observed (bars) and expected (red lines) counts of
586 genes on each chromosome compared to all other chromosomes (Fisher's exact test, $p < 0.05$).

587

588 Our network analysis does not ascribe directions to the edges connecting house fly genes, and it
589 is therefore possible that *Obp56h* has *trans* regulatory effects on chromosome III expression. To
590 test this hypothesis, we examined available RNA-seq data from an experiment comparing wild-
591 type *D. melanogaster* with flies in which *Obp56h* had been knocked down (Shorter et al. 2016).
592 *Obp56h* is on the right arm of the second chromosome in *D. melanogaster* (2R, or Muller
593 element C), which is homologous to house fly chromosome V (Foster et al. 1981; Weller and
594 Foster 1993). House fly chromosome III is homologous to the *D. melanogaster* X chromosome,
595 which is known as Muller element A (Meisel and Scott 2018; Schaeffer 2018). The
596 *D. melanogaster* males in the RNA-seq experiment all share the same X chromosome, and only
597 differ in one copy of their second chromosome (which either carries a UAS-RNAi knockdown
598 construct or does not). If *Opb56h* genes have *trans* regulatory effects on element A genes in
599 males, we would expect an excess of DE *D. melanogaster* X chromosome genes in *Obp56h*
600 knockdown flies. Indeed, we found that *Obp56h* knockdown in *D. melanogaster* resulted in an
601 excess of X chromosome DE genes in male head (Fisher's exact test, $p = 0.011$, Fig. 4B) and
602 body ($p = 0.038$, Fig. S5), but not in either tissue sample in females (Fisher's exact test, both $p >$
603 0.49). These results suggest that there is male-specific *trans* regulatory control of
604 *D. melanogaster* X-linked genes by *Obp56h*.

605 We found multiple similarities between house fly and *D. melanogaster* that suggest the genes
606 that regulate and/or are regulated by *Obp56h* are evolutionarily conserved between the two
607 species. Specifically, genes that were downregulated upon knockdown of *Obp56h* in
608 *D. melanogaster* have house fly orthologs that are more downregulated in III^M male house flies
609 (i.e., lower $\log_2$ fold-change) than expected by chance ($p = 5.60 \times 10^{-3}$, Fig. S6A). However,
610 genes that were upregulated upon *Obp56h* knockdown in *D. melanogaster* were not significantly
611 differentially regulated between Y^M and III^M male genotypes, although the observed trend
612 suggests that these genes may be more downregulated in III^M males than expected ($p = 0.103$,
613 Fig. S6B). The GO term "response to stress" (GO:0033554) is significantly enriched amongst
614 genes with strong connection scores with *Obp56h* expression in *M. domestica* and in the list of
615 DE genes in *D. melanogaster* upon *Obp56h* knockdown (Supplementary Table S7), providing
616 additional evidence for conserved co-regulation. Altogether, our results suggest that there is
617 evolutionarily conserved *trans* regulatory feedback loop involving *Obp56h* expression and
618 Muller element A in *Drosophila* and house fly through similar molecular functions.

619 ***Network analysis reveals candidate regulators of* Obp56h *expression***

620 The house fly co-expression module contains candidate genes and pathways through which
621 *Obp56h* genes, and likely male copulation latency, are regulated. For example, a serine protease
622 gene, *LOC101893264*, orthologous to *D. melanogaster gd* (Konrad et al. 1998), is among the top
623 5 hub genes within the co-expression module (Fig. 3). This gene is predicted to encode a positive

regulator of the Toll signaling pathway (LeMosy et al. 2001; Valanne et al. 2011), suggesting that the *M. domestica* ortholog of *gd* (*Md-gd*) could have an important gene regulatory function within the module via Toll signaling. *Md-gd* is located on chromosome III, and it is upregulated in III^M males (adj. $p = 0.022$). We tested if *Md-gd* is differentially regulated between the III^M chromosome and standard chromosome III by comparing expression in III^M males (i.e., heterozygotes for III^M and a standard chromosome III) with Y^M males that are homozygous for the standard chromosome III. Differential expression of the III^M and III chromosome alleles would implicate *Md-gd* as having a causal effect on *Obp56h* expression. We identified seven polymorphic sites where all RNA-seq reads were mapped to the III^M allele, while no reads were mapped to the standard chromosome III allele in III^M males (Fig. 5A). At all seven diagnostic SNPs in this gene, the III^M allele is significantly more highly expressed than the III allele in III^M males (all $p = 0.021$), and it is more highly expressed than both III alleles in Y^M males at six of seven sites (all $p = 0.021$). Higher expression of the III^M allele is consistent with *cis* regulatory divergence between the III^M and standard chromosome III being responsible for elevated *Md-gd* expression in III^M males. The lack of expression of the III allele in either III^M or Y^M males is consistent with monoallelic gene expression of the III^M allele, although further evidence is required to confirm this hypothesis (see Supporting Information 3 for detailed discussion).

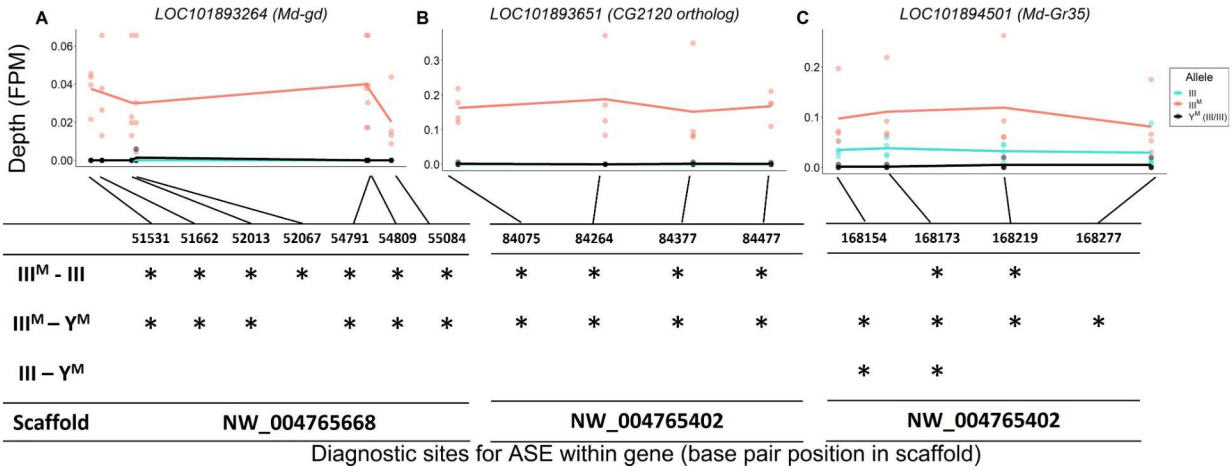


Figure 5 - Allele-specific expression (ASE) in A) *LOC101893264* (*Md-gd*), B) *LOC101893651* (*CG2120* ortholog), and C) *LOC101894501* (*Md-Gr35*). The x-axis depicts base pair positions (scaffold coordinates) of the informative single nucleotide polymorphisms (SNPs) that differ between III^M and standard chromosome III alleles. The y-axis and data points depict the read depth of a given allele normalized by the total mapped reads for a given strain-by-experimental batch group combination (FPM = fragments per million). Lines depict mean read depths at each diagnostic site for III (turquoise) and III^M (salmon) alleles in III^M males, and mean read depths at each site for III alleles in Y^M males (black). Tables under each graph mark significant differences (*: $p < 0.05$) in normalized read depths at each diagnostic site for each of three pairwise comparisons: III^M allele vs. III allele in III^M males (III^M-III), III^M allele in III^M males vs both III alleles in Y^M males (III^M-Y^M), III allele in III^M males vs. both III alleles in Y^M males (III-Y^M).

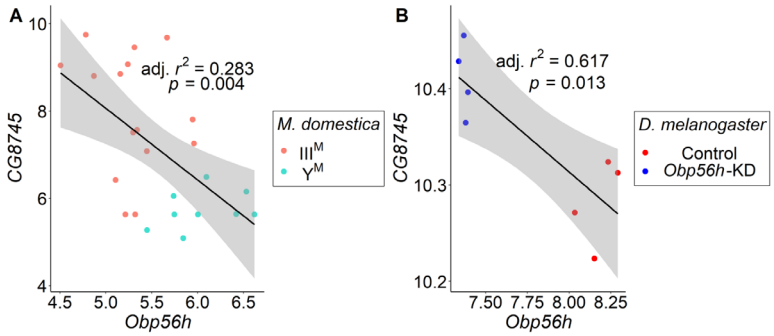
653 We identified similar evidence of monoallelic gene expression within another hub gene,
654 *LOC101893651*, which is orthologous to *D. melanogaster CG2120* (Fig. 5B). *LOC101893651* is
655 among the most central genes within the co-expression module (Fig. 3), and it is strongly
656 upregulated in III^M males ($\log_2$ fold-change: 1.33, $p_{ADJ} = 0.033$). *LOC101893651* is found on
657 house fly chromosome III and is predicted to encode a transcription factor. At all four diagnostic
658 sites within *LOC101893651*, the III^M allele is significantly more highly expressed than the III
659 allele in III^M males (all $p = 0.021$), as well as both III alleles in Y^M males (all $p \leq 0.027$). Within
660 the WGCNA module, *Obp56h* expression is most strongly correlated with *LOC101893651*,
661 suggesting that *LOC101893651* could encode the transcription factor that is directly responsible
662 for the repression of *Obp56h* expression in III^M males. Consistent with this hypothesis, there is
663 evidence that the protein encoded by *CG2120* binds both upstream and downstream of *Obp56h*
664 in *D. melanogaster* (Kudron et al. 2018). Below, we describe experimental results that test the
665 hypothesis that *LOC101893651* (*CG2120*) regulates the expression of *Obp56h* genes.

666 There is also a gustatory receptor gene (*LOC101894501*) that is a hub in the co-expression
667 module and upregulated in III^M males ($p_{ADJ} = 0.037$). *LOC101894501* (also annotated as *Md-*
668 *Gr35*) is a homolog of *D. melanogaster Gr98c* (Scott et al. 2014). We evaluated the relationship
669 between *Obp56h* and the hub genes in the co-expression network module, and we found greater
670 evidence that this module and *Obp56h* plays an important role in chemosensation (Supporting
671 Information 4, Fig. S7). *Md-Gr35* contains 4 exonic SNPs differentiating the III^M and III
672 chromosomes. Within each III^M strain in each RNA-seq experiment, we observed significantly
673 greater expression of the III^M allele than the standard chromosome III allele at two of the four
674 diagnostic SNP sites (Fig. 5C). The other two SNPs showed the same pattern of III^M-biased
675 expression but were not significant (both $p > 0.05$). The III^M allele in III^M males is also expressed
676 higher than both III alleles in Y^M males, consistent with *cis* regulatory divergence between the
677 III^M and standard chromosome III driving elevated *Md-Gr35* expression in III^M males. However,
678 the standard chromosome III allele is expressed significantly higher in III^M males than Y^M males
679 at two of the four diagnostic SNP sites (Fig. 5C); we observe the same pattern at the other two
680 sites without significance ($p > 0.05$). Higher expression of the III allele in III^M males than Y^M
681 males suggests that *trans* regulators further increase the expression of *Md-Gr35* in III^M males.
682 This combination of *cis* and *trans* regulatory effects on *Md-Gr35* expression are consistent with
683 the *trans*-regulatory loop we hypothesize between *Obp56h* and chromosome III that regulates
684 male copulation latency. Future experiments could determine whether *Gr98c* (*Md-Gr35*) and
685 *Obp56h* do indeed interact and, if so, what pheromonal or other chemical compounds they
686 detect.

687 In contrast to the aforementioned three hub genes, the most central gene within the module
688 (*LOC101887703*) may be regulated by *Obp56h*. *LOC101887703* is orthologous to
689 *D. melanogaster CG8745*, which is predicted to encode an ethanolamine-phosphate phospho-
690 lyase and is broadly expressed in many *D. melanogaster* tissues (Chintapalli et al. 2007).
691 *LOC101887703* is upregulated in III^M males ($\log_2$ fold-change: 2.21, $p_{ADJ} = 0.016$), which could

692 have a causal effect on *Obp56h* DE, be caused by *Obp56h* DE, or neither. In both
693 *D. melanogaster* and house fly, *Obp56h* expression is significantly negatively correlated with the
694 expression of *CG8745* or *LOC101887703*, respectively (Fig. 6). Directly comparing *CG8745*
695 expression between control and *Obp56h*-knockdown *D. melanogaster* yields qualitatively similar
696 results, with *Obp56h*-knockdown flies showing greater expression of *CG8745* than controls
697 (Welch's *t*-test, $t = -4.27$, $p = 5.53 \times 10^{-3}$). The negative correlation between *CG8745* and
698 *Obp56h* expression in *D. melanogaster* suggests that *Obp56h* downregulation causes *CG8745*
699 upregulation because *Obp56h* expression was directly manipulated by RNAi in the experiment.
700 The same relationship between *CG8745* and *Obp56h* expression in both house fly and
701 *D. melanogaster* provides evidence that similar molecular mechanisms may underlie the
702 hypothesized *trans* regulatory effect of *Obp56h* in both species.

703



704

705 **Figure 6** - Correlations of gene expression between *Obp56h* (house fly *LOC105261916*) and *CG8745*
706 (*LOC101887703*) in (A) house fly male head tissue, and (B) *D. melanogaster* male head tissue. Values
707 for *D. melanogaster* are from count data as reported in Shorter et al. (2016). Linear regression models
708 were used to determine 95% confidence intervals (shaded in grey) summarizing the effect of *Obp56h*
709 expression on *CG8745* expression in each species.

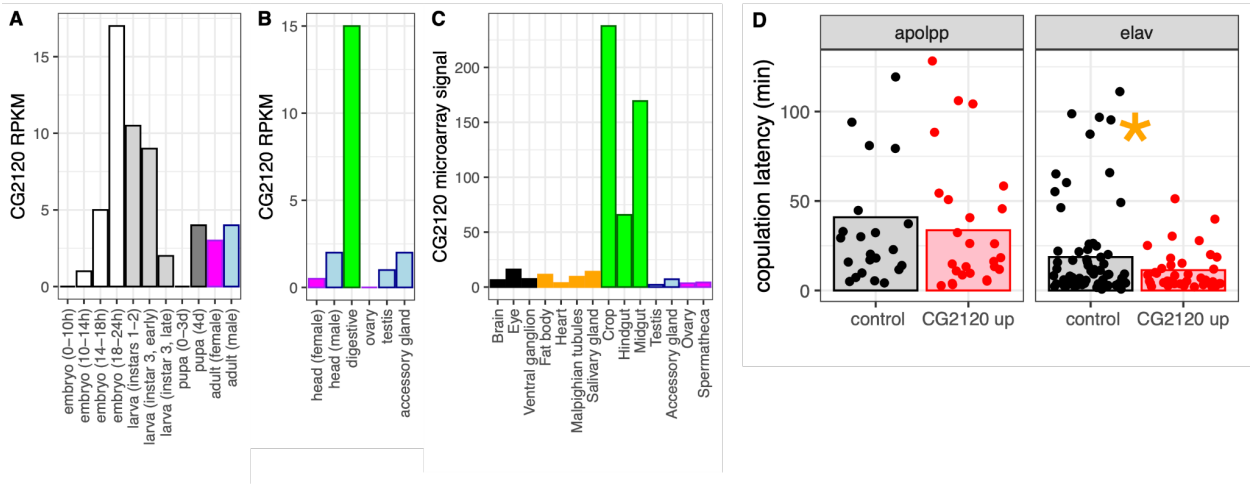
710 **CG2120 *expression affects courtship***

711 We investigated the expression of *CG2120*, one of the hub genes in the house fly co-expression
712 network, across *D. melanogaster* development and tissues. Expression of *CG2120* oscillates over
713 the life history of *D. melanogaster*: increasing throughout embryonic development, decreasing
714 with each subsequent larval instar, and then increasing in pupa and in some adult tissues (Fig
715 7A-C). *CG2120* is predominantly expressed in digestive tissues in adults, with much lower
716 expression in nervous tissues (Fig 7B-C). Notably, *CG2120* is more highly expressed in male
717 than female heads (Fig 7B), consistent with a regulatory effect on male courtship behavior.

718 We tested the hypothesis that *CG2120* affects copulation latency in *D. melanogaster* via negative
719 regulation of *Obp56h* expression. To that end, we used CRISPRa to upregulate expression of
720 *CG2120* in fat body cells or neurons of *D. melanogaster* males, using an *apolpp* and *elav*

721 promoter, respectively. We tested if upregulation of *CG2120* reduces copulation latency, as
722 would be expected if *CG2120* negatively regulates *Obp56h*. We used females from two different
723 strains in our assays (CantonS and OregonR). Neither female strain nor the interaction between
724 female strain and male genotype had a significant effect on copulation latency, regardless of the
725 tissues in which expression was induced (Supplementary Table S8). There was also not a
726 significant difference in copulation latency between control and *CG2120* upregulated flies when
727 we activated expression in fat body cells (Supplementary Table S8, Fig. 7D). In contrast, there
728 was a significant difference between control and *CG2120* upregulated males when expression
729 was activated in neurons using the *elav* driver, with *CG2120* upregulated males mating faster
730 than controls (Supplementary Table S8, Fig. 7D). This is consistent with *CG2120* activation in
731 neurons negatively regulating *Obp56h*, which then reduces copulation latency.

732



733

734 **Figure 7.** CG2120 is broadly expressed and affects copulation latency. Expression levels of *CG2120* are
735 plotted using (A) RNA-seq data from developmental stages (Graveley et al. 2011), (B) RNA-seq data
736 from adult tissues (Brown et al. 2014), or (C) microarray data from adult tissues (Chintapalli et al. 2007).
737 Expression measurements are shown as RPKM values for RNA-seq data and signal intensity for
738 microarray data (Gelbart and Emmert 2013). Bars are colored based on tissue sample categories: embryo
739 (white), larva (light gray), pupa (dark gray), adult female (magenta), adult male (light blue), neuronal
740 (black), digestive tract (green), or other adult tissues (orange). (D) Copulation latency in control male
741 *D. melanogaster* and males in which *CG2120* was upregulated (CG2120 up) in fat body cells (using an
742 apolpp Gal4 driver) or neurons (elav driver). Each dot is a single male for which copulation latency was
743 measured. Vertical bars show the estimated effect of male genotype (control or upregulated) from a linear
744 model in which male genotype is a fixed effect and female strain and batch are random effects. The
745 asterisk indicates a significant difference (P<0.05) between control and CG2120 upregulated for a
746 specific tissue driver.

747

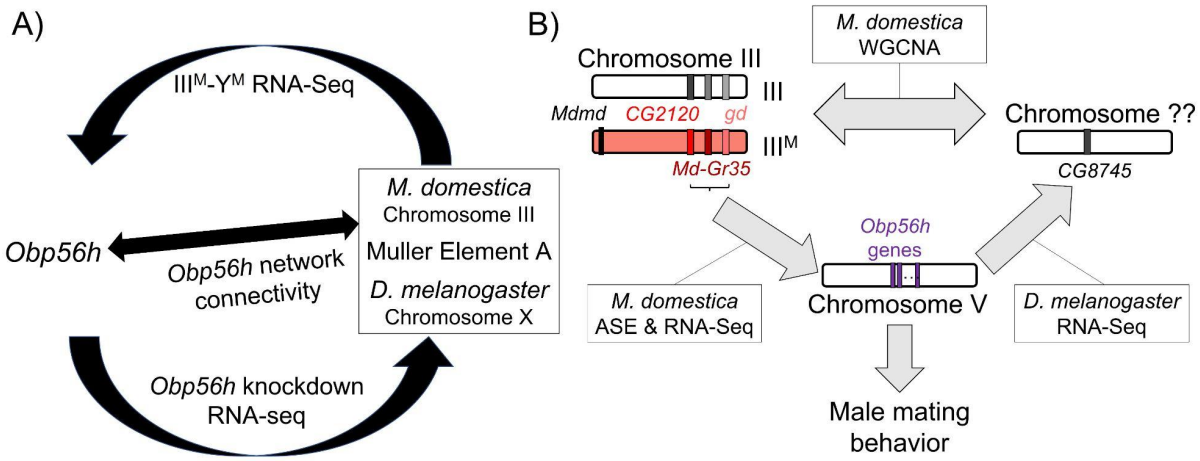


Figure 8 - Hypothesized relationships between *Obp56h* expression, proto-Y chromosome genotype, and male mating behavior based on house fly and *D. melanogaster* gene expression data. A) Summary of evidence for an evolutionarily conserved *trans* regulatory loop between *Obp56h* and Muller Element A (house fly chromosome III, and *D. melanogaster* X chromosome). Our hypothesis is based on differential expression between III^M vs. Y^M male house flies, *Obp56h* knockdown vs. control *D. melanogaster*, and network connectivity of *Obp56h* family gene expression within house fly. B) Summary of candidate genes implicated in conserved *trans* regulatory loop. Three of the top five hub genes of module A are located on house fly chromosome III, are negatively correlated with *Obp56h* expression, and exhibit either allele-specific expression (ASE) or show signs of monoallelic gene expression biased towards the III^M allele. Similar correlations between expression measures of *Obp56h* and *CG8745* (*LOC101887703*) in *D. melanogaster* and house fly male head tissue suggest that *Obp56h* regulates *CG8745*, which is the primary hub gene in the WGCNA module that is differentially expressed between III^M and Y^M male house flies. Shared correlations between *Obp56h* expression and copulation latency in both house fly and *D. melanogaster* also suggest that *Obp56h* regulates male fly mating behavior.

DISCUSSION

We combined analysis of functional genomic data, behavioral experiments, and genetic manipulation to determine the regulatory architecture underlying variation in a courtship trait in flies (Fig 8). Our results provide evidence that *Obp56h* expression affects copulation latency in house fly, consistent with its effects in *D. melanogaster* (Shorter et al. 2016). Based on their locations on house fly chromosome III, positions as hub genes in a house fly co-expression module (Fig. 3), and divergent expression between the III^M and standard III chromosomes (Fig. 5), we hypothesize that *LOC101893651* (*CG2120*), *Md-gd*, and/or *Md-Gr35* (*Gr98c*) negatively regulate *Obp56h* expression (Fig. 8B). In addition, we hypothesize that *Obp56h* negatively regulates *CG8745* (*LOC101887703*) because knockdown of *Obp56h* causes an increase in *CG8745* expression in *D. melanogaster* (Fig. 6B). Together, these results suggest that there is a *trans* regulatory loop in which *Obp56h* is both regulated by and regulates expression of genes on fly Muller element A (house fly chromosome III and *D. melanogaster* chromosome X).

We have experimentally validated that *CG2120* affects copulation latency in a way that is consistent with negative regulation of *Obp56h* in *D. melanogaster* neurons (Fig 7D). We propose the name *speed date* (*spdt*) for *CG2120* because its expression reduces copulation latency, and we refer to the house fly ortholog (*LOC101893651*) as *Md-spdt*. Future manipulative experiments will help in further evaluating the direction of regulation of the other co-expressed genes.

Sexual antagonism, pleiotropic constraints, and sexual selection

Repression of *Obp56h* expression reduces copulation latency, which we show is associated with an advantage in male-male competition for female mates (Fig. 2). The apparently simple correlation between *Obp56h* expression and copulation latency suggests that sexual selection should favor downregulation of *Obp56h* expression. This raises an important question: why do some males not downregulate *Obp56h* expression to gain a sexually selected advantage? III^M males do indeed express *Obp56h* at lower levels than Y^M males (Fig. 1), demonstrating that downregulation is possible. In addition, experimental repression of *Obp56h* in *D. melanogaster* males also reduces copulation latency relative to wild-type males (Shorter et al. 2016), further demonstrating the possibility for downregulation.

We hypothesize that the regulatory architecture associated with *Obp56h* expression creates evolutionary constraints that inhibit selection to reduce expression of *Obp56h* in Y^M males and *D. melanogaster*. Our analysis of coexpressed genes suggests that *Obp56h* exists within a complex regulatory network, including *trans* feedback, that is conserved between *Drosophila* and house fly (Fig. 8). This network architecture likely constrains the evolution of *trans* regulators of *Obp56h* and also creates downstream pleiotropic effects of changes to *Obp56h* expression. These pleiotropic effects likely create correlations between traits and weaken the response to selection (Lande and Arnold 1983).

Selective constraints on *Obp56h* expression may arise because the *trans* regulators are evolutionary constrained. We identified three candidate negative regulators of *Obp56h* (*spdt*, *gd*, and *Gr98c*), one or more of which may be constrained in their ability to increase in expression. Importantly, *trans* regulators are predicted to have pleiotropic effects, which could impede the response to selection on traits they affect (Carroll 2005). For example, *gd* is a serine protease involved in Toll signaling (Konrad et al. 1998; LeMosy et al. 2001). The Toll signaling pathway regulates cellular processes, including embryonic development and immune response to infection (Belvin and Anderson 1996; Valanne et al. 2011). Therefore, upregulation of *gd* could have pleiotropic effects that constrain the evolution of *gd* expression.

The expression profile of *spdt* (*CG2120*) suggests that its upregulation in head or neurons might be similarly constrained. Notably, *spdt* is expressed highly in embryos and in the adult digestive tract, but much lower at other developmental stages or tissues, including adult head (Fig. 7A–C).

816 The high expression in digestive tissues suggests that it is mechanistically possible for *spdt* to
817 evolve higher expression in neurons, which would decrease copulation latency (Fig. 7D). We
818 hypothesize that constraints on the regulators of *spdt*, or the genes it regulates, likely prevent its
819 upregulation in neurons—specifically neurons in the head that affect copulation latency.
820 Consistent with this prediction, there are 6,142 binding locations identified for Spdt from ChIP-
821 seq data (Kudron et al. 2018), suggesting that upregulation of *spdt* would have downstream
822 effects on the expression of a large portion of genes in the genome.
823
824 We additionally hypothesize that downregulation of *Obp56h* may have negative fitness effects
825 that further prevent the evolution of reduced expression, even though it could be advantageous to
826 males. One deleterious effect may arise from the interaction between *Obp56h* and *Gr98c* (*Md-*
827 *Gr35*). *Md-Gr35* is a hub gene in the house fly co-expression module containing *Obp56h*, and
828 the expression levels of *Md-Gr35* and *Obp56h* are negatively correlated (Fig. 3). In addition, we
829 find evidence for both *cis* and *trans* effects on the upregulation of *Md-Gr35* in III^M males (Fig.
830 5C), suggesting a possible feedback on *Md-Gr35* from *Obp56h* downregulation. Obps most
831 typically interact with chemosensory receptors (odorant, ionotropic, and gustatory receptors) in
832 the detection of chemical cues or signals, although they can have other functions as well (Zhou
833 2010; Benoit et al. 2017,Sun et al. 2018b). *Md-Gr35* is the only chemosensory receptor assigned
834 to the co-expression module, suggesting that *Md-Gr35* and *Obp56h* are each other's interacting
835 partners. Chemosensory binding proteins and receptors are also known to co-regulate one
836 another, and a negative correlation between the expression of a chemosensory receptor and its
837 interacting binding protein has previously been reported in a pair of genes that modulate male
838 *Drosophila* mating behavior (Park et al. 2006). If *Obp56h* serves a sensory detection role in fly
839 heads, then *Gr98c* (*Md-Gr35*) is a promising candidate gene with which it interacts. The
840 interactions between these two genes may constrain selection on *Obp56h* expression levels.
841
842 One mechanism by which interactions between *Obp56h*, *Gr98c* (*Md-Gr35*), and other co-
843 expressed genes may create constraints is through the production of cuticular hydrocarbons
844 (CHCs). CHCs are lipid compounds used for both chemical communication and resistance to
845 various environmental stressors, including desiccation (Chung and Carroll 2015). The expression
846 of *Obp56h* is associated with CHC profiles in *D. melanogaster* (Shorter et al. 2016). In addition,
847 *LOC105262120*, another hub gene in the network (Fig. 3), is homologous to *CG10514*, whose
848 expression is correlated with CHC production in *Drosophila serrata* (McGraw et al. 2011).
849 CHCs are often under strong sexual selection across insect systems, with individual or
850 combinations of CHCs serving as important mating cues (Thomas and Simmons 2009,Berson et
851 al. 2019a,b). The correlation of multiple hub genes with CHC profiles in *Drosophila* provides
852 additional evidence that *Obp56h* expression, and the house fly co-expression module more
853 generally, is related to male mating behavior, and possibly under sexual selection.
854

855 Pleiotropic or intersexual conflicts may arise over *Obp56h* expression because CHCs are
856 important for protection against abiotic stressors (Arya et al. 2010; Blomquist and Bagnères
857 2010; Otte et al. 2018,Sun et al. 2018a), including desiccation resistance (Lockey 1988).
858 Disrupted expression of individual genes responsible for CHC production in *D. melanogaster*
859 can result in significant alterations to both mating behaviors and ecologically relevant
860 phenotypes (Marcillac et al. 2005; Shorter et al. 2016), which could impede the response to
861 sexual selection via tradeoffs with natural selection (Rowe and Rundle 2021). This dual role of
862 CHCs in mating and desiccation resistance suggests that sexual selection on *Obp56h* expression
863 could be pleiotropically constrained by trade-offs with stress response. Our GO enrichment
864 analysis on both the house fly and *D. melanogaster* RNA-seq data also revealed that *Obp56h*
865 expression is correlated with the expression of genes involved in general stress responses,
866 supporting this hypothesis. Pleiotropic constraints on *Obp56h* expression (because of correlated
867 changes in CHCs) could therefore reduce the response to selection on male copulation latency.
868
869 Additional pleiotropic constraints are possible because Obps have myriad functions beyond
870 chemical detection (Findlay et al. 2008; Arya et al. 2010; Benoit et al. 2017,Sun et al. 2018a).
871 *Obp56h* expression specifically has been shown to affect mating behavior, avoidance of bitter
872 tastants, and the expression of genes related to immune response and heat stress (Swarup et al.
873 2014; Shorter et al. 2016). Moreover, single base pair changes at the *Obp56h* locus in
874 *D. melanogaster* cause sex-specific effects on a wide range of fitness-related traits, including
875 heat resistance (Mokashi et al. 2021). Future studies should aim at determining whether *Obp56h*
876 expression in III^M and Y^M male house flies is also associated with other phenotypes, such as
877 CHC profiles, desiccation resistance, or tolerance to other environmental stressors.
878
879 In general, our results provide evidence that pleiotropy can reduce the response to sexual
880 selection by creating genetic covariation amongst unrelated traits (Fitzpatrick 2004; Chenoweth
881 and McGuigan 2010). Specifically, selection on *Obp56h* expression and male courtship behavior
882 could be weakened by trade-offs between courtship behavior and stress response. In addition, the
883 regulatory architecture underlying *Obp56h* expression likely creates additional pleiotropic
884 constraints that could impede selection on copulation latency. One expected consequence of
885 these pleiotropic constraints is that there will be genetic variance for sexually selected traits in
886 natural populations (Kirkpatrick and Ryan 1991; Turelli and Barton 2004; Johnston et al. 2013;
887 Heinen-Kay et al. 2020), which we observe in the form of variation in copulation latency
888 between III^M and Y^M males.

889 ***Sexually antagonistic alleles on Y chromosomes***

890 Our results suggest that III^M males may have overcome the pleiotropic constraints associated
891 with *Obp56h* expression via the Y-linkage of at least one *trans* regulator of *Obp56h*. We
892 identified three candidate *trans* regulators of *Obp56h* on the house fly III^M chromosome (*Md-*
893 *spdt*, *Md-gd*, and *Md-Gr35*), all of which could be inhibitors of Obp56h expression (Fig. 8). This

894 is consistent with previous work suggesting that male-beneficial *trans* regulatory alleles could be
895 important for the fitness effects of the house fly III^M chromosome (Adhikari et al. 2021). More
896 generally, *trans* regulation of autosomal and X chromosome expression appears to be an
897 important feature of the phenotypic and fitness effects of Y chromosomes (Chippindale and Rice
898 2001; Lemos et al. 2008, 2010; Brown et al. 2020; Morris et al. 2020; Kawamoto et al. 2021;
899 Sandkam et al. 2021),
900
901 We hypothesize that the pleiotropic constraints associated with the regulation of *Obp56h*
902 expression are a source of sexual antagonism. As described above, deleterious fitness effects
903 could result from both pleiotropic effects of upregulation of the negative regulators of *Obp56h* or
904 downstream effects of the downregulation of *Obp56h*. Regardless of the cause, these fitness
905 costs are likely to affect both males and females. However, downregulation of *Obp56h* confers a
906 fitness benefit to males that may offset any fitness costs. III^M males can realize this fitness
907 benefit because the III^M chromosome carries at least one negative *trans* regulator of *Obp56h*.
908 Therefore, our results provide evidence that sexual antagonism can arise via conflicts between
909 sexual selection in males and opposing pleiotropic effects in females (Lande 1980; Fitzpatrick
910 2004; Mank et al. 2008). It also appears that the intersexual conflict over *Obp56h* expression was
911 (at least partially) resolved with the Y-linkage of the *trans* regulators in III^M males.
912
913 Upregulation of the III^M alleles of *Md-spdt*, *Md-gd*, and/or *Md-Gr35* alone cannot fully resolve
914 the intersexual conflict over *Obp56h* regulation for at least two reasons. First, Y^M males do not
915 realize the fitness benefit because they do not carry a III^M chromosome. Therefore, Y^M males
916 remain evolutionarily constrained in their expression of *Obp56h*, suggesting that other fitness
917 benefits are responsible for the maintenance of the Y^M chromosome. Previous work suggests
918 that temperature-dependent fitness differences could be responsible—specifically Y^M males have
919 higher fitness at colder temperatures (Feldmeyer et al. 2008; Delclos et al. 2021). This suggests
920 that the III^M courtship advantage may be temperature-sensitive, which could create context-
921 dependent fitness effects that favor Y^M males in specific environments. This is consistent with
922 the prediction that genotype-by-environment interactions can maintain variation in sexually
923 selected traits across heterogeneous environments (Kokko and Heubel 2008). In contrast to that
924 prediction, we failed to detect effects of temperature on the III^M courtship advantage (Fig. 2),
925 although we did not comprehensively test for environmental effects. Additional work could
926 evaluate whether trade-offs across environments create opportunities for Y^M males to have
927 courtship advantages in specific contexts.
928
929 A second reason that the intersexual conflict over *Obp56h* regulation is not fully resolved is that
930 female house flies can carry the III^M chromosome. The male determining gene on the Y^M and
931 III^M chromosome (*Mdmd*) is a negative regulator of *Md-tra* splicing, and there is a dominant
932 allele (*Md-tra*^{D}) that is desensitized to the effects of *Mdmd* (Hediger et al. 2010; Sharma et al.
933 2017). This allows female house flies to carry Y^M or III^M if they also carry *Md-tra*^{D}. Females

934 with III^M and *Md-tra*^{D} in their genotype may suffer a fitness cost from upregulation of *Md-spdt*,
935 *Md-gd*, and *Md-Gr35* and/or downregulation of *Obp56h.* Therefore, even though the putatively
936 sexually antagonistic alleles are Y-linked, they can still be carried by females, leaving the
937 conflict unresolved. Consistent with this interpretation, previous work has determined that
938 sexually antagonistic fitness effects of the house fly proto-Y chromosomes could contribute to
939 the maintenance of the Y^M-III^M polymorphism within natural populations (Meisel et al. 2016;
940 Meisel 2021). Our results provide a mechanism by which the III^M chromosome could confer
941 sexually antagonistic effects.

942 ***Commonalities across sex chromosome evolution***

943 Our results have two additional implications for our understanding of the evolution of sex
944 chromosomes. First, our results provide evidence that gene duplication is important for the
945 acquisition of genes with male-specific functions during the evolution of Y chromosomes
946 (Koerich et al. 2008). The most central gene in the co-expression module, *LOC101887703*, has a
947 paralog in the house fly genome (*LOC101890114*) that is predicted to be on chromosome III.
948 These two genes are homologous to *D. melanogaster CG8745*, which we hypothesize is
949 negatively regulated by *Obp56h* based on conserved correlated expression in *D. melanogaster*
950 and house fly (Fig. 6). The two transcripts encoded by *LOC101887703* and *LOC101890114* are
951 <1% diverged in their nucleotide sequences, suggesting a recent duplication event. *CG8745* is
952 broadly expressed in many *D. melanogaster* tissues (Chintapalli et al. 2007), and broadly
953 expressed genes often give rise to paralogs with sex-specific expression (Meisel et al. 2009).
954
955 We hypothesize that one of the two paralogs (*LOC101887703* or *LOC101890114*) has evolved a
956 male-specific function, which resolved an intersexual conflict associated with the coregulation of
957 *CG8745* and *Obp56h*. It is not clear which gene is ancestral and which is derived. However, a
958 derived paralog with sex-specific expression is consistent with duplication of a broadly
959 expressed gene to resolve an inter-sexual conflict (Connallon and Clark 2011; Gallach and
960 Betrán 2011; Van Kuren and Long 2018). Those types of duplications often involve genes on sex
961 chromosomes—either when a broadly expressed X-linked gene gives rise to a an autosomal gene
962 with sex-specific expression, or when a gene is duplicated onto the Y chromosome and evolves
963 male-specific expression (Betrán et al. 2002; Emerson et al. 2004; Koerich et al. 2008; Meisel et
964 al. 2009; Hall et al. 2013; Mahajan and Bachtrog 2017; Ricchio et al. 2021). Consistent with this
965 evolutionary trajectory, the only identified differences between the house fly Y^M and X
966 chromosomes thus far have been autosome-to-Y duplications (Meisel et al. 2017). Future work
967 could address a potential sexually dimorphic subfunctionalization of the two *CG8745* paralogs in
968 the house fly genome.
969
970 Second, our results suggest that intersexual conflict may have been an important factor in the
971 convergent evolution of Muller element A into a sex chromosome in both *Drosophila* and house
972 fly. Muller element F was the X chromosome of the most recent common ancestor of *Drosophila*

973 and house fly, and element A became the *Drosophila* X chromosome after the divergence with
974 most other flies (Vicoso and Bachtrog 2013). Element A also recently became a sex chromosome
975 in house fly when it acquired an *Mdmd* gene, creating the III^M chromosome (Meisel et al. 2017;
976 Sharma et al. 2017; Son and Meisel 2021). Element A has become a sex chromosome in at least
977 two other dipteran lineages: *Glossina* (tsetse flies) and *Anopheles* mosquitoes (Pease and Hahn
978 2012; Vicoso and Bachtrog 2015), raising the possibility that element A is primed to be recruited
979 as a sex chromosome. Similar convergent sex-linkage of the same chromosomal region has been
980 observed in vertebrates (O'Meally et al. 2012; Furman and Evans 2016; Ezaz et al. 2017), which
981 could be explained by the same gene independently acquiring a sex determining allele in
982 multiple independent lineages (Takehana et al. 2014).
983
984 We hypothesize that element A may be convergently recruited to be a sex chromosome because
985 the *trans* regulatory connections with *Obp56h* create sexually antagonistic effects related to male
986 mating behavior. This differs from the hypothesized cause of convergent evolution of sex
987 chromosomes in vertebrates, which is based on the expectation that some chromosomes contain
988 an excess of genes that are predisposed to become sex determiners. Sexually antagonistic alleles
989 are expected to be an important selective force in the formation of new sex chromosomes
990 because sex-limited inheritance can resolve the intersexual conflict (van Doorn and Kirkpatrick
991 2007, 2010; Roberts et al. 2009). Our results suggest that an enrichment of genes involved in a
992 regulatory network with sexually antagonistic effects could promote the sex-linkage of the same
993 chromosome in distantly related species without convergent evolution of a master sex
994 determiner.

995 ***Conclusions***

996 We have identified *Obp56h* expression as the likely cause of differences in courtship
997 performance between Y^M and III^M house flies. We further identified multiple candidate *trans*
998 regulators of *Obp56h* on the III^M chromosome, one of which (*spdt*) we experimentally verified in
999 *D. melanogaster*. Our results demonstrate how the *trans* regulators of gene expression could
1000 have sexually antagonistic effects, which are resolved via the Y-linkage of those *trans* factors.
1001 This is the first time, to our knowledge, that a mechanism has been ascribed to the observation
1002 that the fitness effects of Y chromosomes can manifest via *trans* effects on autosomal genes,
1003 which affect male courtship and other sexually selected traits (Chippindale and Rice 2001;
1004 Morris et al. 2020; Kawamoto et al. 2021; Sandkam et al. 2021).

1005 **ACKNOWLEDGEMENTS**

1006 Rebecca Presley assisted with house fly courtship assays, and Poala Najera assisted with
1007 *Drosophila* courtship assays. This work was completed in part through the use of the Maxwell
1008 Cluster provided by the Research Computing Data Core at the University of Houston. This
1009 material is based upon work supported by the National Science Foundation under Grant No.

1010 DEB-1845686. Any opinions, findings, and conclusions or recommendations expressed in this
1011 material are those of the authors and do not necessarily reflect the views of the National Science
1012 Foundation.

1013 **REFERENCES**

1014 Abbott, J. K., A. K. Nordén, and B. Hansson. 2017. Sex chromosome evolution: historical insights and
1015 future perspectives. *Proc. Biol. Sci*. 284.

1016 Adhikari, K., J. H. Son, A. H. Rensink, J. Jaweria, D. Bopp, L. W. Beukeboom, and R. P. Meisel. 2021.
1017 Temperature-dependent effects of house fly proto-Y chromosomes on gene expression could be
1018 responsible for fitness differences that maintain polygenic sex determination. *Mol. Ecol*. 30:5704-
1019 5720.

1020 Arya, G. H., A. L. Weber, P. Wang, M. M. Magwire, Y. L. S. Negron, T. F. C. Mackay, and R. R. H.
1021 Anholt. 2010. Natural variation, functional pleiotropy and transcriptional contexts of odorant binding
1022 protein genes in *Drosophila melanogaster*. *Genetics* 186:1475–1485.

1023 Atkinson, D. 1996. Ectotherm life-history responses to developmental temperature. Pp. 183–204 *in* A. F.
1024 B. Ian A. Johnston, ed. Animals and Temperature: Phenotypic and Evolutionary Adaptation.

1025 Bachtrog, D., J. E. Mank, C. L. Peichel, M. Kirkpatrick, S. P. Otto, T.-L. Ashman, M. W. Hahn, J.
1026 Kitano, I. Mayrose, R. Ming, N. Perrin, L. Ross, N. Valenzuela, J. C. Vamosi, and Tree of Sex
1027 Consortium. 2014. Sex determination: why so many ways of doing it? *PLoS Biol*. 12:e1001899.

1028 Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. *J.*
1029 *Stat. Software* 67:1–48.

1030 Belvin, M. P., and K. V. Anderson. 1996. A conserved signaling pathway: the *Drosophila* toll-dorsal
1031 pathway. *Annu. Rev. Cell Dev. Biol*. 12:393–416.

1032 Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful
1033 approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol*. 57:289–300.

1034 Benoit, J. B., A. Vigneron, N. A. Broderick, Y. Wu, J. S. Sun, J. R. Carlson, S. Aksoy, and B. L. Weiss.
1035 2017. Symbiont-induced odorant binding proteins mediate insect host hematopoiesis. *eLife* 6.

1036 Berson, J. D., F. Garcia-Gonzalez, and L. W. Simmons. 2019a. Experimental evidence for the role of
1037 sexual selection in the evolution of cuticular hydrocarbons in the dung beetle, *Onthophagus taurus*.
1038 *J. Evol. Biol*. 32:1186–1193.
1039 Berson, J. D., M. Zuk, and L. W. Simmons. 2019b. Natural and sexual selection on cuticular
1040 hydrocarbons: a quantitative genetic analysis. *Proc. Biol. Sci*. 286:20190677.
1041 Betrán, E., K. Thornton, and M. Long. 2002. Retroposed new genes out of the X in *Drosophila*. *Genome*
1042 *Res*. 12:1854–1859.
1043 Beukeboom, L. W., and N. Perrin. 2014. The Evolution of Sex Determination. Oxford University Press.
1044 Bliss, C. I. 1926. Temperature characteristics for prepupal development in *Drosophila melanogaster*. *J.*
1045 *Gen. Physiol*. 9:467–495.
1046 Blomquist, G. J., and A.-G. Bagnères. 2010. Insect Hydrocarbons: Biology, Biochemistry, and Chemical
1047 Ecology. Cambridge University Press.
1048 Brankatschk, M., and S. Eaton. 2010. Lipoprotein particles cross the blood-brain barrier in *Drosophila*. *J.*
1049 *Neurosci*. 30:10441–10447.
1050 Bray, N. L., H. Pimentel, P. Melsted, and L. Pachter. 2016. Near-optimal probabilistic RNA-seq
1051 quantification. *Nat. Biotechnol*. 34:525–527.
1052 Brown, E. J., A. H. Nguyen, and D. Bachtrog. 2020. The *Drosophila* Y chromosome affects
1053 heterochromatin integrity genome-wide. *Mol. Biol. Evol*. 37:2808–2824.
1054 Brown, J. B., N. Boley, R. Eisman, G. E. May, M. H. Stoiber, M. O. Duff, B. W. Booth, J. Wen, S. Park,
1055 A. M. Suzuki, K. H. Wan, C. Yu, D. Zhang, J. W. Carlson, L. Cherbas, B. D. Eads, D. Miller, K.
1056 Mockaitis, J. Roberts, C. A. Davis, E. Frise, A. S. Hammonds, S. Olson, S. Shenker, D. Sturgill, A.
1057 A. Samsonova, R. Weiszmann, G. Robinson, J. Hernandez, J. Andrews, P. J. Bickel, P. Carninci, P.
1058 Cherbas, T. R. Gingeras, R. A. Hoskins, T. C. Kaufman, E. C. Lai, B. Oliver, N. Perrimon, B. R.
1059 Graveley, and S. E. Celniker. 2014. Diversity and dynamics of the *Drosophila* transcriptome. *Nature*
1060 512:393–399.

1061 Bryant, E. H. 1980. Geographic variation in components of mating success of the housefly, *Musca*
1062 *domestica L.*, in the United States. *Am. Nat.* 116:655–669.
1063 Burke, N. W., and G. I. Holwell. 2021. Increased male mating success in the presence of prey and rivals
1064 in a sexually cannibalistic mantis. *Behav. Ecol.* 32:574–579.
1065 Carroll, S. B. 2005. Evolution at two levels: on genes and form. *PLoS Biol*. 3:e245.
1066 Castel, S. E., A. Levy-Moonshine, P. Mohammadi, E. Banks, and T. Lappalainen. 2015. Tools and best
1067 practices for data processing in allelic expression analysis. *Genome Biol*. 16:195.
1068 Charlesworth, D., B. Charlesworth, and G. Marais. 2005. Steps in the evolution of heteromorphic sex
1069 chromosomes. *Heredity* 95:118–128.
1070 Chenoweth, S. F., and K. McGuigan. 2010. The genetic basis of sexually selected variation. Annu. Rev.
1071 Ecol. Evol. Syst. 41:81–101. *Annual Reviews*.
1072 Chintapalli, V. R., J. Wang, and J. A. T. Dow. 2007. Using FlyAtlas to identify better *Drosophila*
1073 *melanogaster* models of human disease. *Nat. Genet*. 39:715–720.
1074 Chippindale, A. K., and W. R. Rice. 2001. Y chromosome polymorphism is a strong determinant of male
1075 fitness in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA*. 98:5677–5682.
1076 Chung, H., and S. B. Carroll. 2015. Wax, sex and the origin of species: Dual roles of insect cuticular
1077 hydrocarbons in adaptation and mating. *Bioessays* 37:822–830.
1078 Connallon, T., and A. G. Clark. 2011. The resolution of sexual antagonism by gene duplication. *Genetics*
1079 187:919–937.
1080 Delclos, P. J., K. Adhikari, J. E. Cambric, A. G. Matuk, R. I. Presley, J. Tran, and R. P. Meisel. 2021.
1081 Thermal tolerance and preference are both consistent with the clinal distribution of house fly proto-Y
1082 chromosomes. *Evol. Lett.* evl3.248.
1083 Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R.
1084 Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.
1085 Edgar, R. C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space

1086 complexity. *BMC Bioinformatics* 5:113.

1087 Emerson, J. J., H. Kaessmann, E. Betrán, and M. Long. 2004. Extensive gene traffic on the mammalian X
1088 chromosome. *Science* 303:537–540.

1089 Ewen-Campen, B., D. Yang-Zhou, V. R. Fernandes, D. P. González, L.-P. Liu, R. Tao, X. Ren, J. Sun, Y.
1090 Hu, J. Zirin, S. E. Mohr, J.-Q. Ni, and N. Perrimon. 2017. Optimized strategy for in vivo Cas9-
1091 activation in *Drosophila*. *Proc. Natl. Acad. Sci. USA*. 114:9409–9414.

1092 Ezaz, T., K. Srikulnath, and J. A. M. Graves. 2017. Origin of amniote sex chromosomes: an ancestral
1093 super-sex chromosome, or common requirements? *J. Hered*. 108:94–105.

1094 Feldmeyer, B., M. Kozielska, and B. Kuijper. 2008. Climatic variation and the geographical distribution
1095 of sex-determining mechanisms in the housefly. *Evol. Ecol. Res.* 10:797-809.

1096 Findlay, G. D., X. Yi, M. J. MacCoss, and W. J. Swanson. 2008. Proteomics reveals novel *Drosophila*
1097 seminal fluid proteins transferred at mating. *PloS Biol.*. 6:e178.

1098 Fitzpatrick, M. J. 2004. Pleiotropy and the genomic location of sexually selected genes. *Am. Nat.*
1099 163:800–808.

1100 Foster, G. G., M. J. Whitten, C. Konovalov, J. T. A. Arnold, and G. Maffi. 1981. Autosomal genetic maps
1101 of the Australian Sheep Blowfly, *Lucilia cuprina dorsalis* R.-D. (Diptera: Calliphoridae), and
1102 possible correlations with the linkage maps of *Musca domestica* L. and *Drosophila melanogaster*
1103 (Mg.). *Genet. Res.* 37:55–69.

1104 Fox, J., and M. S. Carvalho. 2012. TheRcmdrPlugin.survivalPackage: Extending theRCommander
1105 Interface to Survival Analysis. *J. Stat. Software* 49:1-32.

1106 Fox, J., M. Friendly, and S. Weisberg. 2013. Hypothesis tests for multivariate linear models using the car
1107 package. *The R Journal* 5:39.

1108 Franco, M. G., P. G. Rubini, and M. Vecchi. 1982. Sex-determinants and their distribution in various
1109 populations of *Musca domestica L.* of Western Europe. *Genet. Res.* 40:279–293.

1110 Furman, B. L. S., and B. J. Evans. 2016. Sequential turnovers of sex chromosomes in African clawed

1111 frogs (*Xenopus*) suggest some genomic regions are good at sex determination. *G3* 6:3625–3633.
1112 Gallach, M., and E. Betrán. 2011. Intralocus sexual conflict resolved through gene duplication. *Trends*
1113 *Ecol. Evol*. 26:222–228.
1114 Gelbart, W. M., and D. B. Emmert. 2013. FlyBase High Throughput Expression Pattern Data.
1115 Graveley, B. R., A. N. Brooks, J. W. Carlson, M. O. Duff, J. M. Landolin, L. Yang, C. G. Artieri, M. J.
1116 van Baren, N. Boley, B. W. Booth, J. B. Brown, L. Cherbas, C. A. Davis, A. Dobin, R. Li, W. Lin, J.
1117 H. Malone, N. R. Mattiuzzo, D. Miller, D. Sturgill, B. B. Tuch, C. Zaleski, D. Zhang, M. Blanchette,
1118 S. Dudoit, B. Eads, R. E. Green, A. Hammonds, L. Jiang, P. Kapranov, L. Langton, N. Perrimon, J.
1119 E. Sandler, K. H. Wan, A. Willingham, Y. Zhang, Y. Zou, J. Andrews, P. J. Bickel, S. E. Brenner,
1120 M. R. Brent, P. Cherbas, T. R. Gingeras, R. A. Hoskins, T. C. Kaufman, B. Oliver, and S. E.
1121 Celniker. 2011. The developmental transcriptome of *Drosophila melanogaster*. *Nature* 471:473–479.
1122 Hall, A. B., Y. Qi, V. Timoshevskiy, M. V. Sharakhova, I. V. Sharakhov, and Z. Tu. 2013. Six novel Y
1123 chromosome genes in Anopheles mosquitoes discovered by independently sequencing males and
1124 females. *BMC Genomics* 14:273.
1125 Hamm, R. L., J.-R. Gao, G. G.-H. Lin, and J. G. Scott. 2009. Selective advantage for III^M males over Y^M
1126 males in cage competition, mating competition, and pupal emergence in *Musca domestica* L.
1127 (Diptera: Muscidae). *Environ. Entomol*. 38:499–504.
1128 Hamm, R. L., R. P. Meisel, and J. G. Scott. 2015. The evolving puzzle of autosomal versus Y-linked male
1129 determination in *Musca domestica*. *G3* 5:371–384.
1130 Hamm, R. L., T. Shono, and J. G. Scott. 2005. A cline in frequency of autosomal males is not associated
1131 with insecticide resistance in house fly (Diptera: Muscidae). *J. Econ. Entomol*. 98:171–176.
1132 Hediger, M., C. Henggeler, N. Meier, R. Perez, G. Saccone, and D. Bopp. 2010. Molecular
1133 characterization of the key switch *F* provides a basis for understanding the rapid divergence of the
1134 sex-determining pathway in the housefly. *Genetics* 184:155–170.
1135 Heinen-Kay, J. L., R. E. Nichols, and M. Zuk. 2020. Sexual signal loss, pleiotropy, and maintenance of a

1136 male reproductive polymorphism in crickets. *Evolution* 74:1002–1009.
1137 Johnston, S. E., J. Gratten, C. Berenos, J. G. Pilkington, T. H. Clutton-Brock, J. M. Pemberton, and J.
1138 Slate. 2013. Life history trade-offs at a single locus maintain sexually selected genetic variation.
1139 *Nature* 502:93–95.
1140 Kawamoto, M., Y. Ishii, and M. Kawata. 2021. Genetic basis of orange spot formation in the guppy
1141 (*Poecilia reticulata*). *BMC Ecol. Evol.* 21:211.
1142 Kirkpatrick, M., and M. J. Ryan. 1991. The evolution of mating preferences and the paradox of the lek.
1143 *Nature* 350:33–38.
1144 Koerich, L. B., X. Wang, A. G. Clark, and A. B. Carvalho. 2008. Low conservation of gene content in the
1145 *Drosophila* Y chromosome. *Nature* 456:949–951.
1146 Kokko, H., and K. Heubel. 2008. Condition-dependence, genotype-by-environment interactions and the
1147 lek paradox. *Genetica* 134:55–62.
1148 Konrad, K. D., T. J. Goralski, A. P. Mahowald, and J. L. Marsh. 1998. The gastrulation defective gene of
1149 *Drosophila melanogaster* is a member of the serine protease superfamily. *Proc. Natl. Acad. Sci.*
1150 *USA*. 95:6819–6824.
1151 Koštál, V., J. Korbelová, T. Štětina, R. Poupardin, H. Colinet, H. Zahradníčková, I. Opekarová, M. Moos,
1152 and P. Šimek. 2016. Physiological basis for low-temperature survival and storage of quiescent larvae
1153 of the fruit fly *Drosophila melanogaster*. *Sci. Rep*. 6:32346.
1154 Kozielska, M., B. Feldmeyer, I. Pen, F. J. Weissing, and L. W. Beukeboom. 2008. Are autosomal sex-
1155 determining factors of the housefly (*Musca domestica*) spreading north? *Genet. Res*. 90:157–165.
1156 Kudron, M. M., A. Victorsen, L. Gevirtzman, L. W. Hillier, W. W. Fisher, D. Vafeados, M. Kirkey, A. S.
1157 Hammonds, J. Gersch, H. Ammouri, M. L. Wall, J. Moran, D. Steffen, M. Szynkarek, S. Seabrook-
1158 Sturgis, N. Jameel, M. Kadaba, J. Patton, R. Terrell, M. Corson, T. J. Durham, S. Park, S. Samanta,
1159 M. Han, J. Xu, K.-K. Yan, S. E. Celniker, K. P. White, L. Ma, M. Gerstein, V. Reinke, and R. H.
1160 Waterston. 2018. The ModERN resource: Genome-wide binding profiles for hundreds of *Drosophila*

1161 and *Caenorhabditis elegans* transcription factors. *Genetics* 208:937–949.
1162 Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura. 2018. MEGA X: Molecular evolutionary
1163 genetics analysis across computing platforms. *Mol. Biol. Evol*. 35:1547–1549.
1164 Lande, R. 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution*
1165 34:292–305.
1166 Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution*
1167 37:1210–1226.
1168 Langfelder, P., and S. Horvath. 2008. WGCNA: an R package for weighted correlation network analysis.
1169 *BMC Bioinformatics* 9:559.
1170 Langfelder, P., P. S. Mischel, and S. Horvath. 2013. When is hub gene selection better than standard
1171 meta-analysis? *PLoS One* 8:e61505.
1172 Leek, J. T., W. E. Johnson, H. S. Parker, A. E. Jaffe, and J. D. Storey. 2012. The sva package for
1173 removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*
1174 28:882–883.
1175 Lemos, B., L. O. Araripe, and D. L. Hartl. 2008. Polymorphic Y chromosomes harbor cryptic variation
1176 with manifold functional consequences. *Science* 319:91–93.
1177 Lemos, B., A. T. Branco, and D. L. Hartl. 2010. Epigenetic effects of polymorphic Y chromosomes
1178 modulate chromatin components, immune response, and sexual conflict. *Proc. Natl. Acad. Sci. USA*.
1179 107:15826–15831.
1180 LeMosy, E. K., Y. Q. Tan, and C. Hashimoto. 2001. Activation of a protease cascade involved in
1181 patterning the *Drosophila* embryo. *Proc. Natl. Acad. Sci. USA*. 98:5055–5060.
1182 Li, A., and S. Horvath. 2007. Network neighborhood analysis with the multi-node topological overlap
1183 measure. *Bioinformatics* 23:222–231.
1184 Lockey, K. H. 1988. Lipids of the insect cuticle: origin, composition and function. *Comp. Biochem.*
1185 *Physiol. B: Comp. Biochem.* 89:595–645.

1186 Loeb, J., and J. H. Northrop. 1917. On the influence of food and temperature upon the duration of life. *J.*
1187 *Biol. Chem.* 32:103–121.
1188 Love, M. I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for
1189 RNA-seq data with DESeq2. *Genome Biol*. 15:550.
1190 Maere, S., K. Heymans, and M. Kuiper. 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of
1191 Gene Ontology categories in Biological Networks. *Bioinformatics* 21:3448–3449.
1192 Mahajan, S., and D. Bachtrog. 2017. Convergent evolution of Y chromosome gene content in flies. *Nat.*
1193 *Commun*. 8:785.
1194 Mank, J. E., D. J. Hosken, and N. Wedell. 2014. Conflict on the sex chromosomes: cause, effect, and
1195 complexity. *Cold Spring Harb. Perspect. Biol*. 6:a017715.
1196 Mank, J. E., L. Hultin-Rosenberg, M. Zwahlen, and H. Ellegren. 2008. Pleiotropic constraint hampers the
1197 resolution of sexual antagonism in vertebrate gene expression. *Am. Nat*. 171:35–43.
1198 Marcillac, F., Y. Grosjean, and J.-F. Ferveur. 2005. A single mutation alters production and
1199 discrimination of *Drosophila* sex pheromones. *Proc. Biol. Sci*. 272:303–309.
1200 May, C. E., A. Vaziri, Y. Q. Lin, O. Grushko, M. Khabiri, Q.-P. Wang, K. J. Holme, S. D. Pletcher, P. L.
1201 Freddolino, G. G. Neely, and M. Dus. 2019. High dietary sugar reshapes sweet taste to promote
1202 feeding behavior in *Drosophila melanogaster*. *Cell Rep*. 27:1675–1685.e7.
1203 McGraw, E. A., Y. H. Ye, B. Foley, S. F. Chenoweth, M. Higgie, E. Hine, and M. W. Blows. 2011. High-
1204 dimensional variance partitioning reveals the modular genetic basis of adaptive divergence in gene
1205 expression during reproductive character displacement. *Evolution* 65:3126–3137.
1206 McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D.
1207 Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The Genome Analysis Toolkit: a
1208 MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–
1209 1303.
1210 Meisel, R. P. 2021. The maintenance of polygenic sex determination depends on the dominance of fitness

1211 effects which are predictive of the role of sexual antagonism. *G3* jkab149.
1212 Meisel, R. P., T. Davey, J. H. Son, A. C. Gerry, T. Shono, and J. G. Scott. 2016. Is multifactorial sex
1213 determination in the house fly, *Musca domestica (L.)*, stable over time? *J. Hered.* 107:615–625.
1214 Meisel, R. P., C. A. Gonzales, and H. Luu. 2017. The house fly Y Chromosome is young and minimally
1215 differentiated from its ancient X Chromosome partner. *Genome Res.* 27:1417–1426.
1216 Meisel, R. P., M. V. Han, and M. W. Hahn. 2009. A complex suite of forces drives gene traffic from
1217 *Drosophila* X chromosomes. *Genome Biol. Evol*. 1:176–188.
1218 Meisel, R. P., and J. G. Scott. 2018. Using genomic data to study insecticide resistance in the house fly,
1219 *Musca domestica*. *Pestic. Biochem. Physiol*. 151:76–81.
1220 Meisel, R. P., J. G. Scott, and A. G. Clark. 2015. Transcriptome differences between alternative sex
1221 determining genotypes in the house fly, *Musca domestica*. *Genome Biol. Evol*. 7:2051–2061.
1222 Menuz, K., N. K. Larter, J. Park, and J. R. Carlson. 2014. An RNA-seq screen of the *Drosophila* antenna
1223 identifies a transporter necessary for ammonia detection. *PLoS Genet*. 10:e1004810.
1224 Mohapatra, P., and K. Menuz. 2019. Molecular profiling of the *Drosophila* antenna reveals conserved
1225 genes underlying olfaction in insects. *G3* 9:3753–3771.
1226 Mokashi, S. S., V. Shankar, J. A. Johnstun, W. Huang, T. F. C. Mackay, and R. R. H. Anholt. 2021.
1227 Systems genetics of single nucleotide polymorphisms at the *Drosophila Obp56h* locus. *bioRxiv*
1228 2021-06.
1229 Morris, J., I. Darolti, W. van der Bijl, and J. E. Mank. 2020. High-resolution characterization of male
1230 ornamentation and re-evaluation of sex linkage in guppies. *Proc. Biol. Sci.* 287:20201677.
1231 Olafson, P. U., S. Aksoy, G. M. Attardo, G. Buckmeier, X. Chen, C. J. Coates, M. Davis, J. Dykema, S. J.
1232 Emrich, M. Friedrich, C. J. Holmes, P. Ioannidis, E. N. Jansen, E. C. Jennings, D. Lawson, E. O.
1233 Martinson, G. L. Maslen, R. P. Meisel, T. D. Murphy, D. Nayduch, D. R. Nelson, K. J. Oyen, T. J.
1234 Raszick, J. M. C. Ribeiro, H. M. Robertson, A. J. Rosendale, T. B. Sackton, P. Saelao, S. L. Swiger,
1235 S.-H. Sze, A. M. Tarone, D. B. Taylor, W. C. Warren, R. M. Waterhouse, M. T. Weirauch, J. H.

1236 Werren, R. K. Wilson, E. M. Zdobnov, and J. B. Benoit. 2021. The genome of the stable fly,
1237 *Stomoxys calcitrans*, reveals potential mechanisms underlying reproduction, host interactions, and
1238 novel targets for pest control. *BMC Biol*. 19:1-31.
1239 Olafson, P. U., and C. A. Saski. 2020. Chemosensory-related gene family members of the horn fly,
1240 (Diptera: Muscidae), identified by transcriptome analysis. *Insects* 11.
1241 O'Meally, D., T. Ezaz, A. Georges, S. D. Sarre, and J. A. M. Graves. 2012. Are some chromosomes
1242 particularly good at sex? Insights from amniotes. *Chromosome Res*. 20:7–19.
1243 Otte, T., M. Hilker, and S. Geiselhardt. 2018. Phenotypic plasticity of cuticular hydrocarbon profiles in
1244 insects. *J. Chem. Ecol*. 44:235–247.
1245 Pan, J. W., J. McLaughlin, H. Yang, C. Leo, P. Rambarat, S. Okuwa, A. Monroy-Eklund, S. Clark, C. D.
1246 Jones, and P. C. Volkan. 2017. Comparative analysis of behavioral and transcriptional variation
1247 underlying CO2 sensory neuron function and development in *Drosophila*. *Fly* 11:239–252.
1248 Park, S. K., K. J. Mann, H. Lin, E. Starostina, A. Kolski-Andreaco, and C. W. Pikielny. 2006. A
1249 *Drosophila* protein specific to pheromone-sensing gustatory hairs delays males' copulation attempts.
1250 *Curr. Biol*. 16:1154–1159.
1251 Pease, J. B., and M. W. Hahn. 2012. Sex chromosomes evolved from independent ancestral linkage
1252 groups in winged insects. *Mol. Biol. Evol*. 29:1645–1653.
1253 R Core Team. 2019. R: A Language and Environment for Statistical Computing. R Foundation for
1254 Statistical Computing, Vienna, Austria.
1255 Ricchio, J., F. Uno, and A. B. Carvalho. 2021. New genes in the *Drosophila* Y chromosome: lessons from
1256 *D. willistoni*. *Genes* 12.
1257 Rice, W. R. 1996. Evolution of the Y sex chromosome in animals. *Bioscience* 46:331–343.
1258 Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth. 2015. limma powers
1259 differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*
1260 43:e47.

1261 Roberts, R. B., J. R. Ser, and T. D. Kocher. 2009. Sexual conflict resolved by invasion of a novel sex
1262 determiner in Lake Malawi cichlid fishes. *Science* 326:998–1001.
1263 Rowe, L., and H. D. Rundle. 2021. The alignment of natural and sexual selection. Annu. Rev. Ecol. Evol.
1264 Syst. 52:499–517. *Annual Reviews*.
1265 Sandkam, B. A., P. Almeida, I. Darolti, B. L. S. Furman, W. van der Bijl, J. Morris, G. R. Bourne, F.
1266 Breden, and J. E. Mank. 2021. Extreme Y chromosome polymorphism corresponds to five male
1267 reproductive morphs of a freshwater fish. *Nat. Ecol. Evol.* 5:939–948.
1268 Schaeffer, S. W. 2018. Muller "elements" in *Drosophila*: How the search for the genetic basis for
1269 speciation led to the birth of comparative genomics. *Genetics* 210:3–13.
1270 Scott, J. G., P. Sridhar, and N. Liu. 1996. Adult specific expression and induction of cytochrome P450lpr
1271 in house flies. *Arch. Insect Biochem. Physiol*. 31:313–323.
1272 Scott, J. G., W. C. Warren, L. W. Beukeboom, D. Bopp, A. G. Clark, S. D. Giers, M. Hediger, A. K.
1273 Jones, S. Kasai, C. A. Leichter, M. Li, R. P. Meisel, P. Minx, T. D. Murphy, D. R. Nelson, W. R.
1274 Reid, F. D. Rinkevich, H. M. Robertson, T. B. Sackton, D. B. Sattelle, F. Thibaud-Nissen, C.
1275 Tomlinson, L. van de Zande, K. K. O. Walden, R. K. Wilson, and N. Liu. 2014. Genome of the
1276 house fly, *Musca domestica L.*, a global vector of diseases with adaptations to a septic environment.
1277 *Genome Biol*. 15:466.
1278 Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and
1279 T. Ideker. 2003. Cytoscape: a software environment for integrated models of biomolecular
1280 interaction networks. *Genome Res*. 13:2498–2504.
1281 Sharma, A., S. D. Heinze, Y. Wu, T. Kohlbrenner, I. Morilla, C. Brunner, E. A. Wimmer, L. van de
1282 Zande, M. D. Robinson, L. W. Beukeboom, and D. Bopp. 2017. Male sex in houseflies is
1283 determined by *Mdmd*, a paralog of the generic splice factor gene *CWC22*. *Science* 356:642–645.
1284 Shiao, M.-S., W.-L. Fan, S. Fang, M.-Y. J. Lu, R. Kondo, and W.-H. Li. 2013. Transcriptional profiling
1285 of adult *Drosophila* antennae by high-throughput sequencing. *Zool. Stud*. 52:1–10.

1286 Shono, T., and J. G. Scott. 2003. Spinosad resistance in the housefly, *Musca domestica*, is due to a
1287 recessive factor on autosome 1. *Pestic. Biochem. Physiol.* 75:1–7.
1288 Shorter, J. R., L. M. Dembeck, L. J. Everett, T. V. Morozova, G. H. Arya, L. Turlapati, G. E. St Armour,
1289 C. Schal, T. F. C. Mackay, and R. R. H. Anholt. 2016. *Obp56h* modulates mating behavior in
1290 *Drosophila melanogaster*. *G3* 6:3335–3342.
1291 Sink, H., E. J. Rehm, L. Richstone, Y. M. Bulls, and C. S. Goodman. 2001. *sidestep* encodes a target-
1292 derived attractant essential for motor axon guidance in *Drosophila*. *Cell* 105:57–67.
1293 Son, J. H., T. Kohlbrenner, S. Heinze, L. W. Beukeboom, D. Bopp, and R. P. Meisel. 2019. Minimal
1294 effects of proto-Y chromosomes on house fly gene expression in spite of evidence that selection
1295 maintains stable polygenic sex determination. *Genetics* 213:313–327.
1296 Son, J. H., and R. P. Meisel. 2021. Gene-level, but not chromosome-wide, divergence between a very
1297 young house fly proto-Y chromosome and its homologous proto-X chromosome. *Mol. Biol. Evol*.
1298 38:606–618.
1299 Sun, J. S., N. K. Larter, J. S. Chahda, D. Rioux, A. Gumaste, and J. R. Carlson. 2018a. Humidity response
1300 depends on the small soluble protein *Obp59a* in *Drosophila*. *eLife* 7:e39249.
1301 Sun, J. S., S. Xiao, and J. R. Carlson. 2018b. The diverse small proteins called odorant-binding proteins.
1302 *Open Biol*. 8:180208.
1303 Swarup, S., T. V. Morozova, S. Sridhar, M. Nokes, and R. R. H. Anholt. 2014. Modulation of feeding
1304 behavior by odorant-binding proteins in *Drosophila melanogaster*. *Chemical Senses* 39:125-132.
1305 Takehana, Y., M. Matsuda, T. Myosho, M. L. Suster, K. Kawakami, T. Shin-I, Y. Kohara, Y. Kuroki, A.
1306 Toyoda, A. Fujiyama, S. Hamaguchi, M. Sakaizumi, and K. Naruse. 2014. Co-option of *Sox3* as the
1307 male-determining factor on the Y chromosome in the fish *Oryzias dancena*. *Nat. Commun.* 5:4157.
1308 Thomas, M. L., and L. W. Simmons. 2009. Sexual selection on cuticular hydrocarbons in the Australian
1309 field cricket, *Teleogryllus oceanicus*. *BMC Evol. Biol.* 9:1-12.
1310 Tomita, T., and Y. Wada. 1989. Multifactorial sex determination in natural populations of the housefly

1311 (*Musca domestica*) in Japan. *Japanese J. Genetics* 64:373–382.

1312 Turelli, M., and N. H. Barton. 2004. Polygenic variation maintained by balancing selection: pleiotropy,
1313 sex-dependent allelic effects and G x E interactions. *Genetics* 166:1053–1079.

1314 Valanne, S., J.-H. Wang, and M. Rämet. 2011. The *Drosophila Toll* signaling pathway. *J. Immunol.*
1315 186:649–656.

1316 Van De Bor, V., G. Zimniak, L. Papone, D. Cerezo, M. Malbouyres, T. Juan, F. Ruggiero, and S. Noselli.
1317 2015. Companion blood cells control ovarian stem cell niche microenvironment and homeostasis.
1318 *Cell Rep*. 13:546–560.

1319 van Doorn, G. S., and M. Kirkpatrick. 2010. Transitions between male and female heterogamety caused
1320 by sex-antagonistic selection. *Genetics* 186:629–645.

1321 van Doorn, G. S., and M. Kirkpatrick. 2007. Turnover of sex chromosomes induced by sexual conflict.
1322 *Nature* 449:909–912.

1323 Van Kuren, N. W., and M. Long. 2018. Gene duplicates resolving sexual conflict rapidly evolved
1324 essential gametogenesis functions. *Nat. Ecol. Evol.* 2:705–712.

1325 Vicoso, B., and D. Bachtrog. 2015. Numerous transitions of sex chromosomes in Diptera. *PLoS Biol.*
1326 13:e1002078.

1327 Vicoso, B., and D. Bachtrog. 2013. Reversal of an ancient sex chromosome to an autosome in
1328 *Drosophila*. *Nature* 499:332–335.

1329 Weller, G. L., and G. G. Foster. 1993. Genetic maps of the sheep blowfly *Lucilia cuprina*: linkage-group
1330 correlations with other dipteran genera. *Genome* 36:495–506.

1331 Zhang, Y. Q., C. K. Rodesch, and K. Broadie. 2002. Living synaptic vesicle marker: synaptotagmin-GFP.
1332 *Genesis* 34:142–145.

1333 Zhou, J. J. 2010. Odorant-binding proteins in insects. *Vitamins & Hormones* 83:241–272.

1334