*Evolution*, 2023, **77(12)**, 2642–2655
https://doi.org/10.1093/evolut/qpad163
Advance access publication 20 September 2023
**Original Article**

# Pleiotropic fitness effects across sexes and ages in the *Drosophila* genome and transcriptome

**Heidi W.S. Wong[1] and Luke Holman[2,]**

[1]School of Biosciences, University of Melbourne, Parkville, VIC, Australia
[2]School of Applied Sciences, Edinburgh Napier University, Edinburgh, United Kingdom
Corresponding author: School of Applied Sciences, Edinburgh Napier University, Sighthill Court, Edinburgh EH11 4BN, United Kingdom. Email: l.holman@napier.ac.uk

## Abstract

Selection varies between categories of individuals, with far-reaching ramifications: Sex-specific selection can impede or accelerate adaptation, and differences in selection between young and old individuals are ultimately responsible for senescence. Here, we measure early- and late-life fitness in adults of both sexes from the *Drosophila* genetic reference panel and perform quantitative genetic and transcriptomic analyses. Fitness was heritable, showed positive pleiotropy across sexes and age classes, and appeared to be influenced by very large numbers of loci with small effects plus a smaller number with moderate effects. Most loci affected male and female fitness in the same direction; relatively few candidate sexually antagonistic loci were found, though these were enriched on the X chromosome as predicted by theory. The expression level of many genes showed an opposite correlation with fitness in males and females, consistent with unresolved sexual conflict over transcription. The load of deleterious mutations correlated negatively with fitness across genotypes, and we found some evidence for the mutation accumulation (but not the antagonistic pleiotropy) theory of aging.

**Keywords:** distribution of fitness effects, evolution, mutation load, quantitative genetics, GWAS, TWAS

## Introduction

Genetic variation is the raw material of adaptation, yet the majority of new mutations are deleterious (Eyre-Walker & Keightley, 2007; Halligan & Keightley, 2009). Understanding how genetic variation is maintained, despite erosion by directional and stabilizing selection and the rarity of mutation, is a major goal of population and quantitative genetics. Heterogeneity in alleles' fitness effects across different environmental or genomic contexts is thought to be one important driver of variation. Such heterogeneity can create balancing selection, or at least slow the rate at which genetic polymorphism is lost by weakening net selection. In dioecious species, the fitness effect of an allele can differ between males and females (e.g., Barson et al., 2015; Harper et al., 2021; Ruzicka et al., 2019; Singh et al., 2023), which can increase genetic diversity (Connallon & Clark, 2012, 2014; Kidwell et al., 1977). Furthermore, sex-specific selection can have either positive or negative consequences for adaptation and the average fitness of the population (e.g., Agrawal, 2001; Bonduriansky & Chenoweth, 2009; Connallon et al., 2010; Lorch et al., 2003; Whitlock & Agrawal, 2009). For example, if total selection tends to be stronger on males than females due to male-biased sexual selection, yet similar alleles/phenotypes are associated with high fitness in both sexes, females benefit from a gene pool that has been purged of mutations by stronger selection on males (Grieshop et al., 2021; Whitlock & Agrawal, 2009). By contrast, for loci where the highest-fitness allele is not the same in both sexes, individuals inherit an excess of suboptimal alleles that were favored by selection on their opposite-sex ancestors, a phenomenon termed intralocus sexual conflict (Bonduriansky & Chenoweth, 2009). The net effect of selection on males on the fitness of females (and vice versa) depends on the relative portion of genetic variance in fitness that has antagonistic versus concordant effects on males and females, which remains poorly understood. Answering this question has ramifications for human health (Harper et al., 2021), local adaptation to climate change (Connallon & Hall, 2016), and the evolution of sexual reproduction (Agrawal, 2001). Furthermore, genetic variation influences fitness via a long causal chain of intermediates (e.g., transcription, translation, development, the phenotype, and selection), and a complete understanding of (sex-specific) adaptation can only be reached by elucidating these causal links.

Sex-specific adaptation shows many parallels with the evolution of aging (reviewed in Maklakov & Chapman, 2019), though these two topics are rarely studied together. Aging is evolutionarily puzzling because one might expect longer-lived individuals to have higher fitness, yet most organisms undergo senescence despite abundant genetic variation for senescence rate (e.g., Ivanov et al., 2015; Melzer et al., 2019). Evolutionary theories of aging rest on the existence of a "selection shadow" (Haldane, 1941), whereby the efficacy of natural selection starts to decline after the age at reproductive maturity because alleles expressed in later life are exposed to selection less often than alleles expressed in early life. In the "mutation accumulation" theory, mutations with deleterious late-life effects accumulate via genetic drift due

Received May 30, 2023; revisions received August 20, 2023; accepted September 19, 2023
Associate Editor: Robert Unckless; Handling Editor: Tracey Chapman

to this relaxed selection, creating senescence. In the non-mutually exclusive "antagonistic pleiotropy" theory, senescence arises because of selection for alleles that elevate early-life fitness yet reduce late-life fitness, due to the unequal efficacy of selection across age classes. The latter theory brings to mind intralocus sexual conflict, which similarly involves the creation of a genetic load via antagonistic pleiotropy, while the former is similar to the "weak form" of intralocus sexual conflict discussed by Bonduriansky and Chenoweth (2009) in which loci that are not expressed in one sex accumulate more deleterious alleles. For both aging and sex-specific adaptation, it can be instructive to think of males and females, and young and old individuals, as two contrasting environments that impose selection of varying strength (and possibly sign) on alleles that pass through them, and for maladaptation to arise because of differences in the selection surfaces of each environment. Measuring genetic correlations between these environments, and detecting the loci that shape these correlations, is a key goal of both research areas (Connallon & Hall, 2016; Connallon et al., 2010; Maklakov et al., 2015; Wilson et al., 2007).

Here, we investigate these ideas by searching for genotypic and transcriptional predictors of adult male and female fitness in younger and older individuals in the *Drosophila* Genetic Reference Panel (DGRP; Mackay et al., 2012). We focused on adult fitness since this life stage has the most scope for sex differences in selection (Chippindale et al., 2001), and because the selection shadow appears at the point of reproductive maturity (Maklakov & Chapman, 2019). The DGRP is a collection of almost entirely homozygous inbred lines, created by inbreeding the offspring of single wild-caught females from a site in North Carolina to create a persistent sample of the genetic variation segregating in that population. Whole genome sequencing and array-based expression data are available for each DGRP line, allowing one to test for associations between any line mean phenotype and the genotype or average whole-body transcriptome (from males or females) of that line. This study has multiple aims, including to measure the genetic (co)variance of sex- and age-specific fitness, identify fitness-associated loci and transcripts, determine the distribution of fitness effects across loci, and estimate the relative frequency of loci and transcripts that show concordant versus antagonistic associations with fitness across sexes and age classes. We test various predictions, for example, that selection on males purges the genome of deleterious mutations (Whitlock & Agrawal, 2009), that maladaptation arises due to genetic nonindependence of sexes and age classes (Bonduriansky & Chenoweth, 2009; Maklakov & Chapman, 2019), and that aging results from the accumulation of mutations that are especially harmful in older individuals (Maklakov & Chapman, 2019).

We predict that large numbers of genetic polymorphisms will have weak effects on fitness, that alleles with strong effects on fitness will be rare within populations (but each individual will nevertheless carry many of them), and that there will be extensive pleiotropy for fitness among sexes and age classes (evidenced by single alleles/transcripts being associated with fitness in multiple classes of individuals). It is unclear a priori what fraction of loci and transcripts will have concordant versus antagonistic associations with fitness across sexes and age classes, since this fraction depends on many simultaneously acting evolutionary forces (Berger et al., 2014; Collet et al., 2016; Connallon & Hall, 2016; Flatt, 2020; Holman & Jacomb, 2017; Innocenti & Morrow, 2010; Long et al., 2012; Rowe & Houle, 1997; Singh et al., 2023), so we measured this fraction using multiple complementary methods. Furthermore, we predict an increase in the genetic variance in fitness with age, in line with the mutation accumulation theory of aging, and possibly trade-offs between early- and late-life fitness in line with the antagonistic pleiotropy theory. We also predict an increase in the intersex genetic correlation for fitness with age because the hypothesized increase in the number of unconditionally deleterious mutations with increasing age should decrease the fraction of sexually antagonistic alleles in older individuals.

## Methods

### Fly stocks and husbandry

We studied 125 lines of the DGRP (Mackay et al., 2012). All lines had been recently analyzed using *Eco*RI RFLP genotyping (following Mackay et al., 2012) to verify that their genotypes matched expectations. All flies were reared at 25 °C in 25-mm vials containing c. 8 ml of lightly yeasted food medium (Supplementary Table S1). We used two additional stocks carrying the visible markers *brown*[1] and P{FRT(whs)} G13 P{Ubi-GFP.nls}2R1 P{Ubi-GFP.nls}2R2 as standardized mates and competitors for the DGRP flies (hereafter termed *bw* and *GFP*).

### Fitness assays

We measured fitness for adult males and females from each DGRP line using a protocol similar to Innocenti and Morrow (2010). We provide a concise description of the fitness assays here, and a complete version in the supplementary material.

Fitness for both sexes was defined as the quantity of offspring produced by groups of five focal DGRP flies, living in vials that also contained 10 same-sex and 15 opposite-sex individuals of a standard, phenotypically distinguishable genotype (*bw* females and *GFP* males). For females, fitness was defined as the absolute number of offspring produced, and for males as the proportion of offspring sired. We measured fitness twice for both sexes: one fitness measurement was performed on 2- to 5-day-old flies (referred to as "early-life" fitness) and the other on 14- to 17-day-old flies ("late-life" fitness). For context, this species becomes reproductively mature <2 days posteclosion, female fecundity peaks at 3–5 days, and average life span in the wild is thought to be less than a week (Flatt, 2020). DGRP flies (and their same-sex competitors) were not replaced when they died, such that our fitness assays incorporate variation in mortality as well as reproduction. For both the male and female fitness assays, we ran six replicate vials (each containing 5 DGRP flies) per line. All DGRP flies were reared at a standardized density of 100 first-instar larvae ("L1") per vial. To generate *bw* and *GFP* flies, we placed fifteen 1- to 4-day-old mated females into yeasted vials, allowed them to oviposit for 36 hr, then collected virgin offspring on days 10–13.

Importantly, we measured fitness by counting L1 larvae produced by the focal flies, rather than counting their adult offspring as in most comparable studies (e.g. Innocenti & Morrow, 2010; Ruzicka et al., 2019; Singh et al., 2023). This was accomplished by counting GFP larvae for male fitness, and by temporarily placing the DGRP flies into a separate oviposition vial for 24 hr for female fitness. Counting L1 larvae prevents inter-line variation in L1-to-adult survival

(which is considerable; Ellis et al., 2014; Rohde et al., 2016) from confounding estimation of inter-line variation in adult fitness traits such as adult survival, male mating success, and female fecundity. Genetic variation in survivorship from the zygote to the L1 stage remains a confounding factor, which might inflate our heritability and genetic correlation estimates; however, we expect this bias to be small, for example, because offspring carry only half the DGRP flies' alleles.

### Estimating heritability and genetic correlations for fitness traits

We calculated the proportion of variance explained by line for each fitness trait, using the R package rptR to implement the methods of Nakagawa and Schielzeth (2010). Specifically, we fit a univariate mixed model with Poisson or binomial errors (for female and male fitness, respectively), with line and block as random effects, then found the proportion of variance explained by the DGRP line (which approximates the broad-sense heritability; Mackay et al., 2012) and its 95% confidence intervals. We also calculated the Pearson correlations among the line means of the four fitness traits, yielding an estimate of the genetic correlations.

### Genome-wide associations with fitness

We downloaded DGRP genotypes from http://dgrp2.gnets.ncsu.edu/ and used PLINK v1.90 (Purcell et al., 2007) to remove variants with a minor allele frequency (MAF) below 0.05, and those with ≥10% of missing genotypes, then imputed missing genotypes using "Beagle" 5.4 (Browning et al., 2018). Next, we tested for associations between each variant and each of the fitness traits via four linear mixed models implemented in GEMMA (Zhou & Stephens, 2012), using the genomic relatedness matrix to adjust for population stratification. We defined the reference allele as the one that was most common across the entire DGRP, such that positive effect sizes mean that DGRP lines carrying the minor allele have higher average fitness. We also created an LD-pruned subset of SNPs via the PLINK command "--indep-pairwise 100 10 0.2." This LD-pruned subset makes possible computationally intensive analyses using mashr (see below), and ameliorates statistical nonindependence issues for some downstream analyses. We used the variant annotations generated by the creators of the DGRP, who used SnpEff (Cingolani et al., 2012).

### Transcriptome-wide associations with fitness

We performed a transcriptome-wide association study (TWAS), that is, testing for associations between the line mean expression level for each expressed transcript and each of the four fitness traits, using expression data by Huang et al. (2015). For each transcript-trait combination, we fit a linear model to calculate the effect size of transcript abundance on the fitness trait, as well as the associated standard error. For models involving male early- or late-life fitness, the predictor variable was the line mean expression level in male whole-body RNA extracts (scaled to mean 0, variance 1), while for models of female fitness, the predictor was the scaled line mean expression level in female whole bodies. We used limma (Ritchie et al., 2015) to calculate the average expression level (across individuals of both sexes) and the average sex difference in expression (expressed as a log ratio) for each transcript, for use in downstream analyses.

### Adjusting the GWAS and TWAS results using multivariate adaptive shrinkage

We used the R package mashr (Urbut et al., 2019) to adjust and analyze the summary statistics produced by GWAS and TWAS, similarly to Boyle et al. (2017) and Urbut et al. (2021). This package implements multivariate adaptive shrinkage using an empirical Bayes method, and requires as input a matrix of $n$ effect sizes for $m$ conditions, plus another $n \times m$ matrix containing the standard errors for these effect sizes. For the GWAS data, there were $n = 4$ effect sizes for each of the $m$ = 226,581 variants that remained after LD pruning, while for the TWAS data, there were $m$=14,286 transcripts. The main goal of mashr is to shrink the estimated effect sizes towards zero, thereby improving precision and controlling the number of false discoveries, while applying shrinkage in a manner sensitive to the covariances among the $n$ effect sizes (which are estimated from the data) as well as their standard errors. As well as returning an $n \times m$ matrix of adjusted effect sizes, mashr outputs the local false sign rate for each effect size. We used the local false sign rate (LFSR) to calculate the probability that each variant/transcript had a same-signed versus opposite-signed effect on female and male fitness, and did the same for early- and late-life fitness (see next section).

### Estimating the frequencies of antagonistic loci and transcripts

We define a locus or transcript as antagonistic if its true correlation with fitness has opposite signs in males and females, or in the early- and late-life assays within a sex. To answer the question "What fraction of the genome/transcriptome is antagonistic?," it is not enough to simply count the number of loci/transcripts with opposite effect size estimates, because half of all variants with a true effect size of zero (as well as approximately half of variants where the effect size was measured with high uncertainty) should have opposite-signed estimated effect sizes in males and females. Instead, one could count the number of loci/transcripts where the effect is significantly positive for one fitness trait and significantly negative for another, but this approach is either too conservative or too permissive depending on what arbitrary significance threshold is chosen. An alternative approach would be to avoid binary classification for each locus/transcript and instead focus on a quantitative measure of the evidence for antagonism, which avoids the need for an arbitrary threshold but makes interpretation more nuanced. To gain a complete picture of the proportion of candidate antagonistic loci and transcripts, we employed various complementary threshold-based and quantitative measures to examine the numbers of antagonistic and concordant loci and transcripts.

First, we simply tallied the number of loci and transcripts which had significantly positive or negative effects on pairs of fitness components (e.g., male and female early-life fitness, or male early- and late-life fitness), for various different significance thresholds from $p < .01$ to $p < 10^{-7}$. For small $p$-values, this approach is over-conservative because the false negative rate is high for any given locus/transcript, and there are two opportunities to make a false negative error. For larger $p$-values, this approach is under-conservative, leading to false positives.

Second, we binned the mashr-adjusted effect sizes of each locus/transcript on fitness component $i$ into quartiles. Because the median effect size is very close to zero (see Results), the four quartiles represent four equally-sized sets of loci/

transcripts with negative, weakly negative, weakly positive, and positive estimated effect sizes respectively. We can then tabulate the number of loci/transcripts that have (for example) a negative effect on fitness component *i* and a positive effect on component *j* (and vice versa) to gain insight into the relative abundance of antagonistic and nonantagonistic loci. Our aim was to create a simple graphical illustration of the proportion of loci and transcripts that are potentially antagonistic.

Third, we calculated an evidence ratio using the LFSRs produced by mashr, yielding a quantitative measure of the strength of evidence for concordance versus antagonism. The LFSR can be used to calculate the probability that a locus/transcript has a positive relationship with fitness, *P(pos)*, and the converse probability that the relationship is negative, *P(neg)* = 1 - *P(pos)*. We can calculate the probability that a locus/transcript has a concordant effect on two fitness components *i* and *j* as $P(concord) = P(pos)_i \times P(pos)_j + P(neg)_i \times P(neg)_j$, and the probability that it has an antagonistic effect as $P(antag) = P(pos)_i \times P(neg)_j + P(neg)_i \times P(pos)_j$. The evidence ratio is the ratio of these two probabilities, *P(concord)/P(antag)*. For loci where the true effect size is zero or loci where there is high uncertainty about the true sign of the effect size, the LFSR tends to 50% and the evidence ratio tends to 1. An evidence ratio of 10 indicates that the locus/transcript is 10-fold more likely to be concordant than antagonistic, while an evidence ratio of 0.1 indicates it is 10-fold more likely to be antagonistic than concordant.

As well as plotting the evidence ratios to illustrate the evidence for antagonistic and concordant effects, we used the *log2*-transformed evidence ratios in downstream statistical analyses investigating the characteristics of candidate antagonistic loci and transcripts. To improve model fit and balance statistical power with the risk of misclassification, we arbitrarily defined loci/transcripts as antagonistic if their evidence ratio was in the bottom 1% of the total sample, and nonantagonistic otherwise. We then ran a binomial generalized linear model with this classification as the response variable and evaluated the effects of various predictor variables using likelihood ratio tests.

## Results

### Heritability and genetic correlations between fitness traits

Figure 1 illustrates the variance and covariance in line means for the four fitness traits. There was evidence for significantly nonzero heritability for fitness in all sexes and age classes, and the two female fitness measures had significantly higher heritability than the two male fitness measures (Supplementary Table S2). For males, heritability was significantly higher for late-life compared to early life fitness, while for females, heritability did not differ significantly between age classes (Supplementary Table S2).

All pairs of fitness traits showed significantly positive genetic correlations, except for female early-life fitness and male late-life fitness, which showed a nonsignificant positive correlation (Supplementary Table S3). Genetic correlations between early- and late-life fitness within each sex were strongly positive but significantly less than unity, while genetic correlations between male and female fitness measurements were somewhat lower (Supplementary Table S3). The genetic correlation between male and female fitness was slightly more positive when calculated using the late-life ($r$ = .32) compared to early-life ($r$ = .23) fitness measurements, but this difference was not statistically significant (difference in $r$ = .09, bootstrapped 95% CIs: –0.06 to 0.23). Furthermore, the genetic correlation between early- and late-life fitness was significantly higher in males ($r$ = .86) compared to females ($r$ = .77) (difference in $r$: .098, bootstrapped 95% CIs: 0.008 to 0.22).

### Significant variants identified by GWAS

We identified 79 loci (with each "locus" defined either as one SNP/indel, or a group of nearby SNPs/indels in 100% linkage disequilibrium) that were significantly associated with at least one fitness measure with $p < 10^{-5}$ (Supplementary Table S4). These variants overlapped numerous coding and noncoding genes, including *5-hydroxytryptamine (serotonin) receptor 1A* (*5-HT1A*), *rhomboid-5*, and *roundabout 3*. Some of these variants had estimated fitness effects that were opposite in sign in males and females, though none of these putatively sexually antagonistic loci were significant in both sexes with $p < 10^{-5}$.

Supplementary Table S5 gives a breakdown of the numbers of significant loci using various significance thresholds and tallies the number of loci that significantly affected two or more fitness traits for each significance threshold (providing a very conservative test for pleiotropy). At the $p < .01$ level, 30 loci had a positive association with fitness in one sex and a negative association in the other sex, compared with 547 loci with sexually concordant effects significant at $p < .01$. At $p < .001$, there were no antagonistic loci and 10 concordant loci. There were no age antagonistic loci even at the $p < .01$ level, while many loci had concordant effects across age classes (Supplementary Table S5).

### The genetic architecture of fitness

Figures 2A and B plot the estimated effects—from GWAS with the effect sizes adjusted using multivariate adaptive shrinkage (see Methods)—of each of the 1,207,357 variants on the four fitness traits. The associated Manhattan plot (Figure 2C) shows the distribution of these effects across loci and chromosomes.

Consistent with the positive genetic correlations calculated from the phenotypic line means, we observed positive covariance between all four sets of variant effect sizes (Figures 2A and B). This result indicates that alleles that were positively (or negatively) associated with one fitness trait tended to have a similar association with the other three traits as well. The vast majority of loci had small adjusted effect sizes, though many had large adjusted effect sizes on one or more fitness traits.

The mean and median variant effect size was close to zero yet significantly negative for all four fitness traits, indicating that the minor allele was most often associated with lower fitness and the major allele with higher fitness (Supplementary Table S6). Furthermore, the largest effect sizes were observed for loci where the minor allele was the one associated with reduced fitness (as indicated by the asymmetry around $x = 0$ in Figures 2A and B).

Inspired by Boyle et al. (2017), we ordered all of the loci by the fitness effect of the minor allele, placed them in bins of 1,000, and then calculated the average effect size for male- and female early-life fitness for all the loci in each bin (using the unadjusted GWAS effect sizes). There was a very tight correlation between the average effects of the variants in each bin on

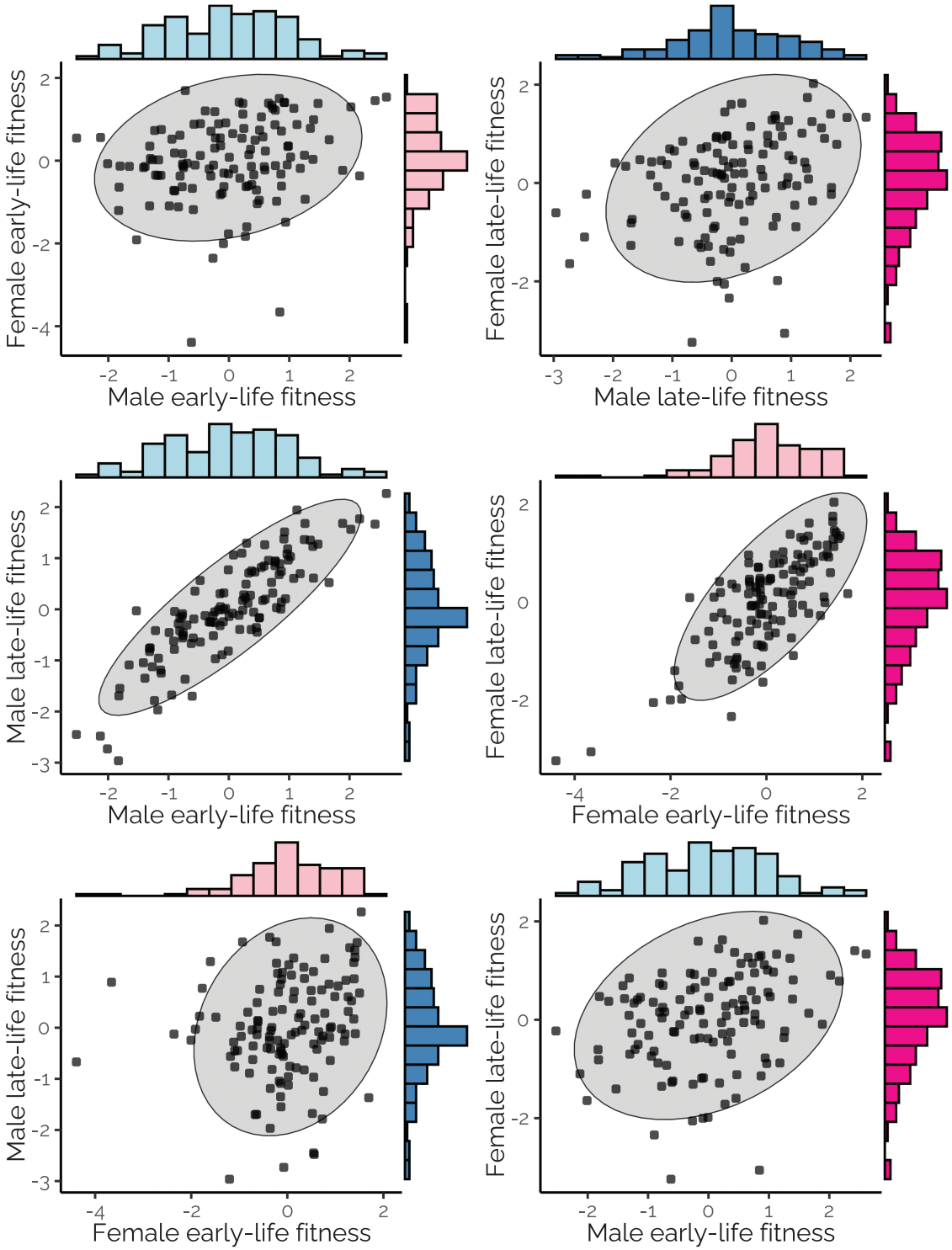


**Figure 1.** Correlations among estimated line means for fitness between sexes and age classes. The line means were estimated from mixed models that account for block effects and the nonindependence of our early- and late-life fitness measurements. Gray ellipses show where 95% of genotypes are expected to fall in bivariate trait space, and histograms show the variation in line means.

male and female fitness (Figure 3). Besides further illustrating the positive intersex genetic correlation for fitness, this finding implies that fitness is highly polygenic or "omnigenic" (Boyle et al., 2017). Our male and female fitness measurements were collected in independent assays, and so Figure 3 allows small but genuine associations between genotype and fitness to be distinguished from statistical noise, despite the low power of our study (and all GWASs) to detect loci with sufficiently small effect sizes. To see why, consider the alternative possibility that fitness is oligogenic, such that the heritability and genetic correlations we observed stem from a modest number of large-effect loci, and almost all the nonzero effect sizes in Figures 2A and B reflect statistical imprecision rather than true weak effects. The plot in Figure 3 would then be flat in the center (since the small nonzero effects on females would be spurious, and thus uncorrelated with the male effects) with

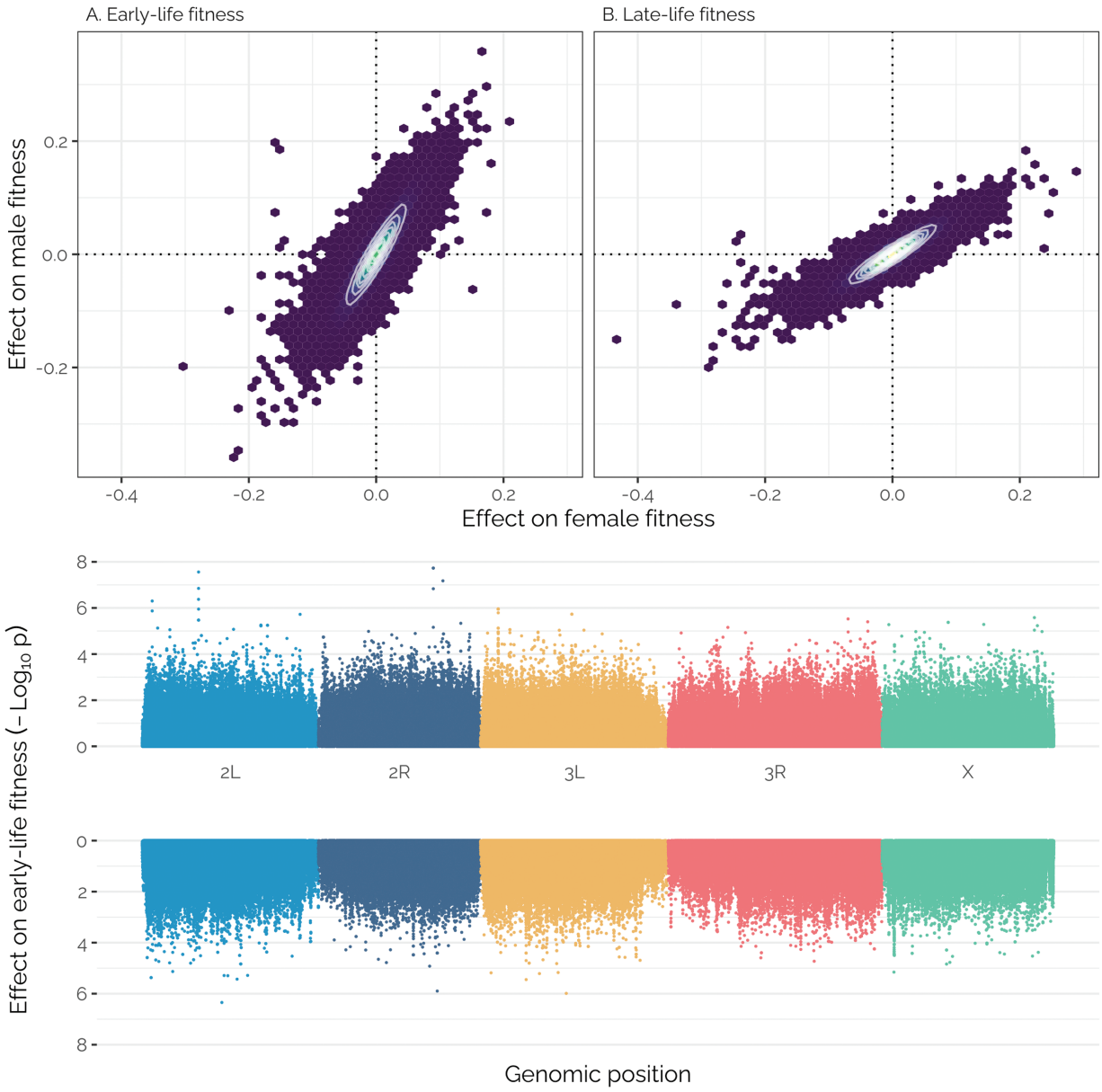


**Figure 2.** (A, B) The mashr-adjusted effect sizes of 1,207,357 polymorphic loci on male and female early- and late-life fitness. The data have been binned into hexagons, with the color and contour lines indicating the number of loci in each bin. Positive effect sizes indicate that the minor allele is associated with higher fitness and the major allele with lower fitness. (C) A pair of Manhattan plots, showing the chromosomal position and $-\text{Log}_{10}$ *p*-value (from linear mixed model GWAS using GEMMA) for each locus's effect on female (top) and male (bottom) early-life fitness.

steep inflections at each end (caused by the few large-effect loci). The fairly straight slope that we see instead suggests that a very large number of loci (all with MAF > 0.05) affect fitness—typically in both sexes in the same direction—with effect sizes ranging from tiny to moderate. We speculate that the nonlinearity visible in Figure 3A may be explained by a growing proportion of sexually antagonistic loci moving left from zero along the *x*-axis, due to the preferential removal by selection of female-harming alleles that do not have a countervailing beneficial effect on male fitness. However, contrary to this interpretation, Figure 3A does not have a corresponding plateau to the right of the *x*-axis, as predicted if female-beneficial alleles that are detrimental in males go to fixation less easily than female-beneficial alleles that do not harm males.

Inspired by Singh et al. (2023), we calculated the mutation load of each DGRP line as the total number of candidate deleterious mutations. We defined these candidates as alleles with 0 < MAF < 0.05 in the DGRP as a whole (implying selection against them), and which also had a relatively major mutational effect (defined as an insertion, deletion, or nonsynonymous substitution inside a coding sequence, as well as gains and losses of start codons) as opposed to a relatively minor effect (modifying an intron, synonymous site, or UTR/flanking/intergenic region). There was a negative correlation between all four fitness traits and mutation load across DGRP lines (Figure 4A–D), which differed significantly from zero for female late-life fitness and was borderline significant for the other traits (Figure 4E; Supplementary Table S7). There was a nonsignificant trend for mutations to more strongly correlate with late-life fitness compared with early-life fitness, in both sexes (Figure 4F; Supplementary Table S7). There was no detectable difference in how strongly mutation

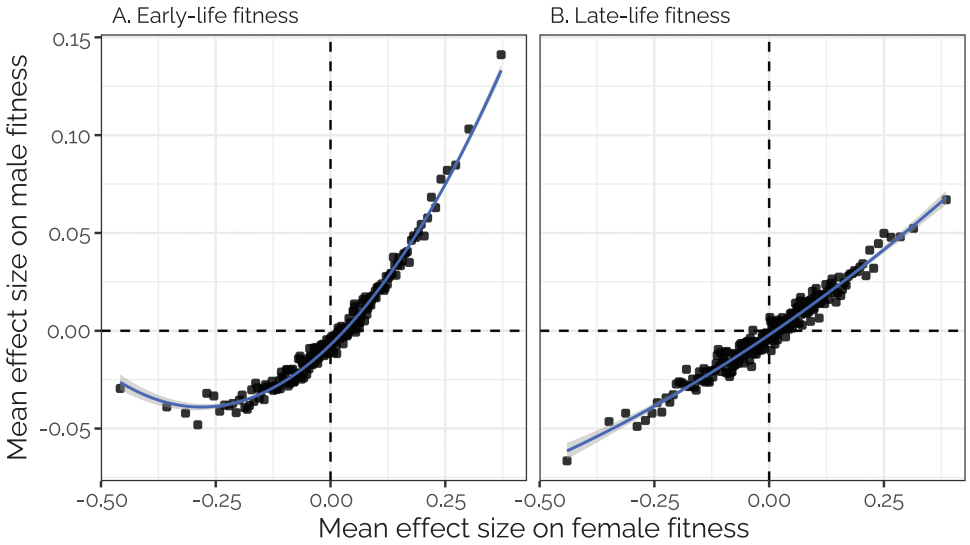


**Figure 3.** Estimated mean effect size for groups of 1,000 variants, on male and female early-life (A) and late-life (B) fitness. The variant groups were created by sorting variants by their estimated effect size on female fitness, then dividing the sorted list into groups of 1,000. This analysis was performed on a pruned set of 208,987 variants in approximate linkage disequilibrium with one another. The observed positive relationships imply that large numbers of loci have small effects on the fitness of both sexes (see main text). The fit lines are from a quadratic linear regression, and the shaded area shows the standard error.

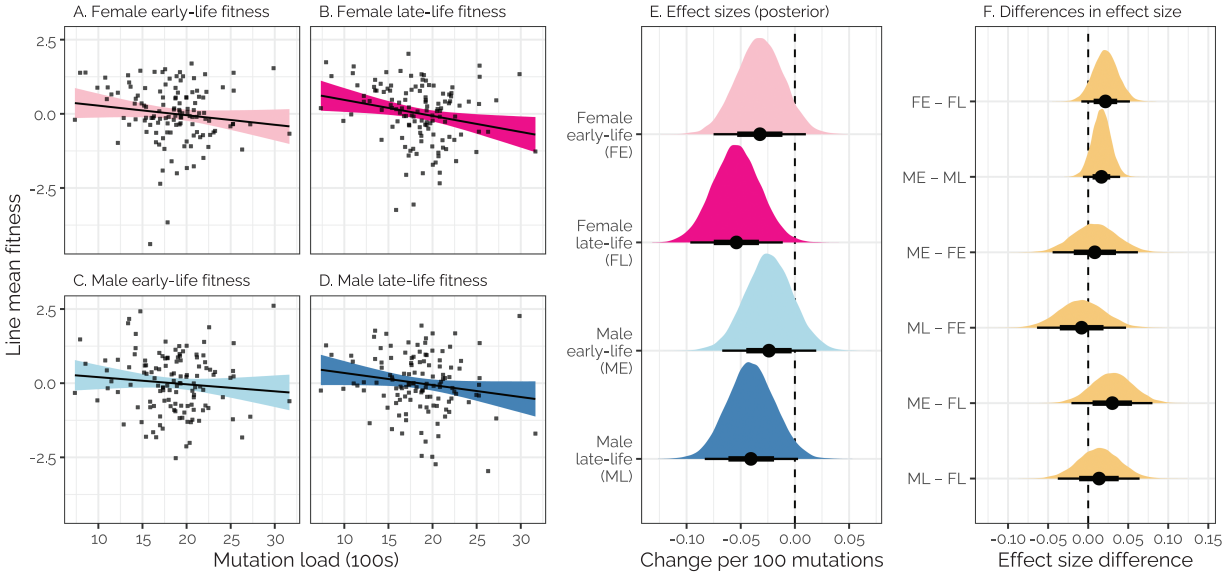


**Figure 4.** (A–D) The relationship across DGRP lines between mutation load and line mean fitness; the regression lines are from a Bayesian multivariate model that accounts for the covariance in the four fitness traits across DGRP lines. (E) The posterior estimates of the four regression slopes (i.e., the effect size of 100 mutations on fitness, where fitness is measured in standard units on the scale of the linear predictor), with the black bars summarizing the median and 66% and 95% credible intervals. (F) The posterior estimates of the differences in this effect size between pairs of fitness traits.

load correlated with male and female fitness (Figure 4F; Supplementary Table S7).

## Significant transcripts identified by TWAS

For many transcripts, mean expression level appeared to be correlated with at least one fitness measure across lines. Using a threshold of $p < .01$, 517 transcripts correlated significantly with at least one fitness trait (Supplementary Dataset S1); there were 6 transcripts if using $p < 10^{-4}$, and 1 transcript (from the gene *rudimentary*, involved in pyrimidine biosynthesis) if using $p < 10^{-5}$ (Supplementary Table S8). A number of transcripts showed opposite relationships with male and female fitness (e.g., 26 transcripts if using $p < .05$; Supplementary Table S8), such that expression level correlates positively with fitness in one sex and negatively for the other sex. There were no age-antagonistic transcripts at any threshold examined (Supplementary Table S8).

## Estimated frequencies of antagonistic variants and transcripts

Figure 5A–B and C–D (and Supplementary Tables S9 and S10) tabulate the relative numbers of loci and transcripts, respectively, falling into each of the four effect size quartiles for males and females. Quartile 1 contained the most

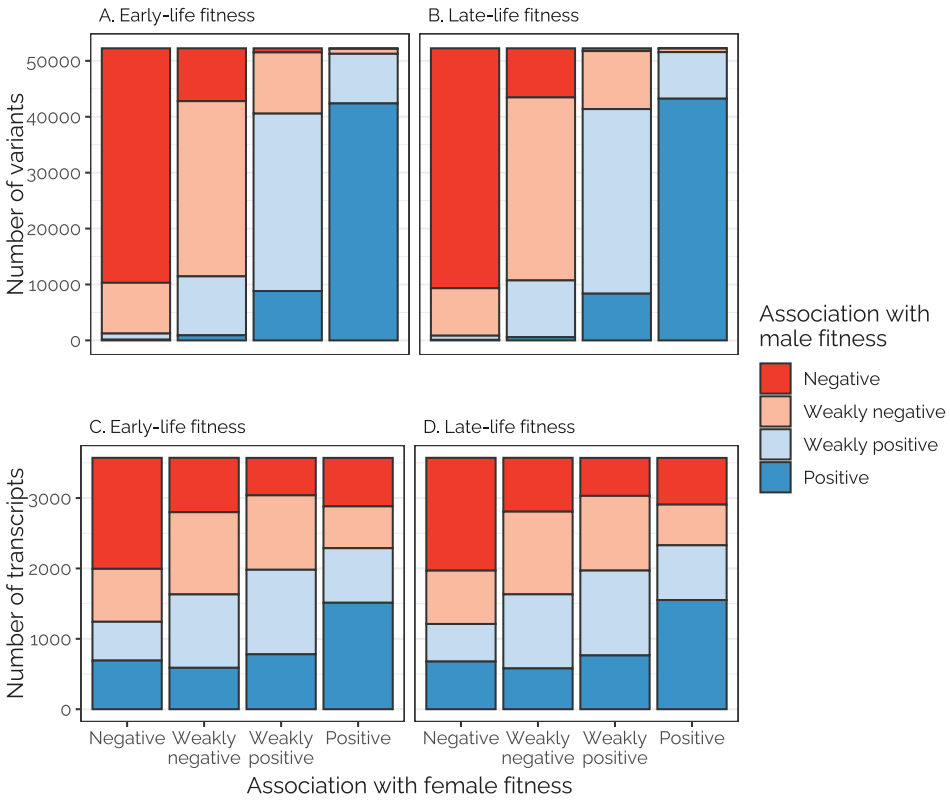


**Figure 5.** This plot tallies the number of loci (A, B) and transcripts (C, D) falling into each quartile in their effect on male and female fitness. Because the median effect size is essentially zero, quartiles 1–4 have been labeled negative, weakly negative, weakly positive, and positive, respectively. For example, the largest red area in (A) illustrates the number of loci whose effect size for female early-life fitness was in quartile 1 (i.e., the 25% most negative effects), which were also in quartile 1 for male early-life fitness.

"negative" associations (i.e., those where the minor allele, or lower gene expression, was associated with reduced fitness), while quartile 4 contained the most "positive" associations (i.e., those where the minor allele, or elevated gene expression, was associated with higher fitness).

Only a small proportion of loci fell into quartile 1 for their association with male fitness and quartile 4 for their association with female fitness (or vice versa), suggesting that sexually antagonistic loci are quite rare. Specifically, of the 208,987 loci examined (i.e., an LD-pruned set with MAF > 0.05), 129 (0.062%) had a "negative" association with female early-life fitness and a "positive" association with male early-life fitness, while a similar number (103, 0.049%) had a positive association with female early-life fitness and a negative association with male early-life fitness. Thus, by this measure, the frequency of candidate sexually antagonistic loci in this sample is 0.11%. When calculated using the associations with late-life fitness, the frequency of candidate sexually antagonistic loci was somewhat lower (0.067%). This 1.6-fold difference was statistically significant ($\chi^2 = 21.7$, df = 1, $p < .0001$), suggesting that a higher proportion of loci have sexually antagonistic effects on early-life fitness compared to late-life fitness.

By contrast, a much higher proportion of transcripts had opposite associations with fitness (Figures 4C and D). For early-life fitness, 1,380/14,286 transcripts (9.66%) were in quartile 1 in one sex and quartile 4 in the other, while for late-life fitness it was 1,341 transcripts (9.39%); there was no significant difference in the proportion of antagonistic loci between early- and late-life ($\chi^2 = 0.58$, df = 1, $p = .44$).

There was essentially no evidence for age antagonism when tabulating loci falling into opposite effect size quartiles. None of the 208,987 loci tested were in quartile 1 in their effect on early-life fitness and quartile 4 for their effect on late-life fitness, for either sex (Supplementary Table S11). Similarly, none of the 14,286 transcripts were in quartile 1 in their effect on early-life fitness and quartile 4 for their effect on late-life fitness (Supplementary Table S12).

Figure 6 shows the distribution of evidence ratios, illustrating the numbers of loci for which concordant or antagonistic effects were more likely, as well as the relative strength of evidence over the alternative scenario. Figure 6A indicates that there were many loci showing strong evidence for a concordant effect on fitness across sexes and age classes (e.g., a 50-fold higher probability relative to antagonism). A smaller number of loci showed some evidence for sexually antagonistic effects, though the largest evidence ratios for candidate antagonistic loci were comparatively small (e.g., a fourfold probability difference relative to concordance). Given that many sexually concordant loci were confidently detected, this result implies that loci with strong sexually antagonistic effects are probably rare. However, we cannot rule out the existence of some (perhaps many) sexually antagonistic loci with weak to moderate fitness effects (whose evidence ratios would be close to 1), since these cannot be confidently distinguished from weak-effect concordant loci without high statistical power. No loci showed evidence of age antagonistic effects.

Figure 6B reveals that for the transcriptome data, there was considerably more evidence for sexual antagonism: many

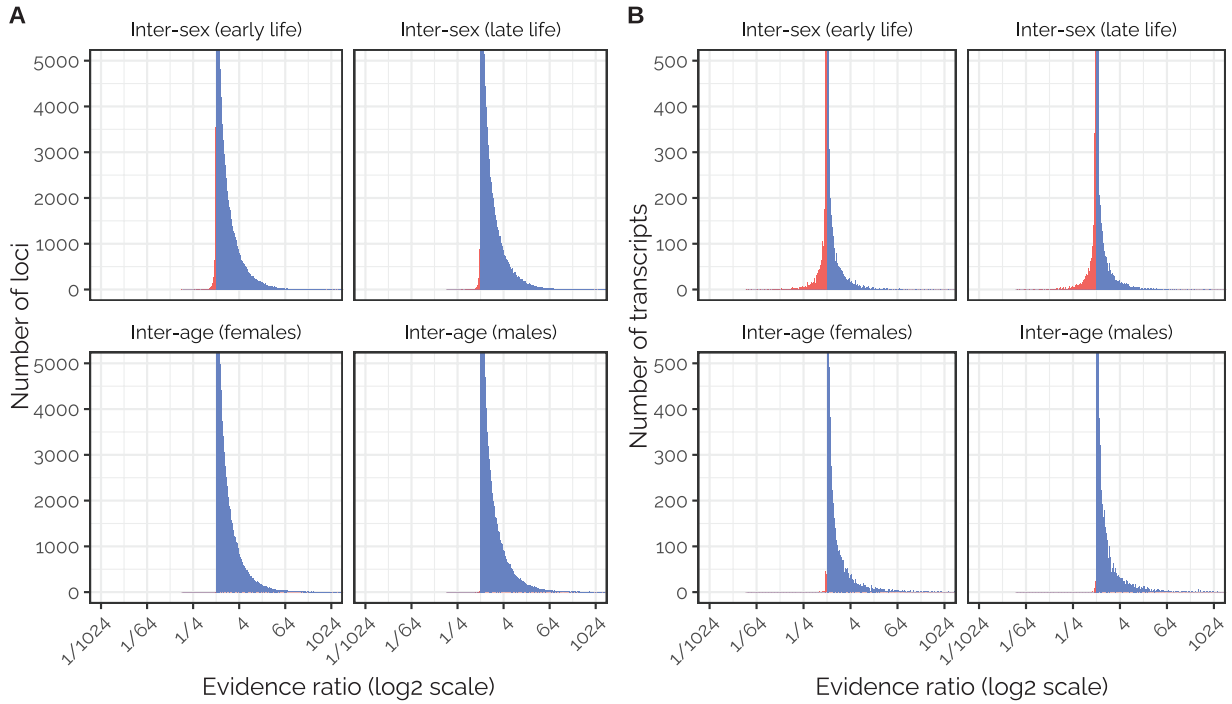


**Figure 6.** Distribution of evidence ratios across loci (A) and transcripts (B), illustrating the strength of evidence for a concordant relationship with fitness (evidence ratio > 1), or an antagonistic relationship (<1). The top row of both panels considers the evidence for concordance/antagonism between the sexes (within each age class), while the bottom row shows concordance/antagonism between age classes (within each sex). The figure illustrates that many loci and transcripts (i.e., those with large evidence ratios) show strong evidence for concordant effects across sexes and age classes. By contrast, there are relatively few candidate sexually antagonistic loci (i.e., those with evidence ratios well below 1), somewhat more sexually antagonistic transcripts, and essentially zero age antagonistic loci or transcripts.

transcripts had strong evidence ratios favoring antagonism (e.g., 31 transcripts had evidence ratios below 1/20). There were also many transcripts showing strong evidence in favor of a sexually concordant (or age concordant) relationship with fitness, and again essentially none showing evidence for age antagonistic effects.

## Characteristics of candidate antagonistic loci and transcripts

We used a binomial GLM to evaluate the effects of three predictor variables on the probability that a locus fell in the top 1% candidate sexually antagonistic loci, as determined by the evidence ratio for early-life fitness. The three predictors were "chromosome arm" (a 5-level factor), MAF (a continuous variable in [0.05, 0.5]), and "mutation type" (a 2-level factor describing whether or not the variant had a major effect, as defined in the section on mutation load). Mutation type was not significant ($\chi^2$ = 0.0, df = 1, $p$ = .97), but there was a significant difference between chromosomes (Figure 7A; $\chi^2$ = 11.28, df = 4, $p$ = .024), such that loci on the X chromosome were more likely to rank among the top 1% candidate sexually antagonistic loci (e.g., when compared with chromosome 2L; log odds ratio (LOR) = −0.26, $p$ = .0012, 2R: LOR = −0.14, $p$ = .073 or 3L: LOR = −0.13, $p$ = .080). There was also a strong positive effect of MAF, such that loci with a common minor allele were more likely to rank among the top 1% candidate sexually antagonistic loci than those with a rare minor allele ($\chi^2$ = 630, df = 1, $p$ < .0001). However, we stress that this correlation with MAF is expected even under the null hypothesis that selection has not shaped allele frequencies, for statistical reasons (*viz.*, loci with low MAF cannot have very high evidence ratios, since their fitness effects are measured less precisely). Nevertheless, MAF is included in the model to control for differences in average MAF when estimating the effects of chromosome and mutation type.

We used a similar binomial GLM to evaluate the effects of three predictor variables on the probability that a locus ranked among the top 1% candidate sexually antagonistic transcripts. The predictors were the chromosome from which the transcript was expressed (a 5-level factor), the mean expression level (averaged over sexes and DGRP lines), and the mean absolute expression bias between sexes (where zero indicates no sex bias and positive numbers indicate unequal expression between sexes). There was a negative effect of mean expression level (Figure 7B; LOR = −0.52, df = 1, $p$ = .0036); chromosome ($\chi^2$ = 3.57, df = 4, $p$ = .47) and sex bias in expression ($\chi^2$ = 0.04, df = 1, $p$ = .83) were not significant and were dropped from the full model.

Finally, we ran a GO:Biological Process enrichment test on the list of transcripts analyzed by TWAS (using the Wilcoxon method, implemented in the R package *GOfuncR*), with genes ordered by their log intersex evidence ratios (at the early-life stage; i.e., the variable in the top left of Figure 5B). This test searches for GO terms that are enriched among the genes which produced the transcripts with strongly negative log evidence ratios (i.e., the top candidate antagonistic transcripts) and also strongly positive evidence ratios (i.e., the top candidate concordant transcripts). Using a significance threshold of FWER < 0.05 (family-wise error rate, computed by permuting the GO graph to adjust for multiple testing), we found no GO terms enriched among the candidate antagonistic transcripts. However, a number of GO terms were significant among the candidate concordant transcripts (Table S13), including "gene expression," "cytoplasmic translation,"

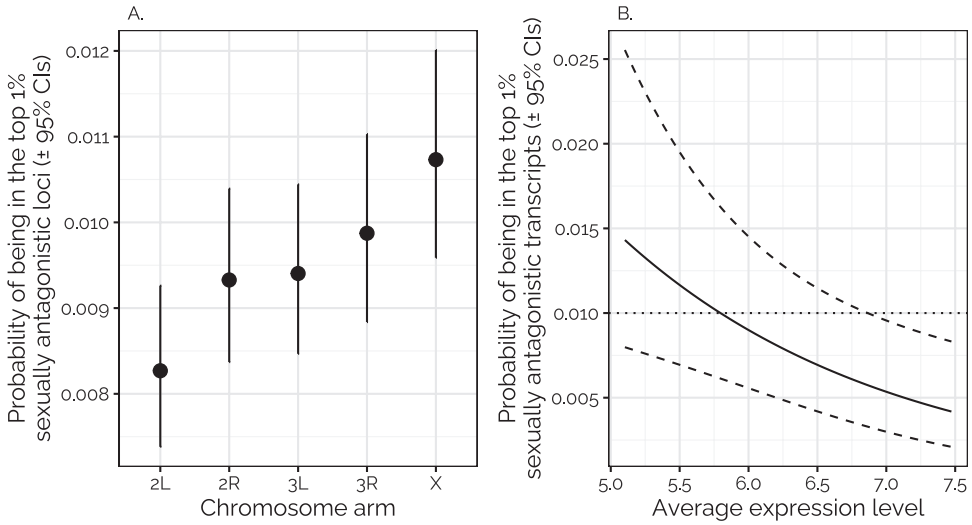


**Figure 7.** (A) The predicted probability that a locus is among the top 1% candidate sexually antagonistic loci, ranked by evidence ratio, for each of the major chromosome arms. (B) The predicted relationship between average log2 gene expression level and the probability that a transcript is among the top 1% candidate sexually antagonistic transcript. The predictions are from binomial GLMs, and the error bars and dashed lines show 95% confidence intervals estimated as 1.96 *SE*.

and "peptide biosynthetic process." This finding suggests that variation in the expression levels of genes that are involved in basic, essential cellular functions correlates with variation in fitness across genotypes, concordantly in both sexes.

## Discussion

Before discussing the results, it is prudent to highlight some limitations of our study. The genetics of fitness depends on the evolutionary history of the population and on how fitness is measured. The DGRP was created by 20 generations of full-sib inbreeding from the offspring of single wild-caught females, which is expected to purge some mutations that have strong fitness effects when homozygous. Furthermore, at least for alleles with strong fitness effects, selection might purge sexually concordant alleles more readily than antagonistic ones (Connallon & Clark, 2012), and the small, high-relatedness populations used to create inbred lines might change the relative efficacy of selection on male- and female-beneficial antagonistic alleles (cf., Li Richter & Hollis, 2021; Pizzari et al., 2015). Using inbred lines might therefore cause underestimation of the genetic variance in fitness and mismeasurement of the prevalence of sexual antagonism, the distribution of fitness effects, and the relative sizes of male and female genetic variance in fitness. Furthermore, it is unknown how closely the fitness traits we measured correlate with fitness in the wild, though as discussed below there are multiple lines of evidence that our fitness measures are sufficiently well-correlated with fitness in the wild to provide useful insights. Additionally, since DGRP lines are homozygous, we cannot separate additive from nonadditive genetic effects, and the allelic fitness effects we report might be different in heterozygous flies (probably lower, since harmful alleles tend to be at least partly recessive; Agrawal & Whitlock, 2011; Ruzicka et al., 2021).

Overall, there was extensive evidence for genetic effects on fitness and positive pleiotropy across fitness components. All four fitness traits showed significant broad-sense heritability, with greater heritability for female fitness than male fitness (perhaps implying stronger past selection on males; Whitlock & Agrawal, 2009), and greater heritability for male late-life fitness than male early-life fitness. Furthermore, there were positive genetic correlations between male and female fitness, and between early- and late-life fitness. Additionally, many loci had nonzero estimated effect sizes for one or more fitness traits in genome-wide association tests, with 79 significant loci significant at $p < 10^{-5}$. Most of these loci lay within or near genes, including *5-HT1A* (a serotonin receptor), *rhomboid-5* (epidermal growth factor receptor signaling), and *roundabout 3* (nervous system and gonad development). There was positive covariance in GWAS effect sizes between fitness components, as predicted if alleles commonly have positive pleiotropic effects on male and female fitness, and on early- and late-life fitness.

At least three of our findings suggest that loci with sexually antagonistic effects are rare, and age-antagonistic loci are rarer still, at least among loci with MAFs greater than 0.05. First, the observed positive genetic correlations between male and female fitness, and between early- and late-life fitness, imply that sex- and age-concordant genetic variance in fitness exceeds antagonistic fitness variance. The sign of a genetic correlation is determined by the relative abundance of positively and negatively pleiotropic loci, as well as the MAFs and effect sizes of those loci (e.g., Roff, 1996). Antagonistic genetic polymorphisms are theoretically expected to have higher MAFs and sex-specific effect sizes than concordant ones (all else equal), because concordant polymorphisms experience stronger net directional selection which depletes genetic variation. Their predicted inflated MAFs and effect sizes mean that sexually antagonistic loci should have a disproportionately large influence on the genetic (co)variance in fitness. Therefore, our finding that concordant genetic variance in fitness nevertheless exceeds antagonistic variance implies that concordant loci substantially outnumber antagonistic loci. This conclusion is perhaps unsurprising, since most new non-neutral mutations probably have unconditionally deleterious (i.e., sexually concordant) effects (Eyre-Walker & Keightley, 2007); however, some studies have nevertheless found evidence of pervasive

sexual antagonism (e.g., Chippindale et al., 2001). Second, using multivariate shrinkage models and a quartile binning method, we estimated the percentage of sexually antagonistic loci to be on the order of 0%–1%, and age-antagonistic loci on the order of 0%. Third, when we computed evidence ratios for each locus from a multivariate model of GWAS effect sizes, very few loci were classified as being more likely to be sexually antagonistic (compared with sexually concordant), and among the few loci that were, none were classified as being sexually antagonistic with high confidence.

We also found evidence that selection, as approximated by our lab-based fitness assays, has shaped allele frequencies in the wild population from which the DGRP was created (as estimated from the allele frequencies in the DGRP). First, alleles that were associated with lower fitness in our GWAS tended to be the minor allele in the DGRP, as expected if the selection surface created by our assays is similar to that imposed by conditions in the wild. Similarly, we found a negative correlation between a DGRP line's load of putatively deleterious mutations (partly defined as those that are rare in the wild) and its fitness in our lab-based assays. Second, the largest absolute effect sizes that we observed tended to be negative, indicating that minor alleles with strong, negative fitness associations outnumber minor alleles with strong, positive fitness associations. Third, we did not find any loci showing very clear associations with fitness, as quantified by the $p$-value; the lowest GWAS $p$-values were $p < 10^{-7}$. Though this partly reflects limited statistical power, some similarly powered DGRP studies have detected very strong associations (e.g., $p < 10^{-20}$), illustrating that the DGRP does contain genetic polymorphisms with major effects on other phenotypic traits, which can be confidently detected using similar sample sizes to ours. Phenotypic traits for which associations at $p < 10^{-20}$ have been found in the DGRP include resistance to insecticides (Duneau et al., 2018; Green et al., 2019) and viruses (Magwire et al., 2012), and the composition of the cuticular hydrocarbon profile (Dembeck et al., 2015) and microbiome (Everett et al., 2020). It is notable that all of these traits are putatively under weak selection compared to the fitness traits studied here (at least in populations where the relevant insecticides and viruses are absent), which might explain the existence of polymorphisms with major phenotypic effect for these traits (Mousseau & Roff, 1987). We are unaware of any counterexamples, i.e., phenotypes that are probably under strong selection in all environments (e.g., body size, fecundity), for which major effect loci have been found in the DGRP; a formal literature review is in preparation. We therefore speculate that our failure to detect segregating alleles with very strong effects on fitness, at least among loci with MAF ≥ 5%, may reflect their genuine absence as a result of past selection, rather than a false negative due to limited power.

Although our results suggest that genetic polymorphisms with strong fitness effects are rare, the data suggest that fitness is affected by vast numbers of polymorphic loci. Figure 3 illustrates that even for groups of loci with an extremely small estimated mean effect on female fitness, their mean effect on male fitness was predictable despite being estimated from independent experiments. Replicating effect sizes like this should only be possible if large numbers of loci have small but genuine positively pleiotropic effects on both traits (Boyle et al., 2017). This conclusion regarding the distribution of fitness effects agrees with estimates from empirical studies using a different method, namely comparing nucleotide substitution rates at putatively neutral versus non-neutral loci (Eyre-Walker & Keightley, 2007). Molecular evolution studies also suggest that most mutations have a weak, deleterious effect on fitness; for example, one concluded that no more than 16% of mutations were effectively neutral in *Drosophila*, and the majority of the remaining 84% were weakly deleterious (Eyre-Walker & Keightley, 2007). Our conclusion also accords with GWAS-based findings from other taxa, including a recent analysis of birth weight (a trait closely correlated with fitness) in red deer, which concluded that this trait is highly heritable and polygenic, yet no large-effect loci were found (Gauzere et al., 2023). It is also clear from first principles that fitness should be affected by large numbers of loci, since fitness depends on many other highly polygenic traits, such as morphology, condition, and life history strategy. Our finding that fitness variation is closely related to the load of deleterious mutations also aligns with arguments that fitness is a large mutational target (e.g., Rowe & Houle, 1997; Whitlock & Agrawal, 2009), and that natural populations carry a large genetic load (Bertorelle et al., 2022).

Although sexually antagonistic loci appear to be rare, we found that the X chromosome contained more loci with a comparatively high evidence ratio favoring antagonism relative to the autosomes. The X is predicted to be a hotspot for sexually antagonistic polymorphisms, at least under certain assumptions (Fry, 2010; Rice, 1984; Ruzicka & Connallon, 2020, 2022). Our result is consistent with this idea, echoing some but not all previous findings in *Drosophila* and other species (reviewed in Ruzicka & Connallon, 2020). We also found that candidate antagonistic loci had higher than average MAFs. This result is also consistent with theory, because antagonistic loci are generally expected to be under relaxed directional selection or balancing selection (Connallon & Chenoweth, 2019; Connallon & Clark, 2012, 2014; Zajitschek & Connallon, 2018). However, the observed correlation between MAF and the evidence for sexual antagonism might be a statistical artifact, because it is easier to detect sexual antagonism for loci which have high MAF, such that our data do not provide a reliable test of this prediction. Experimental evolution would provide a more powerful test: one could identify a sexually antagonistic locus, set up replicate populations carrying either high or low frequencies of the male-beneficial allele, and compare the allele's evolutionary trajectories in each population to theoretical predictions (Jardine, 2022).

Regarding the evolution of aging, none of our results matched predictions from the antagonistic pleiotropy theory. No loci or transcripts had clearly opposing associations with early- and late-life fitness, and the frequency of age-antagonistic loci and transcripts was estimated as 0%, although we cannot rule out the existence of loci with age-antagonistic fitness effects (especially weak ones) with high confidence. Furthermore, the genetic correlation between early- and late-life fitness was high but significantly lower than unity, suggesting some loci have variable effects on fitness (in magnitude, and perhaps also sign) in different age categories. We also found that the positive genetic correlation between early- and late-life fitness was weaker in females than males, as expected if there is a genetically-based trade-off between female fecundity and late-life fitness (e.g., Partridge et al., 2005). By contrast, some results matched predictions from the mutation accumulation theory. First, in males, the heritability

of late-life fitness was significantly higher than for early-life fitness (as found in some other study systems, e.g., Pettay et al., 2008), as predicted if mutations that impair late-life but not early-life fitness are under relaxed selection. Second, we found a statistically significant decline in the proportion of loci that are putatively sexually antagonistic between the early- and late-life fitness measurements, and the intersex genetic correlation for fitness became nonsignificantly more positive with age. These results are what one would predict if Haldane's "selection shadow" relaxes selection on alleles with harmful, sexually concordant effects in later life stages, while alleles with harmful, sexually concordant effects on early-life fitness are purged. Third, we found that the negative correlation between mutation load and fitness became nonsignificantly stronger with age, as expected if aging unmasks deleterious mutations that have little or no effect on early-life fitness. This result aligns with a recent finding that deletion mutations tend to have stronger negative fitness effects with increasing age in *Drosophila* (Brengdahl et al., 2023).

In the "TWAS" analysis of gene expression, the correlation between male-specific mean expression level and mean male fitness was opposite in sign to the correlation between female-specific mean expression level and mean female fitness for many transcripts, across DGRP lines. Innocenti and Morrow (2010) similarly found many putative sexually antagonistic transcripts in a microarray study of 15 hemiclonal lines. Although caution is warranted because this is a correlational result based on genotype mean expression levels in whole bodies, this finding implies that over- or underexpression of some transcripts (relative to the average expression level for that sex) might elevate fitness in one sex and lower it in the other. If true, this would suggest that genetic polymorphisms that affect gene expression (eQTLs; Huang et al., 2015) would be sexually antagonistic under some circumstances, for example if one allele at an eQTL has an unconditionally positive effect on expression yet the optimum expression level differs between sexes. eQTLs have long been hypothesized to be battlegrounds for intralocus sexual conflict (Connallon & Knowles, 2005; Ellegren & Parsch, 2007; Mishra et al., 2022), though to our knowledge we currently lack strong experimental evidence that this is the case; our results provide further circumstantial evidence. It is also notable that sexually antagonistic transcripts appear much more common than sexually antagonistic loci. One speculative explanation is that there are relatively few eQTLs with major, additive effects on expression, and instead the eQTLs' effects are sensitive to the genetic or environmental context (G×G, G×E, or G×G×E effects; Huang et al., 2015), resulting in a stronger correlation between the transcriptome and fitness (across homozygous genotypes) than between single-locus genotype and fitness. Such nonadditive effects could arise if, for example, the effect on expression of mutations in *trans*-acting expression regulators (such as transcription factors) depends upon sequence variation in *cis*-regulatory elements (such as promoters and enhancers), such that the one allele of the *trans* element creates a male-beneficial expression profile in some genetic backgrounds and a female-beneficial profile in others.

In conclusion, we find that selection is largely aligned across sexes and age classes. Genetic variation for fitness is plentiful, and comprises large numbers of relatively common alleles with weak effects, as well as rare alleles with larger effects. We expect that in this population, selection on males causes a correlated response in female traits that is overall beneficial for female fitness (Whitlock & Agrawal, 2009) and that selection on young individuals mitigates senescence (Maklakov & Chapman, 2019), though loci with effects that differ in magnitude and sometimes sign across sexes or age categories are likely to exist.

## Supplementary material

Supplementary material is available online at *Evolution*.

## Data availability

All analysis code and newly collected data can be viewed at https://lukeholman.github.io/fitnessGWAS/.

## Author contributions

H.W.S.W. conducted and codesigned the fitness assays. L.H. conceived the project, analyzed the data, and wrote the manuscript.

## Funding

This work benefited from Australian Research Council funding to L.H. (DE140101481 and DP170100772).

*Conflict of interest*: The author declares no conflict of interest.

## Acknowledgments

We are very grateful to Michael Murray and Charles Robin for providing advice, laboratory space, and Drosophila resources, Paul Battley for performing EcoRI RFLP genotyping, and Tom Keaney for helpful discussion.

## References

Agrawal, A. F. (2001). Sexual selection and the maintenance of sexual reproduction. *Nature*, 411(6838), 692–695. https://doi.org/10.1038/35079590

Agrawal, A. F., & Whitlock, M. C. (2011). Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics*, 187(2), 553–566. https://doi.org/10.1534/genetics.110.124560

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., Jacq, C., Jensen, A. J., Johnston, S. E., Karlsson, S., Kent, M., Moen, T., Niemelä, E., Nome, T., Næsje, T. F., Orell, P., Romakkaniemi, A., Sægrov, H., Urdal, K., … Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, 528(7582), 405–408. https://doi.org/10.1038/nature16062

Berger, D., Grieshop, K., Lind, M. I., Goenaga, J., Maklakov, A. A., & Arnqvist, G. (2014). Intralocus sexual conflict and environmental stress. *Evolution*, 68(8), 2184–2196. https://doi.org/10.1111/evo.12439

Bertorelle, G., Raffini, F., Bosse, M., Bortoluzzi, C., Iannucci, A., Trucchi, E., Morales, H. E., & van Oosterhout, C. (2022). Genetic load: Genomic estimates and applications in non-model animals. *Nature Reviews Genetics*, 23(8), 492–503. https://doi.org/10.1038/s41576-022-00448-x

Bonduriansky, R., & Chenoweth, S. F. (2009). Intralocus sexual conflict. *Trends in Ecology & Evolution*, 24(5), 280–288. https://doi.org/10.1016/j.tree.2008.12.005

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell*, 169(7), 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Brengdahl, M. I., Kimber, C. M., Shenoi, V. N., Dumea, M., Mital, A., & Friberg, U. (2023). Age-specific effects of deletions: Implications for aging theories. *Evolution*, 77(1), 254–263. https://doi.org/10.1093/evolut/qpac027

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *American Journal of Human Genetics*, 103(3), 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Chippindale, A. K., Gibson, J. R., & Rice, W. R. (2001). Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, 98(4), 1671–1675. https://doi.org/10.1073/pnas.98.4.1671

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, 6(2), 80–92. https://doi.org/10.4161/fly.19695

Collet, J. M., Fuentes, S., Hesketh, J., Hill, M. S., Innocenti, P., Morrow, E. H., Fowler, K., & Reuter, M. (2016). Rapid evolution of the intersexual genetic correlation for fitness in *Drosophila melanogaster*. *Evolution*, 70(4), 781–795. https://doi.org/10.1111/evo.12892

Connallon, T., & Chenoweth, S. F. (2019). Dominance reversals and the maintenance of genetic variation for fitness. *PLoS Biology*, 17(1), e3000118. https://doi.org/10.1371/journal.pbio.3000118

Connallon, T., & Clark, A. G. (2012). A general population genetic framework for antagonistic selection that accounts for demography and recurrent mutation. *Genetics*, 190(4), 1477–1489. https://doi.org/10.1534/genetics.111.137117

Connallon, T., & Clark, A. G. (2014). Balancing selection in species with separate sexes: Insights from Fisher's geometric model. *Genetics*, 197(3), 991–1006. https://doi.org/10.1534/genetics.114.165605

Connallon, T., Cox, R. M., & Calsbeek, R. (2010). Fitness consequences of sex-specific selection. *Evolution*, 64(6), 1671–1682. https://doi.org/10.1111/j.1558-5646.2009.00934.x

Connallon, T., & Hall, M. D. (2016). Genetic correlations and sex-specific adaptation in changing environments. *Evolution*, 70(10), 2186–2198. https://doi.org/10.1111/evo.13025

Connallon, T., & Knowles, L. L. (2005). Intergenomic conflict revealed by patterns of sex-biased gene expression. *Trends in Genetics*, 21(9), 495–499. https://doi.org/10.1016/j.tig.2005.07.006

Dembeck, L. M., Böröczky, K., Huang, W., Schal, C., Anholt, R. R. H., & Mackay, T. F. C. (2015). Genetic architecture of natural variation in cuticular hydrocarbon composition in *Drosophila melanogaster*. *eLife*, 4, e09861.

Duneau, D., Sun, H., Revah, J., San Miguel, K., Kunerth, H. D., Caldas, I. V., Messer, P. W., Scott, J. G., & Buchon, N. (2018). Signatures of insecticide selection in the genome of *Drosophila melanogaster*. *G3 (Bethesda, Md.)*, 8(11), 3469–3480. https://doi.org/10.1534/g3.118.200537

Ellegren, H., & Parsch, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics*, 8(9), 689–698. https://doi.org/10.1038/nrg2167

Ellis, L. L., Huang, W., Quinn, A. M., Ahuja, A., Alfrejd, B., Gomez, F. E., Hjelmen, C. E., Moore, K. L., Mackay, T. F. C., Spencer Johnston, J., & Tarone, A. M. (2014). Intrapopulation genome size variation in *D. melanogaster* reflects life history variation and plasticity. *PLoS Genetics*, 10(7), e1004522.

Everett, L. J., Huang, W., Zhou, S., Carbone, M. A., Lyman, R. F., Arya, G. H., Geisz, M. S., Ma, J., Morgante, F., St Armour, G., Turlapati, L., Anholt, R. R. H., & Mackay, T. F. C. (2020). Gene expression networks in the Drosophila Genetic Reference Panel. *Genome Research*, 30(3), 485–496. https://doi.org/10.1101/gr.257592.119

Eyre-Walker, A., & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8(8), 610–618. https://doi.org/10.1038/nrg2146

Flatt, T. (2020). Life-history evolution and the genetics of fitness components in *Drosophila melanogaster*. *Genetics*, 214(1), 3–48. https://doi.org/10.1534/genetics.119.300160

Fry, J. D. (2010). The genomic location of sexually antagonistic variation: Some cautionary comments. *Evolution*, 64(5), 1510–1516. https://doi.org/10.1111/j.1558-5646.2009.00898.x

Gauzere, J., Pemberton, J. M., Slate, J., Morris, A., Morris, S., Walling, C. A., & Johnston, S. E. (2023). A polygenic basis for birth weight in a wild population of red deer (*Cervus elaphus*). *G3 (Bethesda, Md.)*, 13(4), jkad018. https://doi.org/10.1093/g3journal/jkad018

Green, L., Battlay, P., Fournier-Level, A., Good, R. T., & Robin, C. (2019). Cis- and trans-acting variants contribute to survivorship in a naïve *Drosophila melanogaster* population exposed to Ryanoid insecticides. *Proceedings of the National Academy of Sciences of the United States of America*, 116(21), 10424–10429. https://doi.org/10.1073/pnas.1821713116

Grieshop, K., Maurizio, P. L., Arnqvist, G., & Berger, D. (2021). Selection in males purges the mutation load on female fitness. *Evolution Letters*, 5(4), 328–343. https://doi.org/10.1002/evl3.239

Haldane, J. B. S. (1941). *New paths in genetics*. Allen & Unwin.

Halligan, D. L., & Keightley, P. D. (2009). Spontaneous mutation accumulation studies in evolutionary genetics. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 151–172. https://doi.org/10.1146/annurev.ecolsys.39.110707.173437

Harper, J. A., Janicke, T., & Morrow, E. H. (2021). Systematic review reveals multiple sexually antagonistic polymorphisms affecting human disease and complex traits. *Evolution*, 75(12), 3087–3097. https://doi.org/10.1111/evo.14394

Holman, L., & Jacomb, F. (2017). The effects of stress and sex on selection, genetic covariance, and the evolutionary response. *Journal of Evolutionary Biology*, 30(10), 1898–1909. https://doi.org/10.1111/jeb.13149

Huang, W., Carbone, M. A., Magwire, M. M., Peiffer, J. A., Lyman, R. F., Stone, E. A., Anholt, R. R. H., & Mackay, T. F. C. (2015). Genetic basis of transcriptome diversity in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, 112(44), E6010–E6019.

Innocenti, P., & Morrow, E. H. (2010). The sexually antagonistic genes of *Drosophila melanogaster*. *PLoS Biology*, 8(3), e1000335. https://doi.org/10.1371/journal.pbio.1000335

Ivanov, D. K., Escott-Price, V., Ziehm, M., Magwire, M. M., Mackay, T. F. C., Partridge, L., & Thornton, J. M. (2015). Longevity GWAS using the Drosophila genetic reference panel. *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, 70(12), 1470–1478. https://doi.org/10.1093/gerona/glv047

Jardine, M. (2022). *The maintenance of genetic variation by balancing selection* [*PhD thesis*]. University College London.

Kidwell, J. F., Clegg, M. T., Stewart, F. M., & Prout, T. (1977). Regions of stable equilibria for models of differential selection in the two sexes under random mating. *Genetics*, 85(1), 171–183. https://doi.org/10.1093/genetics/85.1.171

Li Richter, X. Y., & Hollis, B. (2021). Softness of selection and mating system interact to shape trait evolution under sexual conflict. *Evolution*, 75(10), 2335–2347. https://doi.org/10.1111/evo.14329

Long, T. A. F., Agrawal, A. F., & Rowe, L. (2012). The effect of sexual selection on offspring fitness depends on the nature of genetic variation. *Current Biology*, 22(3), 204–208.

Lorch, P. D., Proulx, S., Rowe, L., & Day, T. (2003). Condition-dependent sexual selection can accelerate adaptation. *Evolutionary Ecology Research*, 5(6), 867–881.

Mackay, T. F. C., Richards, S., Stone, E. A., Barbadilla, A., Ayroles, J. F., Zhu, D., Casillas, S., Han, Y., Magwire, M. M., Cridland, J. M., Richardson, M. F., Anholt, R. R. H., Barrón, M., Bess, C., Blankenburg, K. P., Carbone, M. A., Castellano, D., Chaboub, L., Duncan, L., & Gibbs, R. A. (2012). The *Drosophila melanogaster* genetic reference panel. *Nature*, 482(7384), 173–178.

Magwire, M. M., Fabian, D. K., Schweyen, H., Cao, C., Longdon, B., Bayer, F., & Jiggins, F. M. (2012). Genome-wide association studies reveal a simple genetic basis of resistance to naturally coevolving viruses in *Drosophila melanogaster*. *PLoS Genetics*, 8(11), e1003057. https://doi.org/10.1371/journal.pgen.1003057

Maklakov, A. A., & Chapman, T. (2019). Evolution of ageing as a tangle of trade-offs: Energy versus function. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 286(1911), 20191604. https://doi.org/10.1098/rspb.2019.1604

Maklakov, A. A., Rowe, L., & Friberg, U. (2015). Why organisms age: Evolution of senescence under positive pleiotropy? *Bioessays*, 37(7), 802–807. https://doi.org/10.1002/bies.201500025

Melzer, D., Pilling, L. C., & Ferrucci, L. (2019). The genetics of human ageing. *Nature Reviews Genetics*, 21(2), 88–101. https://doi.org/10.1038/s41576-019-0183-6

Mishra, P., Barrera, T. S., Grieshop, K., & Agrawal, A. F. (2022). Cis-regulatory variation in relation to sex and sexual dimorphism in *Drosophila melanogaster*. *bioRxiv*. https://doi.org/10.1101/2022.09.20.508724

Mousseau, T. A., & Roff, D. A. (1987). Natural selection and the heritability of fitness components. *Heredity*, 59(Pt 2), 181–197. https://doi.org/10.1038/hdy.1987.113

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. *Biological Reviews of the Cambridge Philosophical Society*, 85(4), 935–956. https://doi.org/10.1111/j.1469-185X.2010.00141.x

Partridge, L., Gems, D., & Withers, D. J. (2005). Sex and death: What is the connection? *Cell*, 120(4), 461–472. https://doi.org/10.1016/j.cell.2005.01.026

Pettay, J. E., Charmantier, A., Wilson, A. J., & Lummaa, V. (2008). Age-specific genetic and maternal effects in fecundity of preindustrial Finnish women. *Evolution*, 62(9), 2297–2304. https://doi.org/10.1111/j.1558-5646.2008.00452.x

Pizzari, T., Biernaskie, J. M., & Carazo, P. (2015). Inclusive fitness and sexual conflict: How population structure can modulate the battle of the sexes. *Bioessays*, 37(2), 155–166. https://doi.org/10.1002/bies.201400130

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3), 559–575.

Rice, W. R. (1984). Sex chromosomes and the evolution of sexual dimorphism. *Evolution*, 38(4), 735–742. https://doi.org/10.1111/j.1558-5646.1984.tb00346.x

Ritchie, M. E., Phipson, B., Wu, D. I., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47–e47. https://doi.org/10.1093/nar/gkv007

Roff, D. A. (1996). The evolution of genetic correlations: An analysis of patterns. *Evolution*, 50(4), 1392–1403. https://doi.org/10.1111/j.1558-5646.1996.tb03913.x

Rohde, P. D., Krag, K., Loeschcke, V., Overgaard, J., Sørensen, P., & Kristensen, T. N. (2016). A quantitative genomic approach for analysis of fitness and stress related traits in a *Drosophila melanogaster* model population. *International Journal of Genomics*, 2016, 1–11.

Rowe, L., & Houle, D. (1997). The lek paradox and the capture of genetic variance by condition dependent traits. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 263(1375), 1415–1421.

Ruzicka, F., & Connallon, T. (2020). Is the X chromosome a hot spot for sexually antagonistic polymorphisms? Biases in current empirical tests of classical theory. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 287(1937), 20201869. https://doi.org/10.1098/rspb.2020.1869

Ruzicka, F., & Connallon, T. (2022). An unbiased test reveals no enrichment of sexually antagonistic polymorphisms on the human X chromosome. *Proceedings Biological Sciences*, 289(1967), 20212314. https://doi.org/10.1098/rspb.2021.2314

Ruzicka, F., Connallon, T., & Reuter, M. (2021). Sex differences in deleterious mutational effects in *Drosophila melanogaster*: Combining quantitative and population genetic insights. *Genetics*, 219(3), iyab143. https://doi.org/10.1093/genetics/iyab143

Ruzicka, F., Hill, M. S., Pennell, T. M., Flis, I., Ingleby, F. C., Mott, R., Fowler, K., Morrow, E. H., & Reuter, M. (2019). Genome-wide sexually antagonistic variants reveal long-standing constraints on sexual dimorphism in fruit flies. *PLoS Biology*, 17(4), e3000244. https://doi.org/10.1371/journal.pbio.3000244

Singh, A., Hasan, A., & Agrawal, A. F. (2023). An investigation of the sex-specific genetic architecture of fitness in *Drosophila melanogaster*. *Evolution*, 77(9), 2015–2028. https://doi.org/10.1093/evolut/qpad107

Urbut, S. M., Selvaraj, M. S., Pampana, A., Dattilo, L., Neale, B., ODonnell, C. J., Peloso, G., & Natarajan, P. (2021). Bayesian multivariate genetic analysis of lipids improves translational insight. *Circulation*, 144(Suppl 1), A9855–A9855.

Urbut, S. M., Wang, G., Carbonetto, P., & Stephens, M. (2019). Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nature Genetics*, 51(1), 187–195. https://doi.org/10.1038/s41588-018-0268-8

Whitlock, M. C., & Agrawal, A. F. (2009). Purging the genome with sexual selection: Reducing mutation load through selection on males. *Evolution*, 63(3), 569–582. https://doi.org/10.1111/j.1558-5646.2008.00558.x

Wilson, A. J., Nussey, D. H., Pemberton, J. M., Pilkington, J. G., Morris, A., Pelletier, F., Clutton-Brock, T. H., & Kruuk, L. E. B. (2007). Evidence for a genetic basis of aging in two wild vertebrate populations. *Current Biology*, 17(24), 2136–2142.

Zajitschek, F., & Connallon, T. (2018). Antagonistic pleiotropy in species with separate sexes, and the maintenance of genetic variation in life-history traits and fitness. *Evolution*, 72(6), 1306–1316. https://doi.org/10.1111/evo.13493

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44(7), 821–824. https://doi.org/10.1038/ng.2310