An investigation of the sex-specific genetic architecture of fitness in *Drosophila melanogaster*

Amardeep Singh, Asad Hasan and Aneil F. Agrawal[1]

Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, Ontario, Canada, M5S 3B2

[1]*E-mail: a.agrawal@utoronto.ca*

Abstract

In dioecious populations, the sexes employ divergent reproductive strategies to maximize fitness and, as a result, genetic variants can affect fitness differently in males and females. Moreover, recent studies have highlighted an important role of the mating environment in shaping the strength and direction of sex-specific selection. Here, we measure the sex-specific outbred fitness of 357 lines from the *Drosophila* Synthetic Population Resource (DSPR) in two different mating environments. We analyze the data using three different approaches to gain insight into the sex-specific genetic architecture for fitness: classical quantitative genetics, genomic associations, and a mutational burden approach. The quantitative genetics analysis finds that, on average segregating genetic variation in this population has concordant fitness effects both across the sexes and across mating environments. We do not find specific genomic regions with strong associations with either sexually antagonistic (SA) or sexually concordant (SC) fitness effects, yet there is modest evidence of an excess of genomic regions with weak associations, both with SA and SC fitness effects. Our examination of mutational burden indicates stronger selection against indels and loss-of-function variants in females than males.

Introduction

The genetic architecture of fitness variation (e.g., the number and identity of loci, and their effect sizes in different contexts) remains poorly understood. Given that any locus in the genome has the potential to contribute to fitness, the genetic architecture of fitness is likely highly polygenic. However, this does not preclude the possibility that some variants have reasonably large effects. Investigating the genetic architecture of fitness may provide useful clues as to the forces responsible for the maintenance of segregating fitness variants (e.g., mutation-selection balance or various forms of balancing selection; Mackay 2001; Barton and Keightley 2002; Frankham 2009)

In dioecious populations, the genetic architecture of phenotypic traits, including fitness may differ substantially between the sexes. Due to anisogamy, the sexes maximize fitness in different ways: in many systems, male fitness is limited by the number of mates that they acquire whereas female fitness is limited by the amount of resources that they can acquire and efficiently convert into offspring (Bateman 1948; Parker 1979; Wade 1979; Wade and Arnold 1980). Thus, traits necessary to maximize fitness, and by extension the genetic architecture underpinning those traits, can differ across the sexes.

Sex differences in the genetic architecture of fitness depend on the sex-specific effect size and direction of segregating variants. For example, some variants may have sex-limited effects because they are sex-linked or contribute to the expression of sex-limited traits (e.g., seminal fluid proteins). Such variants can be important contributors to the variance in fitness in one sex but may have little or no influence on fitness in the other. Variants with purely sex-limited fitness effects may be relatively rare. Instead, alleles may often contribute to variance in fitness in both sexes. The evolutionary fate of such variants will depend on the strength and direction of selection they experience when expressed in each sex. In some cases, alleles at a locus may have opposing,

sexually antagonistic, effects on fitness resulting in so-called 'intra-locus sexual conflict' (reviewed in Bonduriansky and Chenoweth 2009). Under specific conditions, such variants can be maintained by balancing selection, particularly when the effect sizes of variants are large (Kidwell et al. 1977; Connallon and Clark 2012). However, most genes are thought to serve similar functions in both sexes and disruptions to those functions are likely deleterious for both sexes (Rowe and Houle 1996; Whitlock and Agrawal 2009). While sexually concordant selection should remove genetic variation at these sites (Long et al. 2012; Connallon and Clark 2013), this class of loci may serve as a large mutational target and thus contribute substantially to the standing genetic variation in fitness via mutation-selection balance.

The primary objective of this study is to better understand fitness variation within and across the sexes. Ideally, one would catalog every variant and measure its fitness effect in each sex, but that ideal is not logistically feasible. Instead, we pursue three different approaches that each offer a different lens through which to view the relationship of the genetic architecture for fitness across the sexes.

First, we used a classic quantitative genetics approach to provide a summary of how genetic architecture of fitness is related between the sexes. This approach has been used in a number of laboratory (e.g., Vieira et al. 2000; Wayne et al. 2001; Pischedda and Chippindale 2006; Long et al. 2009; Collet et al. 2016) and wild (e.g., Kruuk et al. 2000; Merilä and Sheldon 2000; McCleery et al. 2004; Foerster et al. 2007) populations and is useful for asking how much additive genetic variation in sex-specific fitness exists in a population. In addition, we used this approach to estimate the intersexual genetic correlation for fitness (r_{mf}), which serves as a useful indicator of the overall concordance or antagonism of segregating genetic variation in a population. This correlation can be used to quantify the extent to which improvement in the fitness of one sex will

104 help or hinder the improvement of the other, at least in the short term. Empirical estimates of r_{mf}
105 have been mixed: although a survey of animal species found that, on average, estimates of r_{mf}
106 were small but positive (Poissant et al. 2010), several studies have reported negative estimates of
107 r_{mf} of substantial magnitude (e.g., Chippindale et al. 2001; Fedorka and Mousseau 2004;
108 Brommer et al. 2007; Punzalan et al. 2014). The power and limitation of the intersexual correlation
109 is that it is a summary statistic that is predictive of short-term evolution that neither requires nor
110 reveals information about individual allelic effects. Indeed, the underlying 'genetic details' matter
111 little with respect to short-term evolution but can substantially affect evolutionary outcomes on
112 longer time-scales.

113 Next, we used genomic association mapping techniques in an attempt to gain resolution on
114 specific loci underlying sex-specific fitness. Association mapping provides a route to identifying
115 individual allelic effects and, in principle, can be used to find the number and magnitude of
116 mutations with sexually concordant, antagonistic, and sex-limited effects. Recognizing that the
117 genetic basis of fitness is likely to be highly polygenic, we are unlikely to have the statistical power
118 to identify a multitude of small effect variants. However, we were particularly interested in
119 identifying loci contributing to sexual antagonism and such variants maintained by balancing
120 selection are expected to be of larger effects (Kidwell et al. 1977; Connallon and Clark 2012),
121 possibly detectable via genome association mapping. This approach has recently proven fruitful
122 in identifying candidate sexually antagonistic loci in this species (Ruzicka et al. 2019).

123 Though association mapping is severely limited in its ability to identify small effect
124 variants, other approaches can provide insights. When sequence and fitness data are both available
125 for a collection of genotypes, then one approach is to quantify the aggregate effect of specified
126 classes that are suspected to affect fitness *a priori* (Yang et al. 2017; Brown and Kelly 2019). For

127 example, one can use this approach to ask how the total number of loss-of-function variants (LoFs)
128 in a genome affects fitness, which is equivalent to quantifying the average strength of selection
129 against segregating LoFs. We are particularly interested in comparing the selection strength in
130 males versus females. Strong competition for access to females has been postulated to increase the
131 strength of selection on males relative to females on sexually concordant variants (reviewed
132 Whitlock and Agrawal 2009), potentially increasing population mean fitness by reducing mutation
133 load (Manning 1984; Agrawal 2001; Siller 2001) and accelerating adaptation (Lorch et al. 2003).
134 Experimental studies using phenotypic marker mutations (e.g., Whitlock and Bourguet 2000;
135 Pischedda and Chippindale 2005; Stewart et al. 2005; Sharp and Agrawal 2008; Hollis et al. 2009)
136 and mutation accumulation lines (e.g., Enders and Nunney 2010; Mallet et al. 2011; Mcguigan et
137 al. 2011; Sharp and Agrawal 2013; Grieshop et al. 2016) tend to support the notion of stronger
138 selection in males than females, though there are potential concerns with these approaches (see
139 Discussion).

140 The goals of this study were twofold. The first was to more fully understand the sex-
141 specific genetic architecture of fitness using the three approaches described above. The second
142 goal of this study was to investigate how the genetic architecture of fitness of each sex, as well as
143 the relationship of these architectures across the sexes, was affected by the ecological context in
144 which sexual interactions occur. Here, we estimated the sex-specific outbred fitness for 357
145 *Drosophila melanogaster* recombinant inbred line (RIL) haplotypes from the *Drosophila*
146 Synthetic Population Resource (DSPR; King et al. 2012) in two mating environments: traditional
147 fly culturing vials and larger cages that represent lower density and more structurally complex
148 environments. Previous work has shown males and females interact differently in environments
149 similar to the two used here (Yun et al. 2017; MacPherson et al. 2018). Male harm and sexual

150 conflict are strong features of sexual interactions in standard fly vials but not in the more complex
151 cages; these differences affect the rates of nonsexual adaptation and purging (Singh et al. 2017;
152 Yun et al. 2018).

153 Our quantitative genetic analysis recovers appreciable levels of additive genetic variation
154 in fitness for both male and female fitness and finds that the fitness consequences of this genetic
155 variation tends to be concordant across the sexes as well as across environments. Our association
156 mapping analysis did not reveal any statistically convincing associations between fitness and any
157 specific genomic region. However, there is some evidence that both sexually concordant and
158 sexually antagonistic regions exist. Our mutation burden analysis recovers a significant negative
159 association between the burden of putatively deleterious variants and female, but not male, fitness
160 in both mating environments.

161

162 **Methods**

163 DSPR Inbred Genotypes

164 Recombinant inbred line (RIL) genotypes used for measuring sex-specific outbred fitness
165 were obtained from the *Drosophila* Synthetic Population Resource (DSPR). The creation of these
166 lines has been described in detail elsewhere (King et al. 2012a). Briefly, eight parental genotypes
167 from geographically disparate regions were crossed in a round robin design to generate an outbred
168 population that was allowed to recombine for 50 generations. Inbred genotypes were generated
169 from this outbred, synthetic population following 25 generations of fullsib mating. Following low-
170 resolution RADseq genotyping of inbred genotypes, a hidden Markov model (HMM) was
171 implemented to determine the probable parental ancestry at regularly spaced 10kb intervals (or

172 haplotype blocks) along the genome of each line. The parental ancestry of the vast majority of
173 10kb blocks in most lines can be assigned to a parental source with high certainty. Here, we
174 estimated sex-specific fitness of 357 lines. Although the majority of lines were assayed for fitness
175 in both sexes and both environments, this was not true of all lines for logistical reasons (see Table
176 S1 for number of lines assayed per sex and environment).

177 Measuring Fitness of DSPR Genotypes

178 To measure sex-specific outbred fitness of focal genotypes we first crossed females of each
179 focal genotype to males from a common outbred population that was generated by mixing all
180 DSPR founder genotypes. Focal males and females were then assayed in separate competitive
181 fitness assays. Specifically, for each outbred genotype, we combined four focal males/females with
182 twelve competitors of the same sex from an outbred laboratory population homozygous for a
183 dominant bw^D marker. Focal and competitor genotypes competed for access to sixteen unmarked
184 individuals of the opposite sex that came from an outbred laboratory population. For logistical
185 reasons the source population for these individuals differed between male and female fitness
186 assays: in male fitness assays we used females from an outbred laboratory population originally
187 captured from a West African population (*Dahomey* stock) while in the female fitness assays we
188 used males from an outbred stock generated by mixing DSPR founder genotypes.

189 All flies were collected at approximately the same age (~2-3 days post eclosion) and held
190 separately by sex in vials sprinkled with live yeast in groups of sixteen for 24 hours prior to the
191 start of fitness assays. All flies for a replicate were then combined and allowed to interact for 6
192 days in either a standard *Drosophila* culture vials (28 x 95 mm, approximately 60 mL) that
193 contained 7ml of food sprinkled with live yeast or a larger, complex mating cages. The cages were
194 polypropylene ULINE® deli containers (114.3 x 76.2 mm, approximately 500 mL) with two food

195 cups containing 7mL of food each, sprinkled with live yeast. Replicates were maintained in a
196 12L:12D light cycle at 25°C. Fresh media, sprinkled with live yeast was provided on day one and
197 three (fresh media was not provided on day three in cage assays due to logistical reasons). On day
198 six, all females were removed and, to reduce subsequent larval densities, divided approximately
199 evenly into two vials sprinkled with live yeast and allowed to oviposit for ~24 hours. Starting ten
200 days later, and repeating for a total of three days, offspring from the two oviposition vials were
201 collected, pooled and frozen daily. The vast majority of offspring emerged on the first two days of
202 collection. Following collection, we randomly sampled at least 100 offspring and calculated the
203 proportion of total offspring with wildtype (red) eye colour in this sample. We have five replicate
204 assays for most combinations of line, sex, and environment. Due to human error, some lines have
205 more than five replicates while others have fewer. Fitness assays were conducted in a blocked
206 design; for analyses, fitness estimates were standardized by dividing by the block mean fitness for
207 that sex and environment and zero centered.

208 Quantitative Genetic Analysis of Fitness in the DSPR

209 To estimate the sex-specific heritability in fitness in each mating environment, we fit
210 separate mixed models for each mating environment to estimate the sex-specific additive genetic
211 and residual variances. We fit the model $W_{ijk} = \mu + \phi_{jk} + e_{ijk}$ where W_{ijk} is the within-block
212 standardized and zero-centered fitness of the ith replicate of sex j of genotype k (i.e., RIL k). μ is
213 the grand mean relative fitness of all lines, ϕ_{jk} is the random genetic effect of sex j of genotype k
214 and follows a bivariate Gaussian distribution such that $\phi_{jk} \sim N(0, \boldsymbol{G})$ where

215 $$\boldsymbol{G} = \begin{bmatrix} \sigma^2_{RIL:m} & \sigma_{RIL:m,f} \\ \sigma_{RIL:m,f} & \sigma^2_{RIL:f} \end{bmatrix}$$

is a genetic covariance matrix. The diagonal elements are the variances among RILs for male and female fitness, which are proportional to the additive genetic variances; the off-diagonal is the among-RIL intersexual covariance, which is proportional to the additive genetic covariance. (Our estimated matrix is proportional, rather than equal, to the traditional $\boldsymbol{G}$ matrix because our unit of observation is the fitness of groups of four focal individuals sharing a RIL haploid genome rather than a single individual; see Supplemental Material for derivation). e_{ijk} specifies the residual variance of the ith replicate of sex j of genotype k and are identical and independently Gaussian distributed such that $e_{ijk} \sim N(0, \sigma_{ej}^2)$ where σ_{ej}^2 is the residual variance in fitness in sex j. As shown in the Supplementary Material, heritability of fitness in sex j as

$$h_j^2 = \frac{\sigma_{RIL:j}^2}{2\left(\sigma_{ej}^2 + \frac{\sigma_{RIL:j}^2}{4}\right)}$$

The mixed models were fit using *MCMCglmm* (Hadfield 2010) separately for each mating environment. Models were fit using uninformative priors and run for 1,000,000 iterations after a burn-in period of 10,000 iterations with a thinning interval of 1000. Posterior estimates of variance components from each run were used to calculate the mean and 95% credibility intervals (CI) for each parameter of interest (i.e., $\sigma_{RIL,m}^2$, $\sigma_{RIL,f}^2, h_m^2$ and h_f^2). Specifically, CIs reported were calculated using the *HPDInterval* function in *MCMCglmm*. Estimates of variance components were also used to estimate the intersexual genetic correlation for fitness (r_{mf}) separately within each mating environment as

$$r_{mf} = \frac{\sigma_{RIL:m,f}}{\sqrt{\sigma_{RIL:m}^2\, \sigma_{RIL:f}^2}}$$

235 Though our estimates of r_{mf} are intended to represent the intersexual genetic correlation
236 with respect to "adult" fitness, they are indirectly affected by juvenile survivorship because we
237 assess the fitness of adult males and adult females by measuring the number of offspring that
238 survive to the adult stage. Genetic variation for offspring survivorship will upwardly bias our
239 estimate of the intersexual correlation for adult fitness. However, we expect this effect to be small
240 for several reasons. First, we attempted to keep offspring density from being high so that
241 survivorship should be high. Second, the contribution of the genetic variance in offspring
242 survivorship to the variance in focal adult fitness will be diluted by 50% because offspring only
243 have half the genes from the focal parent. Third, previous studies have indicated that genetic
244 variation for juvenile survivorship is small, especially compared to that of adult fitness components
245 (e.g., Leips and Mackay 2000; Sharp and Agrawal 2018).

246 As a separate question, we were interested in the relationship between the genetic
247 architectures of a given sex across the two mating environments. Using the fitness data for a single
248 sex from each of the two environments, we employed a similar approach to estimate the within-
249 sex cross-environment genetic correlation for fitness (i.e., $r_{env,m}$ and $r_{env,f}$ for males and females,
250 respectively).

251 Association Mapping of Fitness and Indices of Sexual Antagonism and Concordance

252 We briefly outline methods for genome association mapping and provide details below. To
253 identify genome associations with sex-specific fitness as well as loci underpinning sexually
254 antagonistic and concordant variation, we first identified univariate traits for fitness and indices of
255 sexual antagonism and concordance. Next, we defined a statistical model for mapping genotypic

256 variation to trait variation and a significance threshold was established to identify loci putatively
257 associated with fitness or axes of antagonism/concordance.

258 To map phenotypic variation in male and female fitness we calculated the mean relative
259 fitness, separately for each sex, of each RIL. To associate loci with sexual antagonism and sexual
260 concordance, we defined an 'axis of antagonism' (i.e., SA Axis) and an 'axis of concordance' (i.e.,
261 SC Axis) following the method by Berger et al. (2014). Specifically, the matrix of coordinates of
262 mean relative male and female fitness of each RIL are rotated such that estimates are projected
263 onto axes that are 45° and 315° about the horizontal, which correspond to the axis of concordance
264 and antagonism respectively. The value of a RIL along the axis of antagonism is calculated as:

265 $$SA = x\cos(45°) - y\sin(45°)$$

266 where x and y are the estimates of mean female and male relative fitness. RILs with above (below)
267 average values for female fitness but the opposite for male fitness are expected to have positive
268 (negative) values of SA. The value of a RIL along the axis of concordance is computed as:

269 $$SC = x\sin(45°) + y\cos(45°)$$

270 RILs with above (below) average fitness for both sexes are expected to have positive (negative)
271 values of SC. Our primary interest was with respect to the SA index, but we independently mapped
272 four traits—SA index and SC index, as well as male and female fitnesses—separately for each
273 environment (i.e., 8 total traits).

274 A feature of the DSPR is that the mosaic haplotype structure of each RIL is known to a
275 high degree of certainty in most cases. Specifically, at 10kb intervals along the genome of each
276 RIL, the additive probability of each of the eight parental genotypes is known from a Hidden
277 Markov Model (King et al. 2012a). To associate loci with each of the eight traits we regressed, at

278 each 10kb haplotype block, the RIL trait values against the additive probability for each of the
279 eight parental lines. Specifically, we fit the mixed model:

280 $$\bar{Z}_i = \mu + \sum_{j=1}^{l} \beta_j G_{ij} + \psi_i + e_{ij}$$

281 where $\bar{Z}_i$ is the mean trait value of the *i*th RIL and μ is the grand trait mean across all sampled
282 RILs, G_{ij} is fixed effect of the additive probability of the haplotype block of the *i*th RIL having
283 descended from parental genotype *j*, β_j is the additive genetic effect of the *j*th parental genotype
284 on the trait, ψ_i is the random effect of RIL genotype that is assumed to follow a Gaussian
285 distribution such that $\psi_i \sim \sigma_{kl} \sim N(0, \boldsymbol{K})$ where $\boldsymbol{K}$ is the relatedness matrix of the DSPR lines (see
286 below), *l* is the number of parental genotypes (max = 8), that contributed a cumulative additive
287 probability of at least 5% at each haplotype block *i* across all RIL genotypes (see below) and e_{ij}
288 specifies the residual variance of the *i*th RIL at the *j*th haplotype block and is assumed to follow a
289 Gaussian distribution with a mean of zero and unknown variance. At each 10kb haplotype block
290 *i*, we excluded any RIL genotypes whose most probable parental ancestor did not have a
291 cumulative additive probability of at least 5% among all RIL genotypes. *P*-values were obtained
292 from a likelihood ratio test against a null model where $\beta_j = 0$ for all *j*.

293 To estimate the genetic relatedness matrix of lines assayed, we calculated the average
294 pairwise relatedness for all 357 inbred genotypes over all haplotype blocks. Specifically, for each
295 pair of genotypes *i* and *j* we calculated the genetic relatedness (κ_{ij}) as

296 $$\kappa_{ij} = \frac{\sum_{n=1}^{m} \sum_{k=1}^{8} p_{ikn} p_{jkn}}{m}$$

297 where p_{ikn} is the additive probability that haplotype block (i.e., a 10kb interval) n in genotype i
298 originated from ancestral genotype k and m is the total number of haplotype blocks where
299 genotypic information is present in both inbred genotypes.

300 As reported below, we were unable to find strong statistical support for individual
301 haplotype blocks contributing to sexual concordance or antagonism. A plausible explanation is
302 that effect sizes are low relative to environmental variation, so it becomes unlikely to find the very
303 low *p*-values required when controlling for multiple testing. To explore the data, we shifted our
304 focus to testing for an excess of haplotype blocks with *p*-values below a given critical value, p_{crit}.
305 In principle, the expected number of such blocks under the null hypothesis of no true genetic
306 effects (i.e., the number of false positives, *x*) should be $E[x] = p_{crit} n_{total}$ where n_{total} is the
307 number of blocks tested. If the tests were independent, it would be straightforward to calculate the
308 full probability distribution of *x* under the null hypothesis and properly test whether the observed
309 number of 'significant' blocks $n_{p \leq pcrit}$ is significantly larger than expected under the null
310 hypothesis. There are two potential issues with this approach. First, it relies on the *p*-value from
311 each individual test accurately reflecting the chance of a false positive under the null hypothesis
312 (i.e., that all the assumptions of the mixed model are well met). Second, the tests are not
313 independent due to linkage disequilibrium. Permutations offer a conceptually simple (though
314 computationally intensive) solution to both issues. We permuted line identity (i.e., genome-wide
315 genotype) with respect to the set of fitness values (i.e., the 8 traits) and reanalyzed the data. From
316 each permutation, we recorded the *p*-value for each block (and for each trait) and then calculated
317 an *x* value (i.e., number of 'significant' blocks) for a given statistical criteria ($p_{crit} \in$
318 $\{0.001, 0.01, 0.05, 0.1\}$). To reduce computational time, we only tested every 7th block; the same
319 blocks were tested in the unpermuted and permuted data sets. (There is little additional information

320 gained from testing all blocks because of the high linkage disequilibrium among neighboring
321 blocks.) 2077 permutations were performed.

322 False discovery rates are an alternative approach to the multiple testing problem inherent
323 in genomic analyses but are not necessarily a perfect solution. The reliability of these methods
324 depend not only on the validity of the underlying *p*-values, but can depend on the independence of
325 tests as well as the shape of the distribution of *p*-values. The *R* package *qvalue* (Storey et al. 2021)
326 was used to convert *p*-values into *q*-values (i.e., FDR rate) based on a previously described method
327 (Storey and Tibshirani 2003). We calculated *q*-values with the unpermuted data as well as each
328 permuted data set. As we did with *p*-values, we compare the number of blocks with a *q*-value
329 below a given threshold, q_{crit}, to the distribution under the null hypothesis as determined from the
330 permutations.

331

332 *Gene Ontology (GO) Enrichment Analysis*

333 We conducted a gene ontology (GO) enrichment analysis of genes found in "genomic
334 clusters" that are associated with male and female fitness and the SA and SC axes. To help control
335 for strong linkage between loci, we employed a clustering algorithm to define genomic clusters
336 that are associated with each trait. For each chromosome arm, our clustering procedure began at
337 the block furthest away from the centromere. We used a sliding window of 3.5 Mb to identify
338 windows (or clusters) that had at least one haplotype block with $p < 0.1$. When a significant
339 haplotype block was identified, we examined the immediately proceeding 3.5 Mb window. If this
340 window contained any haplotype blocks that reached statistical significance, these blocks was
341 subsumed into the previous cluster. This procedure was repeated until no haplotype blocks in the

342 proceeding 3.5 Mb window reached statistical significance. Thus, a cluster was defined as a set of
343 statistically significant haplotype blocks that are closer than 3.5 Mb to each other. After obtaining
344 genomic clusters associated with each of the eight traits (see above), we used Ensembl's *Variant*
345 *Effect Predictor* (*VEP*; McLaren et al. 2010) to find genes that fall 5Kb upstream or downstream
346 of the leading haplotype block in each cluster (i.e., the haplotype block with the lowest *p*-value).
347 This analysis used a liberal $p < 0.1$ cut off for statistical significance of haplotype blocks but in
348 many cases the leading block for a cluster would have a much lower *p*-value. Next, we used
349 *GOrilla* (Eden et al. 2009) to test for an enrichment of GO terms in genes associated with each
350 trait relative to a background set of all genes annotated in the *Drosophila* genome from our *VEP*
351 analysis. *p*-values for GO terms were converted to *q*-values and a *q*-value threshold of 0.05 was
352 used to assess statistical significance.

353

354 Selection on Putatively Deleterious Variants in the DSPR lines

355 Our pipeline to identify putatively deleterious variants (i.e., single nucleotide
356 polymorphisms [SNPs] and insertion/deletion [indel] variants) involved three steps. First, we
357 imputed the underlying genomic sequence at sites segregating for SNPs and indels in the DSPR
358 for all 357 focal genotypes assayed in this study. Following quality control steps, we attempted to
359 enrich for sites likely to be deleterious by filtering out any variants segregating in either of two
360 geographically disparate outbred populations. Finally, we mapped the variants to functional
361 categories and retained only those variants falling within coding regions and whose functional
362 consequence could reasonably be expected to be deleterious.

363 *Genomic Imputation of Inbred Genotypes*

364 To impute SNPs segregating among inbred genotypes, we used the results of a HMM
365 implemented to uncover the parental ancestry of each genotype at regularly spaced 10kb intervals
366 along the genome of each inbred genotype and a table of SNPs segregating in each parental line
367 from the DSPR webpage (http://wfitch.bio.uci.edu/~dspr/). We filtered out all sites in the parental
368 SNP table where less than 95% of SNP calls were to either the reference or alternate allele in each
369 of the parental lines, sites with less than 15 reads and all multiallelic sites. Within each line, we
370 filtered out any 10kb intervals where the ancestry assignment probability of the most probable of
371 the eight parental lines was less than 95%. This filtering step removed an average of ~8.5% of
372 blocks from each RIL.

373 Imputation of insertion/deletion (indel) variants began by obtaining indel VCFs for each
374 of the eight DSPR parental lines from the *Drosophila* Genome Nexus (DGN) webpage
375 (https://www.johnpool.net/genomes.html). We filtered out indels not supported by at least 30
376 reads. We then imputed indels in the 357 DSPR lines in the same manner as described above for
377 SNPs.

378 *Identifying putatively rare variants from the DSPR*

379 Because deleterious variants are expected to segregate at low frequency at mutation-
380 selection balance, a reasonable approach to identifying putatively deleterious variants might
381 include focusing only on low frequency variants. To this end, we filtered out any variants
382 segregating in the DSPR that were also segregating in two geographically disparate outbred
383 populations of *D. melanogaster*: 197 genotypes from Zambia (*Drosophila* Population Genomic
384 Project Phase 3 [DPGP3]) and 205 genotypes from Raleigh, North Carolina (*Drosophila* Genetic
385 Reference Panel [DGRP]).

386 FASTA files containing whole genome sequences for all individuals sampled in both the
387 DPGP3 and DGRP populations were obtained from the DGN webpage. These genome sequences
388 were generated following a standardized alignment and quality control procedures described
389 elsewhere (Lack et al. 2015, 2016). Using Perl scripts supplied on the DGN webpage, we further
390 filtered sites that showed evidence of high relatedness and recent admixture. We concatenated all
391 genomes in a single population into a single multi-FASTA alignment file and generated a VCF
392 containing SNPs for each population using *SNP-sites* (Page et al. 2016). Next, we filtered out any
393 SNPs in the DSPR that were present in VCFs for either of the two outbred populations (i.e., sites
394 segregating for variants in the DSPR that were also captured as segregating in the two outbred
395 population samples were excluded from downstream analyses).

396 Next, we needed to determine which variant in the DSPR was the major allele in nature.
397 We chose one genome at random from the DPGP3 (ZI-103) and one genome at random from the
398 DGRP (RAL-100). For both genomes, we filtered out any sites that were polymorphic within the
399 populations that they were sampled from and then filtered out any sites that differed between these
400 two sequences including all sites that were masked in either sequence (these sites were also filtered
401 out from the DSPR). We compared this sequence (which we term the 'outbred consensus
402 sequence') back to the DSPR VCF file described above and filtered out any sites where one of the
403 two segregating alleles in the DSPR was not identical to the outbred consensus sequence. Finally,
404 for each site in the DSPR, we deemed the variant matching the outbred consensus sequence as the
405 major allele.

406 Indels were filtered from DSPR lines in a similar manner. We obtained indel VCFs for
407 both outbred populations from the DGN webpage and combined each individual VCF using the
408 '*vcf-merge*' command in *VCFtools* (Danecek et al. 2011). Finally, we filtered out any indels in the

409 DSPR lines that shared a starting coordinate with indels segregating in the either the DGRP or
410 DPGP3 indel VCFs.

411 *Functional annotation of genomic variants in the DSPR and identifying putatively rare deleterious*
412 *variants*

413 We predicted the functional consequences of minor allele SNPs and indels in the DSPR
414 lines using VEP (McLaren et al. 2010). Genomic coordinates for all VCFs were updated to release
415 6 of the Drosophila reference genome using the UCSC liftOver tool (Kent et al. 2002) prior to
416 passing VCFs to VEP. We retained only variants that fell within coding regions. For SNPs, we
417 focused on two classes of variants: those variants labelled by VEP as high 'IMPACT' and missense
418 mutations resulting in radical amino acid changes. Variants with high IMPACT scores include
419 variants predicted to have severe functional consequences including splice acceptor/donor
420 disrupting variants and stop loss/gain and start loss variants. Following Sohail et al. (2017), we
421 refer to this combined set of high IMPACT variants as 'loss-of-function' (LoF) mutations. For
422 each missense variant we characterized the resulting amino acid change as either a conservative
423 or radical change based on an amino acid classification scheme proposed by Zhang (2000) that
424 classifies amino acids into six categories that vary with respect to polarity and volume. An amino
425 acid change is deemed 'radical' if the major and minor alleles code for amino acids that are in
426 different categories with respect to either polarity or volume. Radical substitutions are under
427 stronger purifying selection in *Drosophila* than conservative substitutions (Smith 2003). Finally,
428 we use VEP to map indels to the reference genome and filtered out all indels that did not map to
429 coding regions.

430 In summary, we obtained a list of three classes of putatively rare variants segregating in
431 coding regions among the sample of DSPR lines assayed in our study: LoF mutations, variants

resulting in radical amino acid changes, and indels. This list of variants served as putatively deleterious variants segregating in the DSPR. Finally, we scored each inbred genotype for the number of variants of each class that they carry. We termed the number of putatively deleterious variants of carried by a genotype to be that genotype's "mutational burden". Note that the estimated number of deleterious variants for each RIL will, on average, be slightly below the true number because a small fraction of genomic blocks are excluded from each RIL. The "uncounted" variants are a small source of measurement error in the estimate of burden but are not expected to create biases in the effects of interest (i.e., direction and magnitude of variants or sex differences in these effects). Two genotypes were clear outliers with respect to the number of SNPs causing radical amino acid changes (Figure S1) and as result these two genotypes were removed from further analysis.

Statistical Analyses

The mutational burden for genotype *j* is represented by the numbers of LoFs (X_j), radical amino acid variants (Y_j), and coding region indels (Z_j). We modelled fitness as function of these three components of mutational burden. Our interest was in whether we could detect fitness effects of these mutational burdens and whether these effects may depend on sex or environment. Because of the large number of possible interaction terms, we began with a model selection approach to compare four models encompassing our major interests. In the "full" model (Model 1), we fit the following mixed model:

$$W_{ijk} \sim S_i + V_k + X_j + Y_j + Z_j + S_iX_j + S_iY_j + S_iZ_j + V_kX_j + V_kY_j + V_kZ_j + S_iV_kX_j + S_iV_kY_j + S_iV_kZ_j + \phi_j + e_{ijk}$$

453 Here, W_{ijk} is the mean relative fitness of the ith sex of the jth genotype assayed in mating
454 environment k. Relative fitness was standardized to the within-block mean and zero centered; for
455 each line, the average fitness across replicates was used for this analysis. *S* and *V* indicate sex and
456 mating environment, respectively. The model includes genotype as a random effect (ϕ_j) that is
457 assumed to follow a Gaussian distribution with a mean of zero and unknown variance. e_{ijk}
458 specifies the residual error of the ith sex of the jth genotype assayed in mating environment k. In
459 Model 2, we excluded the three three-way interactions (i.e., sex-by-environment-by-burden
460 component) but retained all six two-way interactions. In Model 3, we retained the interactions
461 between sex and each burden component but excluded all interactions involving the environment.
462 In Model 4, we retained the interactions between environment and each burden component but
463 excluded all interactions involving the sex.

464 Statistical analyses were conducted in *R* (version 3.6.2; R Development Core Team 2019).
465 Models were fit using the *lmer* function of the *lme4* package (Bates et al. 2015) after loading
466 *lmerTest* (Kuznetsova et al. 20) so that *t*-values were calculated using Satterwaite's method. We
467 confirmed the significance of the reported results by comparing each parameter estimate to a null
468 distribution created from analyses in which we permuted the set of fitness values (for each sex and
469 environment) with respect to genotype (RIL ID and three burden components).

470 **Results**

471 Quantitative Genetic Analysis

472 We recover modest additive genetic variation (σ^2_{RIL}) for fitness in both sexes when assayed
473 in simple (Figure 1; $\sigma^2_{RIL,f}$ = 0.020 [95% C.I.: 0.015, 0.025]; $\sigma^2_{RIL,m}$ = 0.027 [0.020, 0.035]) and
474 complex ($\sigma^2_{RIL,f}$ = 0.032 [0.023, 0.040]; $\sigma^2_{RIL,m}$ = 0.035 [0.025, 0.046]) mating environments. We

confirmed, by visual inspection of quantile-quantile plots of posteriors distributions, that the posterior distributions for all variance components estimates were approximately Gaussian and not right skewed. Estimates for the heritability h^2 in fitness are modest in males (simple environments: $h_m^2 = 0.102$ [0.074, 0.134]; complex environments: $h_m^2 = 0.124$ [0.089, 0.161]) and in females (simple environments: $h_f^2 = 0.150$ [0.111, 0.186]; complex environments: $h_f^2 = 0.164$ [0.121, 0.208]).

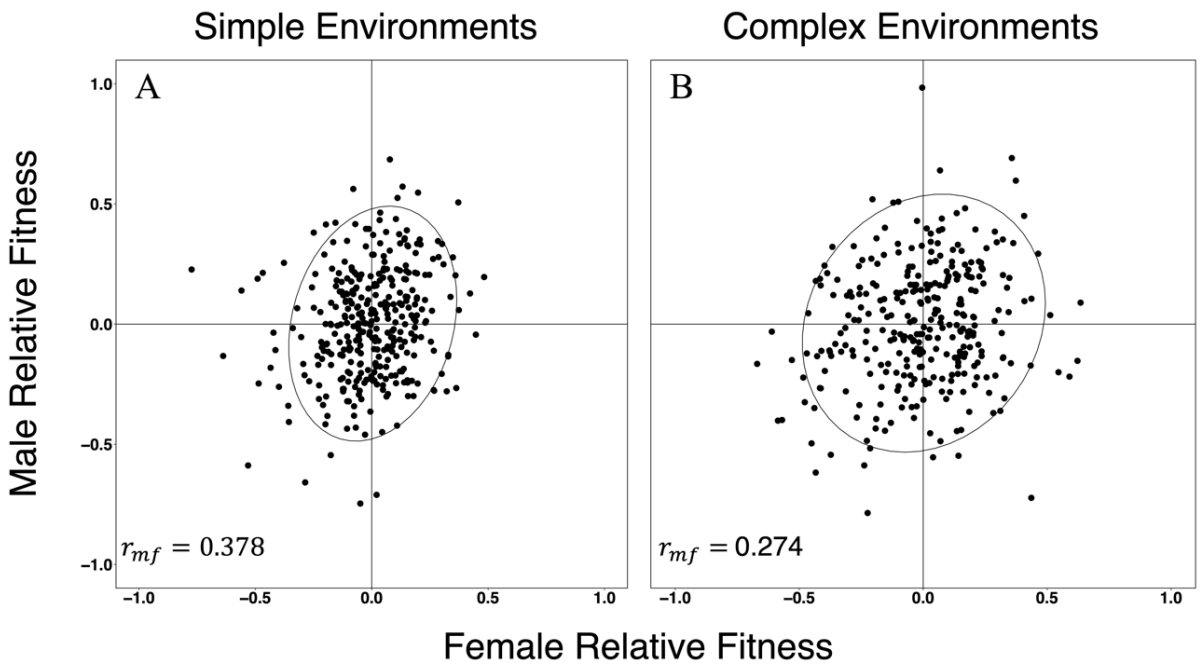

Figure 1: Male and female mean relative fitness of DSPR genotypes measured in **(A)** simple and **(B)** complex mating environments. Fitness measures are standardized to the within-block mean fitness for each sex and environment respectively and zero centered. Ellipses represent the 95% confidence ellipse for the distribution of sex-specific mean fitness estimates.

The intersexual genetic correlation for fitness was positive and significantly different from zero in both mating environments (Figure 1; simple: $r_{mf} = 0.378$ [0.158, 0.583]; complex: $r_{mf} = 0.274$ [0.086, 0.465]). Though significantly positive, these correlations are far from unity, suggesting a fair degree of genetic independence between the sexes. The cross-environment

correlation for fitness was also positive and significantly different from zero in both sexes (Figure 2; $r_{env,m}$ = 0.423 [0.246, 0.636]; $r_{env,f}$ = 0.322 [0.150, 0.471]).

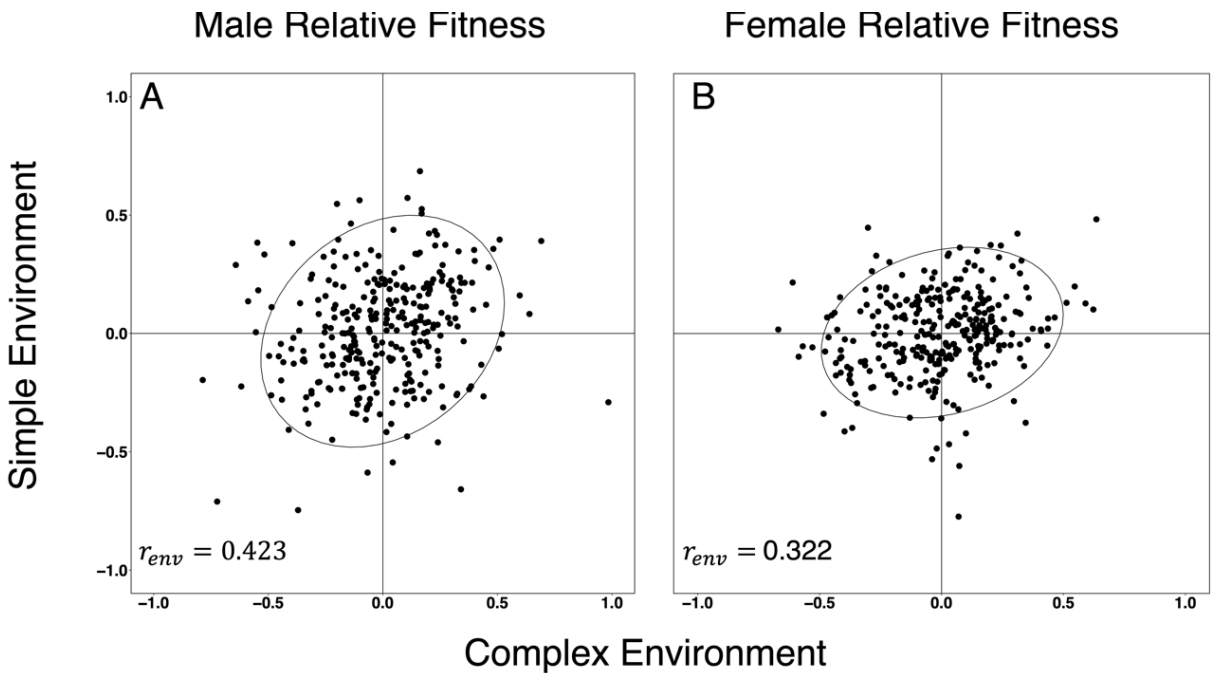

Figure 2: Mean relative fitness measured in simple and complex mating environments in **(A)** males and **(B)** Females. Fitness measures are standardized to the within-block mean fitness for each sex and environment respectively and zero centered. Ellipses represent the 95% confidence ellipse estimated from the sex-specific mean fitness estimates.

Association Mapping of Sexually Antagonistic and Sexually Concordant Loci

We present the results for sexual antagonism (SA) and sexual concordance (SC) here and results for male and female fitness are given in the Supplemental Materials. Manhattan plots for SA and SC in the simple and complex environments are shown in Figure 3. Few haplotype blocks have strong statistical associations with SA or SC; thus, we have little confidence in specific regions affecting SA or SC. Consequently, we shifted our focus to exploring whether the there is any evidence at all of regions associated with SA or SC.

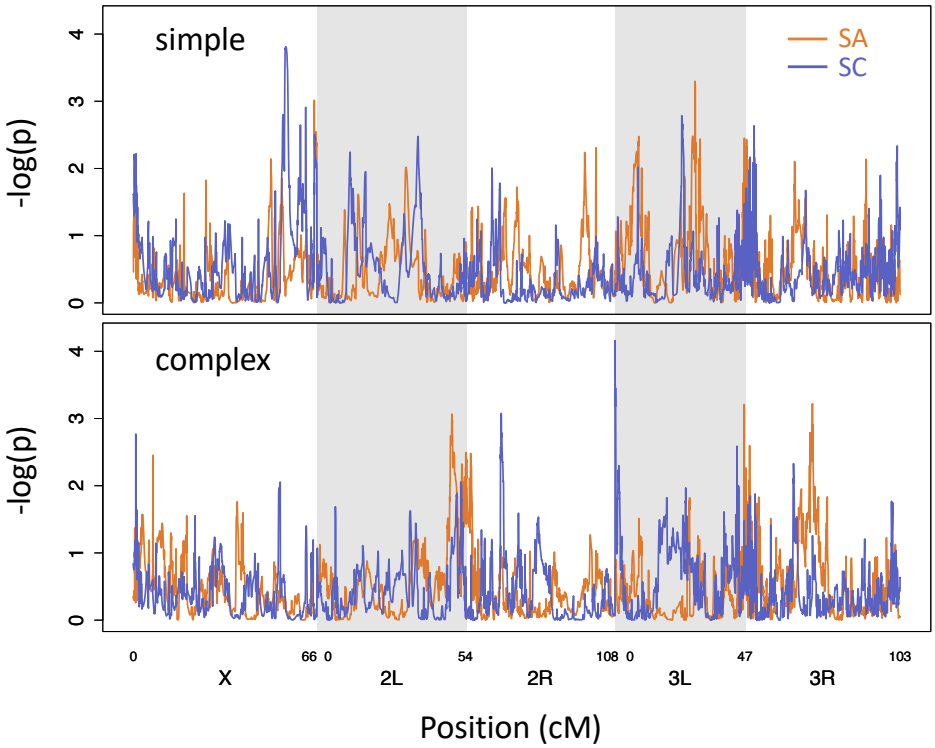

Figure 3: Manhattan plots. Strength of statistical support for associations between 10kb haplotype blocks along the genome with sexual antagonism (SA; orange) and sexual concordance (SC; purple) in DSPR lines when fitness was measured in the simple (top) or complex (bottom) mating environment.

515 Given that the classical quantitative genetic analysis indicates the existence genetic
516 variation, it is plausible that many blocks have modest effects that do not yield extreme p-values
517 (e.g., $p < 10^{-5}$). To explore this possibility, we counted the number haplotype blocks $n_{p \leq pcrit}$ that
518 passed a given statistical requirement (e.g., $p_{crit} = 0.001$, $p_{crit} = 0.05$) and compared this number
519 to the distribution under the null hypothesis of no genetic effects as generated by permutation.

520 Consider the results using $p_{crit} = 0.05$. We first ask if there is evidence of genetic
521 associations of any sort, considering the tests for all four traits (i.e., SA and SC in both
522 environments). We find 610 tests that pass the criterion of $p \leq 0.05$ ("significant blocks"). In
523 permutations, the average number of significant blocks is about half this number (353.3) and fewer
524 than 1% of permutations result in 610 or more blocks (i.e., there are more associations than
525 expected by chance). Examining SA and SC separately, there are 338 significant associations with
526 SA considering both environments. In permuted data sets, the average number of SA blocks is
527 176.2 and fewer than 2% of permutations have 338 or more. There are 272 significant SC blocks
528 across environments though ~10% of permutations have as many or more. Examining SA and SC
529 in each of the environments separately, there are more observed blocks than the corresponding
530 average number observed for each of the four traits, but only for SA in the simple environment is
531 the observed number in the upper 5% of the tail of its corresponding permutation distribution
532 (Table 1).

533 These results are sensitive to the statistical criterion (p_{crit}) used to identify significant blocks
534 (Table 1). In most cases, the observed number of blocks is greater than expected (i.e., the means
535 of the permutation distributions) but whether the excess is substantially greater than expected
536 depends on which trait is being considered and what statistical criterion is used. Whereas the
537 results based on $p_{crit} = 0.05$ indicate an excess of associations for SA but not SC, with $p_{crit} = 0.001$

538 or p_{crit} = 0.1, there appears to be an excess of associations for SC but not SA. With p_{crit} = 0.01,
539 neither type of association is in the upper 5% of the tail of the permutation distribution.

540 The use of permutations as a basis for a null distribution is important in evaluating the data
541 from the DSPR. As an example, consider the associations tests for SA in the simple environment
542 using p_{crit} = 0.05. 1682 blocks are tested. Under the null hypothesis of no true associations, we
543 expect 1682*0.05 = 93.1 'significant' blocks on average; the mean of the permutations is close to
544 this value at 88.1. If the tests were independent, then the variance among permutations should be
545 1682*0.05*0.95 = 88.5; however, the observed variance—1952—is much larger. This higher
546 variance is likely due to the non-independence of tests because of linkage disequilibrium, which
547 can result in the false positive rate being much higher (or much lower) than expected under
548 independence in any given instance (i.e., individual permutation or real data). The use of q-values
549 does not solve this issue (Table S9). For example, applying the significance criterion of $q \leq 0.3$,
550 only 73% of permutations result in 0 'significant' blocks and, of the remaining 27% of
551 permutations, the average number of 'significant' blocks is 129 (range: 1-1675).

552

553 Table 1. The number haplotype blocks passing a specified statistical criteria. For this analysis, only every 7th haplotype block was tested (1682
554 blocks tested). *n* is the observed number in the real (i.e., unpermuted data). $\bar{n}_{perm}$ is the mean number of associations in permutated data sets.
555 P_{perm} is the proportion of permutations with as many as or more than the observed number of associations (*n*); cases where the observed value
556 is the upper 5% of the distribution (i.e., $P_{perm} < 0.05$) are shown in bold. Results using a *q*-value criterion are shown in the Supplementary
557 Material, as are results for male and female fitness (Tables S2-4).

Criterion for 'significant' association	Metric	Any trait	SA	SC	SA in simple environment	SC in simple environment	SA in complex environment	SC in complex environment
	n	24	3	21	3	7	0	14
p_{crit} = 0.001	$\bar{n}_{perm}$	7.5	3.9	3.7	2.1	1.8	1.7	1.9
	P_{perm}	0.056	0.396	**0.025**	0.232	0.079	1	**0.025**
	n	117	54	63	34	34	20	29
p_{crit} = 0.01	$\bar{n}_{perm}$	73.2	36.2	36.9	18.3	18.7	18.0	18.3
	P_{perm}	0.117	0.184	0.131	0.148	0.152	0.333	0.201
	n	710	338	272	195	162	143	110
p_{crit} = 0.05	$\bar{n}_{perm}$	353	176	178	88.1	89.1	88.0	88.5
	P_{perm}	**0.0087**	**0.018**	0.095	**0.031**	0.084	0.127	0.274
	n	1064	479	585	261	298	218	287
p_{crit} = 0.1	$\bar{n}_{perm}$	699	348	351	173	174	174	176
	P_{perm}	**0.0077**	0.091	**0.016**	0.100	0.055	0.230	0.062

558

559
560 Following the procedure outlined in the Methods to generate lists of candidate regions, we
561 conducted GO enrichment analyses (Tables S5-S12). Under the statistical criterion used, no GO
562 terms were enriched among genes associated with the SA and SC in the simple environment.
563 However, we find evidence of enrichment of GO terms among genes associated with SA and with
564 SC in the complex environment. For SA, the enriched categories are biological processes that
565 affect basic cell function as well as higher levels of regulation (e.g., post-translational regulation,
566 epigenetic regulation; Table S7). For SC, a single molecular process was enriched: dipeptidyl-
567 peptidase activity (Table S8). Though the terms enriched are not readily interpretable, the existence
568 of enriched terms is another line of evidence that the association mapping results are more than
569 mere noise.

570

571 Mutational Burden

572 The assayed DSPR genotypes had on average 5.6 (range: 1-12) LOFs, 274.8 (39-363)
573 radical amino acid missense variants, and 125.7 (75-172) indels in coding regions. We first
574 compared four models that differed in their inclusion of interactions of burden with sex and/or
575 environment. Based on AIC, the best of these models included interactions of burden with sex but
576 burden not with environment ("Model 3"). The results of this model are shown in Table 2 (see
577 Tables S13-S16 for results of the other three models and their AICs). Figure 4 shows the model fit
578 relationship between mutation burden and fitness for the three classes of putatively deleterious
579 variants from Model 3. This model indicates significant selection against LoFs and indels but also
580 significant interactions of these two burden components with sex. In both cases, the interaction
581 term indicates that selection is significantly stronger against the burden component in females than

males. Because of the significant interaction terms involving sex, we subsequently fit models for each sex separately. These models were specified in the same way but all terms with 'sex' were omitted. In females there is a significant negative relationship between relative fitness and both LoF burden and indel burden (Table S16). Female fitness is reduced by ~ 1% per LoF and ~0.1% per indel. There is no significant association of radical amino acid burden in females. In males, we cannot detect significant associations between male fitness and any component of mutational burden (Table S17).

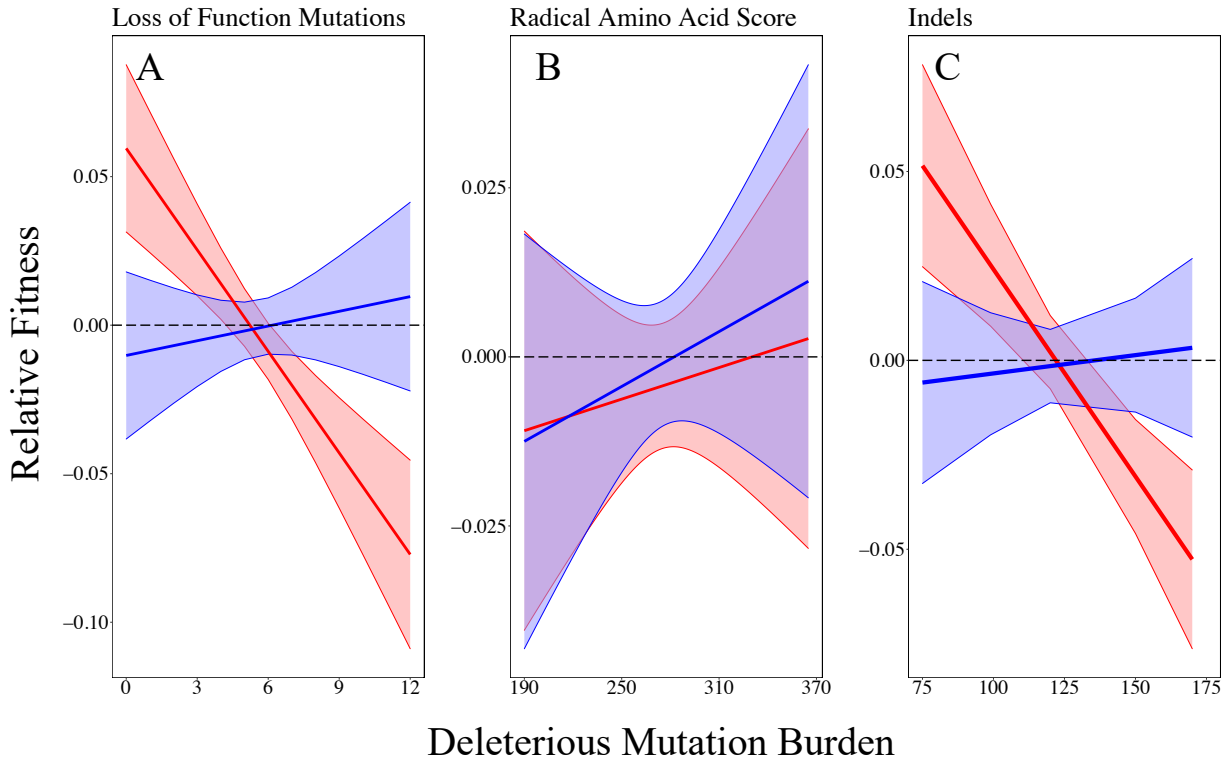

Figure 4: Model fits of the relationship between deleterious variant burden of **(A)** loss-of-function (LoF) mutations **(B)** radical amino acids **(C)** coding region indels and male (blue) or female (red) relative fitness (after zero-centering relative fitness). Shading represents the 95% confidence intervals for model fits.

596

597 **Table 2:** Parameter estimates from mixed model fit to evaluate the fitness effects of each of three
598 components of mutational burden—loss-of-function (LoF) burden, radical amino acid burden, and
599 indel burden—and their interactions with sex. Bold terms indicate statistically significant effects
600 at $\alpha = 0.05$.

Effect	Estimate	Standard Error	*P*
Intercept	0.17	0.11	0.10
Sex (Male)	-0.24	0.13	0.08
Environment (Simple)	0.005	0.011	0.69
LoF Burden	**-0.011**	**0.005**	**0.02**
Radical Amino Acid Burden	0.00008	0.00033	0.81
Indel Burden	**-0.0011**	**0.0005**	**0.03**
LoF Burden × Sex (Male)	**0.013**	**0.006**	**0.03**
Radical Amino Acid Burden × Sex (Male)	0.00006	0.00041	0.89
Indel Burden × Sex (Male)	**0.0012**	**0.0006**	**0.05**

601

602

603

604 **Discussion**

605 The sexes represent different selective environments: surveys of phenotypic selection have
606 found that shared traits experience persistent sex differences in both the strength (Singh and
607 Punzalan 2018) and direction (Cox and Calsbeek 2009; De Lisle et al. 2018) of selection. Thus,
608 there is considerable scope for segregating genetic to affect fitness of the two sexes differently. On
609 the other hand, the two sexes share many similar requirements so selection on many variants will
610 be similar. In this study, we used three approaches to explore different aspects of the sex-specific
611 genetic architecture of fitness in a 'synthetic' laboratory population of *Drosophila melanogaster*.

612 Quantitative Genetics Analysis

613 Our quantitative genetics analysis recovers significant levels of additive genetic variance
614 for fitness in both sexes in the two mating environments where we assayed fitness. The
615 heritabilities of fitness were modest in both sexes in the two mating environments. These results
616 are similar in magnitude as reported in other laboratory populations of *Drosophila* (e.g., Fowler et
617 al. 1997; Gardner et al. 2005; Pischedda and Chippindale 2006; Long et al. 2009; Ruzicka et al.
618 2019) and other insect species (e.g., Archer et al. 2013; Berger et al. 2014) but our estimates of the
619 heritability of fitness are much higher than those reported from studies of wild populations (e.g.,
620 Gustafsson 1986; Kruuk et al. 2000; Merilä and Sheldon 2000; McCleery et al. 2004; Teplitsky et
621 al. 2009; Mcfarlane et al. 2014). Low heritability estimates in wild populations are likely the result
622 of high levels σ_E^2 in fitness (Price and Schluter 1991; Houle 1992) rather than a lack of genetic
623 variation. In addition, the population used in our study has only experienced ~50 generations of
624 adaptation to laboratory conditions in which we measured fitness. Thus, there may be more fitness-

625 affecting variants segregating at intermediate frequency relative to an adapted population where
626 directional selection is expected to have reduced σ_A^2.

627 The intersexual genetic correlation for fitness r_{mf} gives us a genome-wide summary of the
628 extent to which the shared genome constrains the sexes from reaching their sex-specific trait
629 optima (Lande 1980) at least in the short term. Our estimates of r_{mf} were positive and significantly
630 different from 0 in both mating environments (Figure 1). This result implies that, on average,
631 genetic variants have sexually concordant fitness effects and, as a result, evolution is not strongly
632 constrained from improving the fitness of both sexes simultaneously, at least in the short-term.
633 Though estimates for r_{mf} in both environments are significantly greater than 0, they are also far
634 from 1. This intermediate value for r_{mf} suggests that while sexually concordant variants may be
635 predominant, other types of variants (e.g., sex-limited and/or sexually antagonistic variants) are
636 likely also segregating. This view is consistent with our genomic association mapping analysis
637 (though the latter should be regarded with caution because of limited power): the Manhattan plots
638 (Figure 3) suggest that sexually antagonistic genomic regions are present and are only modestly
639 less frequent than sexually concordant ones.

640 r_{mf} is not an immutable property of a species or a population. The evolutionary history of
641 a population in relation to ecological context in which fitness is measured can substantially alter
642 r_{mf}. For example, the genetic variation in populations where both sexes are approaching their sex-
643 specific fitness optima may become enriched for genetic variants with sexually antagonistic fitness
644 effects as selection will have eroded sexually concordant genetic variation (Long et al. 2012;
645 Connallon and Clark 2013), resulting in a negative r_{mf}. However, in a novel environment where
646 the fitness landscape is perturbed, the phenotypic mean values for both sexes may be above or

647 below the new fitness optima and thus segregating genetic variation may experience sexually
648 concordant selection resulting in a more positive r_{mf}. The 'synthetic population' used here was
649 created from mixing isogenic lines and allowing for 50 generations of recombination and,
650 presumably, lab adaptation (King et al. 2012a,b). Thus, it is plausible that our estimates of positive
651 r_{mf} reflect an overall concordance of selection on average as a consequence of this population
652 being relatively maladapted to the laboratory conditions in which we measured fitness, which was
653 not identical to how the lab population was originally created by King et al. (2012a,b).

654 Our significantly positive estimates of r_{mf} stands in contrast to a classic study using the
655 same species by Chippindale et al. (2001) who reported a significantly negative r_{mf}. That study
656 measured fitness in a lab context to which their fly population was well adapted. However, not all
657 studies neatly fit within the narrative in which the degree of adaptedness explains variation in r_{mf}.
658 Sharp and Agrawal (2018) estimated a positive though non-significant values for r_{mf} using
659 another *D. melanogaster* population that was well-adapted to the lab; assay conditions were similar
660 to but did not exactly replicate maintenance conditions. A particularly interesting case is that of
661 Collet et al. (2016) who measured r_{mf} in two descendant populations of the fly population studied
662 by Chippindale et al. (2001). Despite being maintained and assayed similarly, the two descendant
663 populations had significantly different values of r_{mf}, with r_{mf} significantly negative in one and
664 non-significantly positive in the other. That study demonstrates that r_{mf} can evolve substantially
665 over a few hundred generations in ways that are not completely captured by "adaptedness".

666 In addition to evolutionary changes in r_{mf} (i.e., changes in the underlying genetic
667 architecture), r_{mf} can be directly affected by the environment. Studies have found that r_{mf} may
668 be highly sensitive to the ecological context in which genetic variation is expressed (e.g., Berger

669 et al. 2014; Punzalan et al. 2014; Duffy et al. 2019). However, our estimates of r_{mf} in the two
670 mating environments are reasonably similar (and confidence intervals are broadly overlapping).
671 Given that sexual interactions are quite different in these two types of environments (Yun et al.
672 2017), it was not obvious from the outset that the two r_{mf} estimates would be so similar.

673 The similarity of r_{mf} estimates would be unsurprising if the genetic architecture of fitness
674 for each sex was the same across the two environments (i.e., if r_{env} was close to 1 for both sexes).
675 Indeed, r_{env} was significantly positive for both sexes. Nonetheless, both correlations were
676 substantially less than 1, indicating that the genetic architecture for fitness differs somewhat across
677 environments. This is not surprising given that these two environments likely select for different
678 phenotypic traits in both sexes. We have particularly strong reasons to believe this for males. In a
679 different experiment, populations that evolved for a few tens of generations in environments
680 similar to those examined here had diverged significantly for male fitness in a genotype-x-
681 environment dependent manner (Yun et al. 2019), presumably because these alternative mating
682 environments modulate the opportunity for male harassment to result in successful mating (Yun
683 et al. 2017). In contrast to males, females showed much weaker signs of divergence following
684 adaptation to these alternative environments. As such, it is perhaps surprising that here we have
685 found the cross-environment genetic correlation is of similar magnitude for both sexes.

686 It is notable that the intersexual genetic correlations within an environment (r_{mf}) and the
687 cross-environment genetic correlations for each sex (r_{env}) are all significantly positive and of
688 similar magnitude. As mentioned above, a simple explanation for all correlations being positive is
689 that this population contains an excess of variants that are unconditionally deleterious in the lab
690 environment, perhaps because of its history (i.e., insufficient time to adapt to general lab

conditions). Although particular suites of phenotypes may maximize fitness for each sex and mating environments, it is plausible that many of the same morphological, physiological and behavioural traits necessary to acquire fitness are similar between the sexes in both environments. Nonetheless, it is surprising that r_{mf} and r_{env} estimates are of similar magnitude; roughly speaking, this suggests that the differences between sexes are as large as the differences between environments with respect to segregating variation. For example, knowing that a genotype confers high fitness to males in one environment is approximately as good a predictor of conferring high fitness to females in that same environment as it is of conferring high fitness to males in the other environment.

Association Mapping Analysis

Our association mapping analysis revealed few, if any, genomic regions with very strong statistical associations with the traits examined. Consequently, we took a more exploratory approach to evaluating whether there was any evidence of genetic associations beyond what would be expected by chance, using permutations to create appropriate null distributions. The results of these analyses were somewhat inconclusive. On the one hand, in almost all cases the observed number of associations was larger than expected. However, the permutation-based null distributions have high variance and, in only some cases, was the observed number of associations unusually high relative to the corresponding null distribution. With some criteria, there was a substantial excess of associations with SA but not SC; with other criteria, there was an excess of associations with SC but not SA.

It is clear our power is limited. In general, the power to detect a significant genomic association is inversely related to the minor allele frequency (MAF) at loci underlying trait

713 variation. An advantage of this synthetic population is that the variants segregating in this panel
714 should have high MAFs (i.e., a neutral expectation of 12.5%) compared to natural populations.
715 However, in reality, the observed MAFs are often lower than 12.5%, presumably due to 50
716 generations of drift and selection. Of course, statistical power is also directly related to effect size.
717 Because of selection, the standing variation will be depauperate of variants with large directional
718 effects (e.g., unconditionally deleterious) segregating at high frequency. It is likely there are
719 numerous variants that make small contributions to fitness variation, but we would not expect to
720 detect them via this association mapping approach. Nonetheless, it seemed plausible we could
721 detect large-effect sexually antagonistic variants if such alleles were segregating. Theory indicates
722 that large-effect sexually antagonistic variants are more readily maintained by balancing selection
723 than small effect ones (Kidwell et al. 1977; Connallon and Clark 2012). Our failure to observe any
724 strong associations implies that the loci underlying fitness are unlikely to be of large effect.

725 Our tentative interpretation of the association study is that effect sizes are small to moderate
726 for both SA and SC as we tend to see more modest associations than expected yet fail to find very
727 strong associations. In addition, these data suggest that SA variation is not particularly rare relative
728 to (detectable) SC variation as we see only modestly stronger evidence for associations with SC
729 than SA (Table 2). Broadly speaking, these interpretations are consistent with results from the
730 classical quantitative genetic analyses that indicate genetic variance for both male and female
731 fitness and only a modest positive genetic correlation.

732 Using a permissive threshold to generate larger lists of candidate regions, we found some
733 evidence of enriched GO terms for SA and SC in the complex but not the simple environment. In
734 their genomic analysis of sexual antagonism, Ruzicka et al. (2019) found little evidence of
735 enriched GO categories and those they did where not obviously linked to sexually dimorphic traits.

736 One interpretation of these results is that sexual antagonism is diffuse affecting many types of
737 processes and/or is especially prone to persisting in genes affecting a variety of basic functions
738 that are under selective constraint. However, another interpretation is that studies such as ours lack
739 power to meaningfully identify genomic regions experiencing sexually antagonistic selection. On
740 the other hand, even though the terms identified are not readily biologically interpretable, the
741 existence of significantly enriched GO terms for both SA and SC in the complex environment
742 provide another form of evidence that the association analysis is not just noise.

743

744 Mutational Burden

745 We find evidence of selection against mutation burden in females but not males. All of the
746 statistically significant estimates of selection are, reassuringly, negative (i.e., selection disfavours
747 variants that we *a priori* expected to be deleterious). Examining entire classes of variants, rather
748 than individual variants, allows us to detect smaller effects. Specifically, we find a negative
749 association between female relative fitness and both LoF burden and indel burden, with each
750 additional LoF reducing female fitness by, on average, ~1%, and each additional indel by ~0.1%.
751 In contrast, we find no significant associations between any of the three classes of putatively
752 deleterious variants and relative fitness in males in either mating environment. These results imply
753 that selection acting on mutational burden is stronger in females than in males.

754 These results are surprising considering many studies have found evidence for stronger
755 selection acting on variants expressed in males. Several studies have estimated the sex-specific
756 selection acting on marker mutations in *Drosophila* (Whitlock and Bourguet 2000; Pischedda and
757 Chippindale 2005; Stewart et al. 2005; Sharp and Agrawal 2008; Hollis et al. 2009). These studies

758 have generally found evidence in support of the hypothesis that the strength of selection is male
759 biased but because visible marker mutations are unlikely to be representative of the typical
760 mutations that arise in a population, it is difficult to generalize these results. Several studies
761 examining the fitness consequences of novel mutations via mutation accumulation (MA)
762 experiments in *Drosophila* and other species (e.g., Enders and Nunney 2010; Mallet et al. 2011;
763 Mcguigan et al. 2011; Sharp and Agrawal 2013; Grieshop et al. 2016) have also generally
764 supported the hypothesis of stronger selection through males. While an advantage of using MA
765 lines is that they capture the fitness effects of more 'typical' mutations, one disadvantage is that
766 the causative mutations are typically completely unknown. Fitness results from MA studies may
767 be dominated by a small subset of the mutations that have large effects. This is problematic if such
768 mutations are unrepresentative in other ways. For example, perhaps mutations with male-limited
769 effects are disproportionately likely to be of large effect.

770 It remains unclear why our analyses suggest stronger selection in females than males
771 whereas past studies using phenotypic marker mutations or MA lines found the opposite. Our study
772 focuses on variants with specific molecular characteristics. Perhaps more importantly, the genetic
773 variation in the DSPR has been filtered by selection in the wild in contrast to MA studies that
774 capture the fitness effects of all novel mutations (excluding only extremely large effect variants
775 e.g., lethal, or sterilizing variants). Variants causing strong effects in the wild (in either sex) are
776 expected to be underrepresented relative to mutational input. One possibility for the discrepancy
777 of experimental results would be that fitness effects on males are reasonably consistent between
778 field and lab conditions whereas effects on females are more variable between environments. Thus,
779 variants with large fitness effects in males in the lab would be underrepresented whereas those
780 causing large female effects in the lab would be not be so underrepresented due to past selection

781 in the wild. Perhaps selection is more consistent on males between field and lab because their
782 fitness is so heavily determined by their interactions with other flies via mate competition whereas
783 female fitness may be more directly affected by the environment itself (e.g., temperature, humidity,
784 food availability).

785 A second possibility for the discrepancy between this and past study is from differences in
786 how fitness was assayed. First, viability is a potentially important source of fitness variation in our
787 study but minimized in many other studies. We measured fitness of adults via offspring produced
788 after a six day 'mating interaction' period whereas other fly studies typically use much shorter
789 mating periods. Our protracted period of mating interaction allows for more fitness variation via
790 adult mortality, especially so for females; males that die during this period can still contribute
791 offspring via matings prior to death but females dying prior to the day we allowed for egg laying
792 would prevent any contribution to offspring production in our study (i.e., fitness effect of mortality
793 during the mating period is stronger on females than males).

794 While the type of variants we studied are often presumed to be deleterious, very few studies
795 have attempted to validate this presumption by examining associations with fitness. Two recent
796 studies have fruitfully employed similar methods to investigate the aggregate fitness consequences
797 of putatively deleterious variants. Yang et al (2017) employed a comparatively conservative
798 criteria to filter for putatively deleterious variants in maize (*Zea mays*) by filtering SNPs based on
799 both their genomic evolutionary rate profiling (GERP) scores—a measure of evolutionary
800 constraint on a SNP (Cooper et al. 2005)—and the predicted functional consequence of variants.
801 Brown and Kelly (2019) took a more liberal approach to identify deleterious variants in yellow
802 monkeyflower (*Mimulus guttatus*) by filtering out any variant with a minor allele cut-off greater

803 than 5%. Both studies we able to identify significant associations between mutation burden and
804 measures of fitness.

805 Conclusion

806 Any study of the genetic variation in fitness should be considered with respect to the history of the
807 population in relation to the environment in which fitness was assayed. Our study likely represents
808 a modestly, but not exquisitely, well adapted population. Relative to an adapted population at
809 equilibrium in a constant environment, we would expect the genetic lines used here to harbour
810 more sexually concordant variation. Despite this expected "bias" towards concordance, both the
811 quantitative genetic and association mapping results suggest sex differences in the genetic
812 architecture of fitness. The observed intersexual genetic correlations are significantly greater than
813 zero, yet they are also substantially less than 1. Within the limits of our statistical power, there is
814 evidence for genomic regions that are sexually concordant but also regions that are sexually
815 antagonistic. While the history of the population is expected to affect the relative amounts of
816 concordant versus antagonistic variation, there is no *a priori* reason it should differentially affect
817 the sexes with respect to the fitness effects of putatively deleterious variants. Nonetheless, we find
818 evidence of stronger selection in females than males, contrary to studies of selection against
819 phenotypic markers and new mutations. While caution should be used in interpreting these results,
820 they add to the existing empirical evidence of sex differences in selective effects.

821

822

823 **Acknowledgements**: We are grateful to many undergraduate assistants who helped with data
824 collection, especially Nina Adler, Malak Bayoumi, Elenore Breslow, and Matthew Lindsay. We
825 thank Karl Grieshop, Julia Kreiner, George Sandler, and Santiago Sánchez-Ramírez for

826 statistics/bioinformatics help and advice. DSPR lines were generously provided by Stuart
827 MacDonald. Funding for this research was provided in the form of an Ontario Graduate
828 Scholarship (OGS) to AS and an NSERC Discovery Grant to AFA.

829

830

831

832

833

834 **Literature Cited**

835 Agrawal, A. F. 2001. Sexual selection and the maintenance of sexual reproduction. Nature
836 411:692–695.

837 Archer, C. R., S. K. Sakaluk, C. Selman, N. J. Royle, and J. Hunt. 2013. Oxidative stress and the
838 evolution of sex differences in life span and ageing in the decorated cricket, *Gryllodes*
839 *sigillatus*. Evolution. 67:620–634.

840 Arnqvist, G., and L. Rowe. 2005. Sexual Conflict. Princeton University Press, Princeton NJ.

841 Barton, N. H., and P. D. Keightley. 2002. Understanding quantitative genetic variation. Nat. Rev.
842 Genet. 3:11–21.

843 Bateman, A. J. 1948. Intra-sexual Selection in *Drosophila*. Heredity. 2:349–368.

844 Bates, D., M. Mächler, B. M. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models
845 using lme4. J. Stat. Softw. 67:1–48.

846 Berger, D., K. Grieshop, M. I. Lind, J. Goenaga, A. A. Maklakov, and G. Arnqvist. 2014.
847 Intralocus sexual conflict and environmental stress. Evolution. 68:2184–2196.

848 Bonduriansky, R., and S. Chenoweth. 2009. Intralocus sexual conflict. Trends Ecol. Evol. 24:280–
849 288.

850 Brommer, J. E., M. Kirkpatrick, A. Qvamström, and L. Gustafsson. 2007. The intersexual genetic
851 correlation for lifetime fitness in the wild and its implications for sexual selection. PLoS One
852 2: e744.

853 Brown, K. E., and J. K. Kelly. 2019. Severe inbreeding depression is predicted by the "rare allele
854 load" in *Mimulus guttatus*. Evolution. 74:587–596.

855 Chippindale, A. K., J. R. Gibson, and W. R. Rice. 2001. Negative genetic correlation for adult
856 fitness between sexes reveals ontogenetic conflict in *Drosophila*. Proc. Natl. Acad. Sci.

857 98:1671–1675.
858 Collet, J. M., S. Fuentes, J. Hesketh, M. S. Hill, P. Innocenti, E. H. Morrow, K. Fowler, and M.
859 Reuter. 2016. Rapid evolution of the intersexual genetic correlation for fitness in *Drosophila*
860 *melanogaster*. Evolution. 70:781–795.
861 Connallon, T., and A. G. Clark. 2012. A general population genetic framework for antagonistic
862 selection that accounts for demography and recurrent mutation. Genetics 190:1477–1489.
863 Connallon, T., and A. G. Clark. 2014. Evolutionary inevitability of sexual antagonism. Proc. Biol.
864 Sci. 281:20132123.
865 Connallon, T., and A. G. Clark. 2013. Evolutionary inevitability of sexual antagonism. Proc. R.
866 Soc. B Biol. Sci. 281.
867 Cooper, G. M., E. A. Stone, G. Asimenos, E. D. Green, S. Batzoglou, and A. Sidow. 2005.
868 Distribution and intensity of constraint in mammalian genomic sequence. Genome Res.
869 15:901–913.
870 Cox, R. M., and R. Calsbeek. 2009. Sexually Antagonistic Selection, Sexual Dimorphism, and the
871 Resolution of Intralocus Sexual Conflict. Am. Nat. 173:176–187.
872 Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G.
873 Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin. 2011. The variant call format
874 and VCFtools. Bioinformatics 27:2156–2158.
875 De Lisle, S., D. Goedert, A. Reedy, and E. Svensson. 2018. Climatic factors and species range
876 position predict sexually antagonistic selection across taxa. Philos. Trans. R. Soc. Lond. B.
877 Biol. Sci. 373:20170415.
878 Duffy, E., C. R. Archer, M. D. Sharma, M. Prus, R. A. Joag, J. Radwan, N. Wedell, and D. J.
879 Hosken. 2019. Wolbachia infection can bias estimates of intralocus sexual conflict. Ecol.
880 Evol. 9:328–338.
881 Enders, L. S., and L. Nunney. 2010. Sex- specific effects of inbreeding in wild- caught *Drosophila*
882 *melanogaster* under benign and stressful conditions. J. Evol. Biol. 23:2309–2323.
883 Fedorka, K. M., and T. A. Mousseau. 2004. Female mating bias results in conflicting sex-specific
884 offspring fitness. Nature 429:65–67.
885 Foerster, K., T. Coulson, B. C. Sheldon, J. M. Pemberton, T. H. Clutton-Brock, and L. E. B. Kruuk.
886 2007. Sexually antagonistic genetic variation for fitness in red deer. Nature 447:1107–1110.
887 Fowler, K., C. Semple, N. H. Barton, and L. Partridge. 1997. Genetic variation for total fitness in

Drosophila melanogaster. Proc. R. Soc. B 264:191–199.

Fox, J., and S. Weisberg. 2019. An R Companion to Applied Regression. Third. Sage, Thousand Oaks, CA.

Frankham, R. 2009. Genetic Architecture of Reproductive Fitness and its Consequences. Pp. 15–40 *in* J. van der Werf, H.-U. Graser, and R. Frankham, eds. Adaptation and fitness in animal populations: evolutionary and breeding perspectives on genetic resource management. Springer US, New York.

Fry, J. D. 2010. The genomic location of sexually antagonistic variation: Some cautionary comments. Evolution. 64:1510–1516.

Gardner, M. P., K. Fowler, N. H. Barton, and L. Partridge. 2005. Genetic variation for total fitness in *Drosophila melanogaster*: Complex yet replicable patterns. Genetics 169:1553–1571.

Grieshop, K., J. Stångberg, I. Martinossi-Allibert, G. Arnqvist, and D. Berger. 2016. Strong sexual selection in males against a mutation load that reduces offspring production in seed beetles. J. Evol. Biol. 29:1201–1210.

Gustafsson, L. 1986. Lifetime Reproductive Success and Heritability: Empirical Support for Fisher's Fundamental Theorm. Am. Nat. 128:761–764.

Hadfield, J. D. 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. J. Stat. Softw. 33:1–22.

Hollis, B., J. L. Fierst, and D. Houle. 2009. Sexual selection accelerates the elimination of a deleterious mutant in *Drosophila melanogaster*. Evolution. 63:324–333.

Houle, D. 1992. Comparing evolvability and variability. Genetics 130:195–204.

Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, and a. D. Haussler. 2002. The Human Genome Browser at UCSC. Genome Res. 12:996–1006.

Kidwell, J. F., M. T. Clegg, F. M. Stewart, and T. Prout. 1977. Regions of stable equilibria for models of differential selection in the two sexes under random mating. Genetics 85:171–183.

King, E. G., S. J. Macdonald, and A. D. Long. 2012a. Properties and power of the *Drosophila* synthetic population resource for the routine dissection of complex traits. Genetics 191:935–949.

King, E. G., C. M. Merkes, C. L. McNeil, S. R. Hoofer, S. Sen, K. W. Broman, A. D. Long, and S. J. Macdonald. 2012b. Genetic dissection of a model complex trait using the *Drosophila* Synthetic Population Resource. Genome Res. 22:1558–1566.

919 Kruuk, L. E. B., T. H. Clutton-Brock, J. Slate, J. M. Pemberton, S. Brotherstone, and F. E.
920 Guinness. 2000. Heritability of fitness in a wild mammal population. Proc. Natl. Acad. Sci.
921 U. S. A. 97:698–703.
922 Lack, J. B., C. M. Cardeno, M. W. Crepeau, W. Taylor, R. B. Corbett-Detig, K. A. Stevens, C. H.
923 Langley, and J. E. Pool. 2015. The *Drosophila* genome nexus: A population genomic resource
924 of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range
925 population. Genetics 199:1229–1241.
926 Lack, J. B., J. D. Lange, A. D. Tang, R. B. Corbett-Detig, and J. E. Pool. 2016. A thousand fly
927 genomes: An expanded *Drosophila* genome nexus. Mol. Biol. Evol. 33:3308–3313.
928 Lande, R. 1980. Sexual Dimorphism, Sexual Selection, and Adaptation in Polygenic Characters.
929 Evolution. 34:292–305.
930 Long, T. A. F., A. F. Agrawal, and L. Rowe. 2012. The effect of sexual selection on offspring
931 fitness depends on the nature of genetic variation. Curr. Biol. 22:204–208.
932 Long, T. A. F., P. M. Miller, A. D. Stewart, and W. R. Rice. 2009. Estimating the heritability of
933 female lifetime fecundity in a locally adapted *Drosophila melanogaster* population. J. Evol.
934 Biol. 22:637–643.
935 Lorch, P. D., S. Proulx, L. Rowe, and T. Day. 2003. Condition-dependent sexual selection can
936 accelerate adaptation. Evol. Ecol. Res. 5:867–881.
937 Mackay, T. F. C. 2001. The Genetic Architecture of Quantitative Traits. Annu. Rev. Genet.
938 35:303–339.
939 MacPherson, A., L. Yun, T. S. Barrera, A. F. Agrawal, and H. D. Rundle. 2018. The effects of
940 male harm vary with female quality and environmental complexity in *Drosophila*
941 *melanogaster*. Biol. Lett. 14:6–9.
942 Mallet, M. A., J. M. Bouchard, C. M. Kimber, and A. K. Chippindale. 2011. Experimental
943 mutation-accumulation on the X chromosome of *Drosophila melanogaster* reveals stronger
944 selection on males than females. BMC Evol. Biol. 11:156.
945 Manning, J. T. 1984. Males and the advantage of sex. J. Theor. Biol. 108:215–220.
946 McCleery, R. H., R. A. Pettifor, P. Armbruster, K. Meyer, B. C. Sheldon, and C. M. Perrins. 2004.
947 Components of variance underlying fitness in a natural population of the great tit *Parus*
948 *major*. Am. Nat. 164:E62–E72.
949 Mcfarlane, S. E., J. C. Gorrell, D. W. Coltman, M. M. Humphries, S. Boutin, and A. G. Mcadam.

950 2014. Very low levels of direct additive genetic variance in fitness and fitness components in
951 a red squirrel population. Ecol. Evol. 4:1729–1738.
952 Mcguigan, K., D. Petfield, and M. W. Blows. 2011. Reducing mutation load through sexual
953 selection on males. Evolution. 65:2816–2829.
954 McLaren, W., B. Pritchard, D. Rios, Y. Chen, P. Flicek, and F. Cunningham. 2010. Deriving the
955 consequences of genomic variants with the Ensembl API and SNP Effect Predictor.
956 Bioinformatics 26:2069–2070.
957 Merilä, J., and B. C. Sheldon. 2000. Lifetime reproductive success and heritability in nature. Am.
958 Nat. 155:307–310.
959 Page, A. J., B. Taylor, A. J. Delaney, J. Soares, T. Seemann, J. A. Keane, and S. R. Harris. 2016.
960 SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb.
961 genomics 2:e000056.
962 Parker, G. A. 1979. Sexual selection and sexual conflict. Pp. 123–166 *in* M. S. Blum and B. N. A.,
963 eds. Sexual Selection and Reproductive Competition in Insects. Academic Press, New York.
964 Pischedda, A., and A. Chippindale. 2005. Sex, mutation and fitness: Asymmetric costs and routes
965 to recovery through compensatory evolution. J. Evol. Biol. 18:1115–1122.
966 Pischedda, A., and A. K. Chippindale. 2006. Intralocus sexual conflict diminishes the benefits of
967 sexual selection. PLoS Biol. 4:2099–2103.
968 Poissant, J., A. J. Wilson, and D. W. Coltman. 2010. Sex-specific genetic variance and the
969 evolution of sexual dimorphism: A systematic review of cross-sex genetic correlations.
970 Evolution. 64:97–107.
971 Price, T., and D. Schluter. 1991. On the Low Heritability of Life-History Traits. Evolution.
972 45:853–861.
973 Punzalan, D., M. Delcourt, and H. D. Rundle. 2014. Comparing the intersex genetic correlation
974 for fitness across novel environments in the fruit fly, *Drosophila serrata*. Heredity. 112:143–
975 148. Nature Publishing Group.
976 R Development Core Team. 2019. R Development Core Team. R Foundation for Statistical
977 Computing, Vienna.
978 Ruzicka, F., M. S. Hill, T. M. Pennell, I. Flis, F. C. Ingleby, R. Mott, K. Fowler, E. H. Morrow,
979 and M. Reuter. 2019. Genome-wide sexually antagonistic variants reveal long-standing
980 constraints on sexual dimorphism in fruit flies. PLoS Biol. 17: e3000244.

Sharp, N. P., and A. F. Agrawal. 2018. An experimental test of the mutationselection balance model for the maintenance of genetic variance in fitness components. Proc. R. Soc. B Biol. Sci. 285.

Sharp, N. P., and A. F. Agrawal. 2013. Male-biased fitness effects of spontaneous mutations in *Drosophila melanogaster*. Evolution 67:1189–1195.

Sharp, N. P., and A. F. Agrawal. 2008. Mating density and the strength of sexual selection against deleterious alleles in *Drosophila melanogaster*. Evolution. 62:857–867.

Siller, S. 2001. Sexual selection and the maintenance of sex. Nature 411:689–692.

Singh, A., A. F. Agrawal, and H. D. Rundle. 2017. Environmental complexity and the purging of deleterious alleles. Evolution. 71:2714–2720.

Singh, A., and D. Punzalan. 2018. The strength of sex-specific selection in the wild. Evolution. 72:2818–2824.

Smith, N. G. C. 2003. Are Radical and Conservative Substitution Rates Useful Statistics in Molecular Evolution? J. Mol. Evol. 57:467–478.

Sohail, M., O. A. Vakhrusheva, J. H. Sul, S. L. Pulit, L. C. Francioli, L. H. Van Den Berg, J. H. Veldink, P. I. W. De Bakker, G. A. Bazykin, A. S. Kondrashov, and S. R. Sunyaev. 2017. Negative selection in humans and fruit flies involves synergistic epistasis. Science. 356:539–542.

Storey, J.D., A.J. Bass, D. Robinson (2021). *qvalue*: Q-value estimation for false discovery rate control. R package version 2.26.0. https://github.com/jdstorey/qvalue.

Storey, J.D., and Tibshirani R. 2003. Statistical significance for genomewide studies. Proc. Nat. Acad. Sci. U.S.A. 100:9440-9445.

Stewart, A. D., E. H. Morrow, and W. R. Rice. 2005. Assessing putative interlocus sexual conflict in *Drosophila melanogaster* using experimental evolution. Proc. R. Soc. B Biol. Sci. 272:2029–2035.

Teplitsky, C., J. A. Mills, J. W. Yarrall, and J. Merilä. 2009. Heritability of fitness components in a wild bird population. Evolution. 63:716–726.

Wade, M. J. 1979. Sexual Selection and Variance in Reproductive Success. Am. Nat. 114:742–747.

Wade, M. J., and S. J. Arnold. 1980. The intensity of sexual selection in relation to male sexual behaviour, female choice, and sperm precedence. Anim. Behav. 28:446–461.

1012 Whitlock, M. C., and A. F. Agrawal. 2009. Purging the genome with sexual selection: Reducing
1013 mutation load through selection on males. Evolution. 63:569–582.
1014 Whitlock, M. C., and D. Bourguet. 2000. Factors affecting the genetic load in *Drosophila*:
1015 Synergistic epistasis and correlations among fitness components. Evolution. 54:1654–1660.
1016 Yang, J., S. Mezmouk, A. Baumgarten, E. Buckler, K. Guill, M. McMullen, R. Mumm, and J.
1017 Ross-Ibarra. 2017. Incomplete Dominance of Deleterious Alleles Contributes Substantially
1018 to Trait Variation and Heterosis in Maize. PLoS Genet. 13:e086132.
1019 Yun, L., M. Bayoumi, S. Yang, P. J. Chen, H. D. Rundle, and A. F. Agrawal. 2019. Testing for
1020 local adaptation in adult male and female fitness among populations evolved under different
1021 mate competition regimes. Evolution. 73:1604–1616.
1022 Yun, L., P. J. Chen, A. Singh, A. F. Agrawal, and H. D. Rundle. 2017. The physical environment
1023 mediates male harm and its effect on selection in females. Proc. R. Soc. B Biol. Sci.
1024 284:20170424.
1025 Yun, L., P. J. Chen, K. E. Kwok, C. S. Angell, R. D. Rundle, and A. F. Agrawal. 2018. Competition
1026 for mates and the improvement of nonsexual fitness. Proc. Natl. Acad. Sci. U.S.A. 115: 6762-
1027 6767.
1028 Zhang, J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in
1029 mammalian nuclear genes. J. Mol. Evol. 50:56–68.
1030
1031
1032
1033
1034