

Sexually discordant selection is associated with trait-specific morphological changes and a complex genomic response

Tyler Audet¹, Joelle Krol¹, Katie Pelletier^{1,3}, Andrew D. Stewart², Ian Dworkin^{1,4}

¹Department of Biology, McMaster University, Hamilton, ON, Canada

²Department of Biology, Canisius University, Buffalo, NY, United States

³Present address: Institut de Biologie de l'Ecole Normale Supérieure, CNRS, Paris, France

Corresponding authors: Department of Biology, McMaster University, Hamilton, ON, Canada. Email: dworkin@mcmaster.ca; Department of Biology, Canisius University, Buffalo, NY, United States. Email: stewar34@canisius.edu

Abstract

Sexes often have differing fitness optima, potentially generating intra-locus sexual conflict, as each sex bears a genetic “load” of alleles beneficial to the other sex. One strategy to evaluate conflict in the genome is to artificially select populations discordantly against established sexual dimorphism (SD), reintroducing attenuated conflict. We investigate a long-term artificial selection experiment reversing sexual size dimorphism in *Drosophila melanogaster* during ~350 generations of sexually discordant selection. We explore morphological and genomic changes to identify loci under selection between the sexes in discordantly and concordantly size-selected treatments. Despite substantial changes to overall size, concordant selection maintained ancestral SD. However, discordant selection altered size dimorphism in a trait-specific manner. We observe multiple possible soft selective sweeps in the genome, with size-related genes showing signs of selection. Patterns of genomic differentiation between the sexes within lineages identified potential sites maintained by sexual conflict. One discordant selected lineage shows a pattern of elevated genomic differentiation between males and females on chromosome 3L, consistent with the maintenance of sexual conflict. Our results suggest visible signs of conflict and differentially segregating alleles between the sexes due to discordant selection.

Keywords: evolutionary genomics, genetic variation, morphological evolution, artificial selection, sexual selection, sex

Introduction

While in many species, the sexes show phenotypic differences, they must also use nearly the same genome during development to express these phenotypes. This sexual dimorphism (SD) evolves despite the fact that the sexes have a high genetic correlation (r_{MF}) for many traits, which can hinder the evolution of SD (Lande, 1980). The extent to which SD can evolve depends on the strength and direction of selection, additive genetic variance, and the r_{MF} (Delph et al., 2011; Lande, 1980). Many organisms show high r_{MF} , which could limit the potential for sexually discordant evolution (Lande, 1980; Poissant et al., 2010). Despite this, SD is very common in nature, particularly, sexual size dimorphism (SSD; reviewed in Fairbairn et al., 2007). In many insects, including the pomace fly *Drosophila melanogaster*, females are most often the larger sex (Ashburner, 1989). This dimorphism is likely due to the relative contribution of increased fecundity with increased size in females (Honěk, 1993; Reeve & Fairbairn, 1999). Male-biased SSD occurs in some insects, often due to the relative contribution of sexual selection, and can evolve rapidly within a clade (Emlen et al., 2005; Luecke & Kopp, 2019; Moczek et al., 2006). When loci impact the phenotype in a way favoring one sex but disadvantageous to the other,

it may create intra-locus sexual conflict (IASC). This IASC combined with a high r_{MF} between the sexes may impede the rate of evolution if genetic variation is not sufficiently high. Despite high r_{MF} values, family-based artificial selection experimental designs, where phenotypic selection is based on either within or among family trait values, have demonstrated that sex-specific responses can occur relatively rapidly (Alicchio & Palenzona, 1971; Bird & Schaffer, 1972; Eisen & Hanrahan, 1972; Kaufmann et al., 2021). Not only can changes in SSD occur rapidly with family-based artificial selection, but r_{MF} has been directly selected upon and degraded in just a few generations (Delph et al., 2011). This reduction of r_{MF} allowed one sex to be selected for size with minimal response in the unselected sex, while sex-specific selection in high r_{MF} lines resulted in a strongly correlated response in the unselected sex. Although family-based selection experiments demonstrate additive genetic variation for a response to sexually discordant selection, these do not reflect the transmission of allelic effects in most natural populations. As such, approaches based on mass artificial selection may better reflect the transmission of allelic effects in natural populations because alleles of interest may be rarer, and each round of selection will be subject to drift and recombination with random haplotypes from other individuals. Using strong, long-term, mass artificial selection,

Received August 31, 2023; revisions received April 12, 2024; accepted May 7, 2024

Associate Editor: Charles Baer; Handling Editor: Tim Connallon

© The Author(s) 2024. Published by Oxford University Press on behalf of The Society for the Study of Evolution (SSE).

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

[Stewart and Rice \(2018\)](#) demonstrated that a response to a sex-discordant selection pressure can occur. [Stewart and Rice \(2018\)](#) successfully selected body size in *D. melanogaster* in a sex-discordant manner over 250 generations, with a measurable phenotypic response requiring more than 100 generations of artificial selection. In comparison, sex-concordant selection for body size resulted in rapid phenotypic responses. Despite discordant selection responding in family and mass selection experimental designs, one previous study using *Tribolium castaneum* found little response to sex-discordant selection ([Tigreiros & Lewis, 2011](#)) despite a rapid response to sexually concordant selection on pupal mass. This suggests barriers to divergent response in SD may exist in some populations, although it may reflect modest genetic variation in the ancestral populations and the limited number (7) of generations of artificially selected applied.

The sex-concordant (hereafter, concordant) selection lineages established by Stewart and Rice have had their genomes sequenced at generation 100 to identify candidate SNPs associated with body size variation ([Turner et al., 2011](#)). The sex-discordant (hereafter, discordant) lineages have not previously been sequenced and are the focus of this current study. To date, functional genetic analyses in *Drosophila* have implicated a few pathways involved with sex determination and growth that can influence SSD. Manipulation of sex-specific splice variants of the *transformer* (*tra*) gene reduces female size, reducing (but not eliminating) SSD ([Rideout et al., 2015](#)). Increasing *tra* expression showed sex-specific increases in size in females ([Rideout et al., 2015](#)). A duplication of the *diminutive* (*dyc*) gene on the X chromosome resulted in males 12%–14% larger, and when paired with constitutive expression of *tra*, SSD was substantially diminished ([Mathews et al., 2017](#)). Upstream of *tra*, tissue-specific depletion of *sex-lethal* (*sxl*) in neurons led to a reduction in female body size ([Sawala & Gould, 2017](#)). Inhibition of insulin signaling had sex-specific impacts on body size, largely reducing female body size. In contrast, upregulation of inhibition of insulin-signaling increased male body size ([Millington et al., 2021b](#)). These experiments demonstrate pathways involved with phenotypic expression of SD, but it is unclear whether segregating variation in these pathways contributes to natural phenotypic variation. Although alleles of large effect in any of the above genes would be exciting to find in a natural population, it is not a safe assumption that genes showing a phenotypic response in a lab setting will be the genes selected upon if a population undergoes discordant selection. Further, rapid response to selection on size, as demonstrated by [Stewart and Rice \(2018\)](#) and [Bird and Schaffer \(1972\)](#), among others, suggests that a polygenic response on segregating genetic variation, rather than *de novo* mutations, likely mediates short-term responses for body size evolution. In natural populations, segregating alleles contributing to variation in SD interest could be maintained in the population by selection or simply reflect mutation-selection-drift balance.

Populations of *D. melanogaster* harbour alleles with both sexually discordant and concordant effects on fitness ([Rice & Chippindale, 2001](#)). Variation in body size in *D. melanogaster* is highly polygenic ([Carreira et al., 2009](#); [Turner et al., 2011](#)). Sex-limited selection of males has been shown to incur a fitness cost to females ([Prasad et al., 2007](#)). This limited male evolution also results in a change in body size in nonselected females, which is closer to the male optimum ([Prasad et al., 2007](#)). A decrease in female fitness (and body

size) when males are allowed to evolve toward their own optimum without parallel female evolution suggests that there is unresolved sexual conflict in the genome of *D. melanogaster* pertaining to body size. Alleles potentially under conflict in the genomes of the outbred LH_M population of *D. melanogaster* have been identified via examination of male mating success and female fecundity ([Ruzicka et al., 2019](#)). Alleles under IASC influencing body size can help understand how the genome responds to selection with divergent phenotypic optima. The selection of body size in a discordant manner has the potential to answer questions about how the shared genome overcomes high r_{MF} when selection favors divergent phenotypes across the sexes.

Using lineages evolved under sexually discordant selection for size, first described in [Stewart and Rice \(2018\)](#), we demonstrate two important findings. First, despite selection for a trait-agnostic (i.e., selection on general size rather than an increase in mass or thorax length) measure of size, trait-specific patterns of SSD reversal are the norm. Second, despite the long-term sexually discordant selection, we see relatively weak evidence for sexually antagonistic alleles being maintained. We do, however, identify one region potentially segregating differentially between the sexes, a possible sign of unresolved (or reintroduced) sexual conflict. Using an evolve and sequence approach, we examined lineage patterns of genomic differentiation and within-lineage among the sexes. We discuss our findings within the context of both the evolution of SD and the potential role of ongoing intra-locus sexual conflict.

Methods

Lineages

The populations used are part of a long-term experiment on size evolution ([Stewart & Rice, 2018](#); [Turner et al., 2011](#)). The selection lineages examined were started using the outbred population, LH_M , previously adapted to the lab for over 350 generations. These flies are maintained in discrete 2-week generations at moderate density (approximate 200 eggs per 10 ml standard molasses food vial; [Supplementary Table S1](#)). The complete methodology for the maintenance of the base population has been published previously ([Rice et al., 2005](#)). For size selection in each generation, flies are anesthetized using CO_2 and sorted using a motorized stacked sieving device in which each successive sieve is 5% smaller than the sieve above. The largest sieve used had aperture diameters of 2,000 μm , and the smallest sieve had apertures measuring 850 μm . In each generation, all flies (~1,800 individuals) from a selection lineage were sieve sorted, and 10 vials of 16 mating pairs (320 total flies per lineage) selected based on phenotype were used to generate the next generation. One treatment used only the smallest flies of both sexes (S; concordant selection), one treatment the largest of each sex (L; concordant selection), and the reversal of dimorphism (discordant selection) treatment used the smallest females and largest males (E). Finally, a control treatment (C) was populated with flies that passed through the sieves but were not selected based on size. Two independent replicate lineages were maintained for each selection treatment. During subsequent selection for size with the established protocol, it was observed that selection could be more stringent on small flies than large ones because of the nature of the sieve sorting (hindrances due to appendages sticking out, blocking the passage of flies through the

sieve). This would imply that in the sex-discordant selection, females selected to be small were under increased selective pressure compared to males selected to be large. This suggests greater pressure for selection in one direction, and although this does not explain all the results, it may influence responses to sexually discordant selection.

Morphological measures and analysis

At generation 367 (August 19, 2019), flies from each lineage were collected and stored in 70% ethanol for dissection and measurement of traits. Individuals were chosen randomly before and after the selection treatment to get an accurate estimate of size in the overall lineage. Individuals were dissected under a Leica M125 microscope, and legs and thoraces were imaged at 63 \times using a DCF-400 camera. A minimum of 20 flies (for each sex) were dissected from each lineage by dissecting off the first (pro-thoracic) right leg, then imaging the left side of the thorax of each individual. The leg and left side of the thorax were mounted on slides (in 70% glycerol in PBS, with a small amount of phenol as a bacteriostatic agent). The femur, tibia, and first tarsal segment of each leg and each thorax were measured using ImageJ version 1.52q (Schneider et al., 2012). We measured thorax length as a proxy for overall size and a proxy for the general pattern of the female-biased SSD found in *Drosophila melanogaster*. We also measured each leg segment; measurement was completed from the center of the beginning of the segment to the center of the end of the segment.

Analysis of the leg and thorax measurements was completed using R v4.1.3 (R Core Team, 2021). One individual measure of the femur was removed as its measure was ~100 \times smaller than the mean for the trait. For all analyses of morphology, we used \log_2 transformed trait values to facilitate inferences of proportional changes in dimorphism. We fit general linear mixed models for individual traits, including sex, selection, and sampling, and their 2nd order interactions as fixed effects. We allowed sex effects to vary as a random effect of replicate lineage (i.e., random slopes). Models were fit using *lmer* in the *lme4* package v1.1.30 (Bates et al., 2015). We confirmed the results with a fully multivariate mixed model for all traits. For the multivariate mixed model, we allowed for random effects of sex and traits by lineage, as well as accounting for within-individual variation across traits. The initial fit of this model using *lmer* was singular, likely due to variance estimates getting “stuck” on a boundary (0). We employed two approaches to deal with this. First, we employed a Bayesian extension of our model using *blmer* in *blme* v1.0.5 (Chung et al., 2013). This employs weak regularizing priors (away from 0 for the variances). We confirmed the stability of fixed effects using a second approach, fitting a general latent-variable mixed model, to estimate reduced-rank covariance matrices for random effects, as implemented in *glmmTMB* v1.1.7 (Brooks et al., 2017; Kristensen & McGillyuddy, 2023; Niku et al., 2019). Estimated marginal means, custom contrasts, and associated confidence intervals were estimated using *emmeans* v1.8.0 (Lenth et al., 2018). Visualization was done using *ggplot2* v3.3.6 (Wickham, 2018). These approaches provided very similar fixed effects estimates to each other and to the single trait models, which were used for downstream analysis.

Despite the artificial selection being “trait-agnostic” (selecting on a composite cross-sectional area, with possible hindrance from appendages) and a relatively consistent response

as outlined below in the results, we examined changes in multivariate allometry among traits within each selective treatment. Commonly, the first principal component derived from a variance-covariance matrix of log-transformed morphological measures captures a measure of overall size (Blackith & Reymont, 1971; Jolicoeur, 1963; Klingenberg, 1996). It is common to assess whether the log-transformed traits contribute approximately equally to the size of PC1 (isometry), with expected loadings of $1/\sqrt{p}$, where p is the number of traits (dimensions). In addition to examining (and visualizing) these vectors from principal component analyses conducted by each selective treatment, we compared aspects of the orientation and structure of the variance-covariance matrices across treatments.

Evaluating sex ratio of lineages and crosses among them

While phenotyping flies for an unrelated experiment, a deviation of the expected 50/50 sex ratio was observed in a cross between control lineage 1 and discordant lineage 1 samples. As summarized in the results, discordant lineage 1 is the lineage with potential evidence of maintained conflict on chromosome 3L. As such, we conducted experimental crosses to examine adult sex ratios and evaluate whether there is a consistent deviation in sex ratio, potentially due to genomic conflict. At generation 464, 15 males and 15 females were taken from each treatment for single-pair matings to an opposite-sex individual from the LH_M population. Single-pair reciprocal matings were allowed to lay until larvae were visible in the food (~5 days) before F₀ pairs were placed in 70% ethanol. From each F₁ vial, a single-pair was used to generate an F₂, while the rest of the F₁ individuals were stored in 70% ethanol after being allowed to eclose until most pupal casings were visibly empty. Adults from the F₂ generation, once mostly eclosed, were stored in 70% ethanol. The number of male and female offspring in F₁ and F₂ generations were counted.

We modeled the data for the sex ratio crosses using logistic regression (*glm* in R), with counts of males and females from each cross as the unit of sampling, with lineage and cross direction and their interactions as predictors. From the model fits, we computed estimates and their confidence intervals on the response scale using *emmeans* to facilitate interpretability. We did this both with and without the reciprocal direction of genetic crosses (whether or not the individual was treated as a sire or dam) in the model. During the experiment to examine adult sex ratios, we noticed substantial differences in the number of individual offspring. While this was not a planned analysis (and should be treated as such), we examined differences in fecundity (assessed by the census of adults) fitting a general linear model (fit using *lm*) of the number of offspring regressed onto the direction of the cross, selection treatments, and their interactions.

Genomic sample preparation

At generation 378 (February 17, 2020), flies from each lineage were collected in 70% ethanol and stored at -20 °C (Figure 1). For each lineage and sex, flies were separated into four pools of 25 individuals, and DNA was extracted using a column-based DNA extraction kit (DNeasy Qiagen kit, Cat # 69506). The extracted DNA from each of the four pools of 25 for distinct samples was combined so that the same concentration of DNA was added from each pool (equimolar). This resulted in pools of 100 individuals for each sequenced



Figure 1. Example samples from control and discordant lineages. Stereomicroscope images of Control and discordant males and females (25 \times magnification). Flies chosen at random from populations stored and imaged in 70% ethanol.

combination of sex and unique lineage. This 25-individual pooling was done due to the size of the columns not being capable of extracting from 100 individuals at once. Library preparation and sequencing was done by Génome Québec (Centre d'expertise et de services, Génome Québec). Libraries were prepared with IDT dual index adapters. Sequencing was done using Illumina NovaSeq 6000 to an average coverage of 200 \times with 151 bp sequence fragments. Initial sequencing fell short of 200 \times coverage, so additional sequencing from the same libraries was conducted to “top-up” coverage on some samples and were merged with their corresponding samples after mapping.

Bioinformatic pipeline

Supplementary Figure S1 shows a summary of the bioinformatics pipeline. A detailed summary can be found at (https://github.com/DworkinLab/Audet_et_al_Evolution_2024). Reads were trimmed using bbduk v38.90 (Bushnell, 2021), aligned using BWA-mem version 0.7.8 (Li, 2013), GATK v3.8 was used to mark indels and perform local realignment. SAMtools v1.5 (Li et al., 2009) was used to convert SAM files to BAM, extract out the core genome, mark/remove read groups, and create mpileups. Three distinct mpileups were made: one with sexes and lineages separate, one with sexes merged, maintaining lineages, and one where treatments were pooled together by merging both sexes and lineages together; each of these three went through the same SNP calling and file generation methods. Popoolation v1.2.2 (Kofler et al., 2011) was used to mask repetitive regions identified with Repeat Masker v4 (Smit et al., 2013–2015), as well as indels for SNP calling. PoolSNP v1 (Kapun et al., 2020) was used to call variants. From this VCF file, indels were masked using custom scripts from the DrosEU pipeline (Kapun et al., 2020). Due to issues generating sync files from poolSNP VCF output files, a

sync file was also generated from the mpileup using Grenedalf version 0.2.0 (Czech et al., 2023, pre-print). This sync file was filtered for sites present in our SNP-called VCF file to create an SNP-called sync file. For “sexes separate” as well as “treatments pooled” sync files. Through testing, we observed that changes in the bioinformatic pipeline, such as which program was used, changed the results slightly, pointing at artifacts introduced by bioinformatic programs. This suggests that great hesitancy and meticulousness must be applied when selecting programs for bioinformatic analyses, as these choices certainly change final results in, at best, a small way. Bioinformatics was done on Compute Canada servers on the Graham cluster.

Among population genomic analyses

To identify variants that potentially contribute to phenotypic divergence between treatments, we examined three measurements of population differentiation between populations or genetic diversity within populations. To assess variation between populations, we used two related approaches, F_{ST} as a measure of the magnitude of change in allele frequencies and a modified Cochran–Mantel–Haenszel (CMH) statistic that incorporates sources of sampling variation and genetic drift common to evolve and re-sequence experiments. To assess variation within populations and to identify potential selective sweeps, we performed windowed computation of nucleotide diversity (π). We calculated F_{ST} between control treatments and discordant treatments using Grenedalf v0.2.0 in 10 kb windows along each chromosome (Czech et al., 2023), which corrects for sample sizes and pooled sequencing size, known sources of error in parameter estimation for pooled sequencing. We also examined F_{ST} between Large and Small treatments to follow up on the initial analysis in Turner et al. (2011) to see if their findings were replicated after an

additional ~275 generations of artificial selection on body size. To do so, we ran the sequence data from generation 100 through our pipeline and compared genes of interest from F100 to the genes identified at F378.

To calculate effective population size (N_e) we used poolSeq v0.3.5 (Taus et al., 2017). We estimated N_e between Control 1 and Discordant 1 treatments ($N_e = 181$) and between large 1 and small 1 ($N_e = 69$) and used these to calculate CMH statistics with ACER version 1.0 (Spitzer et al., 2020). We also evaluated masking all sites that were fixed in our time point 0 (control for control vs. discordant, large for large vs. small) using the checkSNP() function. This resulted in higher N_e estimated (C1 vs. E1 = 207.90, C2 vs. E2 = 224.51; L1 vs. S1 = 90.65, L2 vs. S2 = 139.76) so we used the previous smaller estimates. N_e calculations for the first replicate comparison were done using formulae proposed by Jónás et al. (2016) and are intended to work with pooled data as well as control for additional sources of sampling variance in allele frequencies that could bias N_e . Since an ancestral sample was not preserved, drift between control and discordant treatments (or large versus small) was used, and the generations of drift were set at 750 from the control treatment (~generations to convergence ×2).

For our samples, given the large number of generations of artificial selection, selective sweeps would likely result in regions of low nucleotide diversity, which we measured as π using Grenedalf v0.2.0 in 10 kb windows for our control, discordant, and concordant samples. This resulted in us having per sample π values for 10 kb windows, F_{ST} values for 10 kb windows, and SNP-by-SNP adjusted p -values from a CMH test. Using these values, regions with a small π value (<0.0005) were extracted as interesting, and were extracted along with the top 5% of F_{ST} (discordant vs. control cutoff was $F_{ST} 0.51$ with 621 windows meeting these criteria; large vs. small cutoff was $F_{ST} 0.78$ with 621 windows meeting that criteria) and SNPs with a CMH adjusted p -value <0.01 (using *p.adjust* in R (Benjamini & Hochberg, 1995); 16,218 SNPs in control vs. discordant, 115,747 SNPs in large vs. small) were extracted. The overlap between all windows or SNPs of interest was intersected with Bedtools v2.31.0 (Quinlan & Hall, 2010) to get SNPs of interest. Unfortunately, most software (including Grenedalf) does not account for the hierarchical structure of populations (common in experimental designs like artificial selection), nor the expected sources of drift and other forms of sampling variance among replicate lineages (within treatments) in evolve and re-sequence experiments. As such, the approach we used above (merging replicate lineages) does not account for sampling variation. To partially account for this, for the subset of “candidate” SNPs we identified, we fit SNP-specific logistic regression models, with SNP counts for each replicate lineage representing units of sampling and with selective treatments as the predictor. Treatment contrast between either discordant versus control or large versus small treatments was obtained using *emmeans*. As this approach is computationally slow, we only did this for the sites we identified as likely candidates (discordant = 20,014, large = 10,649, small = 13,952). From the overlap of all these analyses, we created a file containing a region of interest on chromosome 3L in discordant replicate 1, the interesting discordant sites (Supplementary File 2), interesting large sites (Supplementary File 3), and interesting small sites (Supplementary File 4). These sites were extracted from the SNP-called VCF file using bedtools v2.30.0 (Quinlan & Hall, 2010). Using this VCF file,

a sync file was created, from which an allele frequency table for all sites of interest was made using Grenedalf, with the reference column set as the control replicate 1 major allele for the above-mentioned modeling.

Sites of interest were annotated (Supplementary Files 1, 2, 3, and 4). Gene ontology (GO) enrichment analyses were conducted Gowinda v1.12 (Kofler & Schlötterer, 2012), which controls for gene length and uses permutations to reduce false positives. Gene annotations were conducted using SNPeff version 5.1 (Cingolani et al., 2012).

Within population genomic analyses to identify regions of genomic conflict

To identify genomic regions that may harbour variants contributing to IASC, we examined patterns of genomic differentiation between males and females from within the same experimental lineage and generation for each of the evolved lineages (Cheng & Kirkpatrick, 2016; Kasimatis et al., 2019; Lucotte et al., 2016). Previous simulation analyses have suggested that identifying IASC using intersex F_{ST} generally has low power (Kasimatis et al., 2019). However, the experimental design used for artificial, sexually discordant selection is well suited for this particular approach given strong and persistent selection across many generations. We computed allele frequency tables for each sex within each lineage, along with F_{ST} between males and females within lineages in Grenedalf (window size 5 kb). As summarized in the results, we observed a region on chromosome 3L showing elevated F_{ST} , in a single discordant treatment lineage (E1).

In addition to the possibility that this elevated region on chromosome 3L was due to maintained sexual conflict, there are several other possible explanations that we evaluated. To assess whether this elevated region of F_{ST} could be accounted for simply by sampling variation, we simulated 1,000,000 sites, sampling the full range of allele frequencies used in our analyses and simulating allele frequencies for male and female samples drawn from a common allele frequency in each simulation. We then allowed sequencing depth (based on the approximate empirical distribution) to vary for each sex. We plotted simulated allele frequencies for males and females and over-plotted observed allele frequencies for each lineage to determine whether the male-female differences in allele frequencies in this chromosomal region were extreme relative to distributions under our simulations that modeled sources of sampling variance. Additionally, we conducted a logistic regression (*glm*) of major and minor allele counts between the sexes. From this, we obtained contrasts and confidence intervals using *emmeans*.

To determine the extent to which demographic influences of lab adaptation and artificial selection in the context of empirically derived estimates of recombination along the chromosome arm could account for the elevated region of F_{ST} between males and females, we performed evolutionary simulations using SLiM v4.1-4.2 (Haller & Messer, 2023). We broke the simulation into two phases. The “burn-in” phase to simulate patterns of variation along chromosome 3L in the natural population from which LH_M was derived to serve as the ancestral population, followed by simulations to assess demographic impacts due to the founding and maintenance of the LH_M population and artificial selection lineages derived from it. To account for variation in recombination rates and their influence on evolutionary dynamics on chromosome 3L, we used empirically derived estimates of a recombination rate map for chromosome

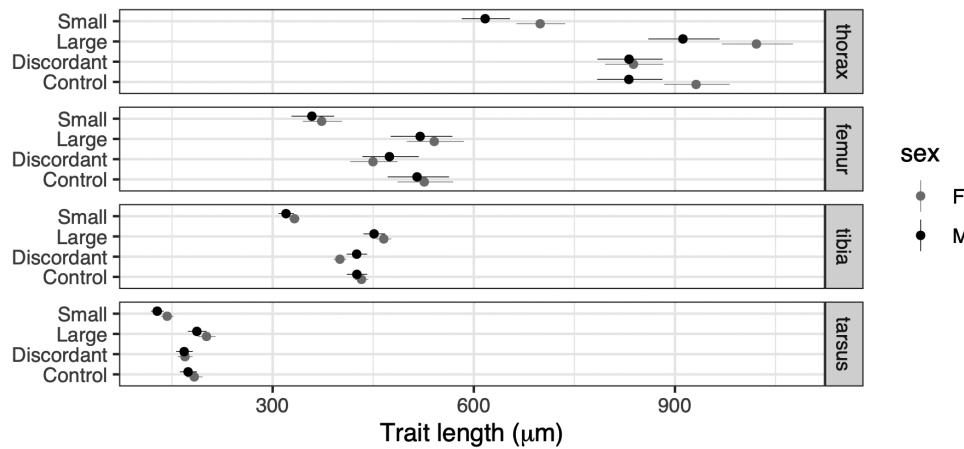


Figure 2. Trait-specific evolution of size among artificially selected lineages. Model estimates for each trait across treatments and sex. Measurements were \log_2 transformed for model fit back-transformed for plotting. Error bars are 95% CIs.

3L (Comeron et al., 2012). As recombination does not naturally occur in *D. melanogaster* males, we simulated sex-specific recombination rates, using the above recombination map for females and 0 for males. For mutation rates, we used the average of empirically derived estimates, $3.2e-9$ (Cingolani et al., 2012; Keightley et al., 2009, 2014) as our per-site estimate. To capture the expected variation for chromosome 3L from the ancestral North American population used to find LH_M , we used an estimated $N_e = 5.54 \times 10^5$ (Arguello et al., 2019). This phase of the simulation was done for 10,000 generations. To speed up computation, we use the protocol suggested in the SLiM manual, rescaling mutation and recombination rates so that the number of individuals that needed to be simulated each generation was reduced by a factor of 20. The recombination rate was rescaled as $r' = 0.5 \times (1 - (1 - 2r)^n)$, where r is the recombination rate, and n is the rescaling factor. This initial simulation is found in the script “SeparateSexesBurnIn.slim.” Thus, this first simulation represents the simulation of genomes (chromosome 3L) of individuals from the ancestral North American population from which all samples were derived. For the second part of the simulations (accounting for demographic effects of the founding and maintenance of LH_M and the artificial selection lineages), we sampled 1,792 individuals (1:1 sex ratio) from the ancestral population and maintained this population size for 360 discrete generations (capturing demographics of the LH_M population). Mutation and recombination rates were used as described above, without the need for parameter rescaling. Following this, we then simulated a population maintained at 320 individuals (1:1 sex ratio) to capture the population size of artificial selection lineages for an additional 377 generations. At the end of each simulation run (“SeparateSexesOutputSamples_Run_OnePop.slim”), we randomly sampled 200 chromosomes each for males and females, outputted allele counts for segregating sites, and after parsing the data into an appropriate format, computed windows of F_{ST} across the length of the simulated chromosome 3L using the same approach as discussed above for empirical data. We performed 100 simulations to assess the variation in variability and within-generation, within-lineage F_{ST} among males and females.

Results

Selection for size resulted in a trait-dependent response when the selection was sex-discordant

Consistent with the previous results for overall “body size” (Stewart & Rice, 2018), all measured traits responded

to artificial selection in the expected directions. Thorax length was measured as a proxy for body size and showed a clear relationship between size and selection lineage ($\chi^2 = 242.72$, $p < 0.0001$; Figure 2; Supplementary Table S2; Supplementary Figures S2 and S3) with some lineage-specific variation (Supplementary Figure S3). Concordant artificial selection for decreased size reduced thorax length relative to control lineages by ~25% in females (26% in males), while selection for increased size increased thorax length by ~9.5% for both sexes. In discordant selection lineages, we found a ~10% decrease in female thorax length and a minimal (~0%) change in males. We observed similar patterns for the length of leg segments. Small treatment males and females decreased in length relative to controls to a similar degree (~30% for femur, ~24% for tibia and tarsus; Figure 3). Large treatment males and females increased in size somewhat more modestly relative to controls (~2% for femur, 6% for tibia, 8% for tarsus). For sexually discordant selection, females decreased in size relative to controls (~14% for the femur, ~7% for the tibia and tarsus). Males from the discordant lineage decreased more modestly relative to controls (~8% for the femur, <1% for the tibia, ~3.5% for the tarsus). While some trait changes have confidence intervals that overlap zero, all traits and treatments seemed to respond in size in the expected direction of effect (Figures 2 and 3; Supplementary Tables S3, S4, and S5).

SD and multivariate allometry diverged only under discordant, sex-specific selection

Despite substantial changes in overall size, SSD remained female-biased in all concordant selection treatment groups (Figures 2 and 3), showing little change in dimorphism relative to control (Figure 3). In contrast, the sexually discordant selection treatment resulted in a substantial reduction in the amount of ancestral female-biased dimorphism relative to the controls (Figure 3). Thorax and tarsus lengths have evolved to be essentially monomorphic, while femur and tibia lengths are now male-biased (~5% increase) in size. The change in the amount of dimorphism varied somewhat by trait but with consistent patterns of change (Figure 3).

Given the results described above, it is not surprising we observed substantial changes in patterns of multivariate allometry (Figure 4; Supplementary Figures S2, S3, and S4). We observe a substantial change in the sex-discordant

lineages, which deviates from the isometry vector (Table 1) that is observed for other *Drosophila melanogaster* populations (Shingleton et al., 2009). We further compared covariance matrices across selective treatments using the Krzanowski correlation. The discordant selection lineages show a reduced correlation to the control lineages ($r_{Krz} = 0.86$) compared with the patterns observed for both sexually concordant selection

treatments ($r_{Krz} = 0.97$ and $r_{Krz} = 0.99$ for large and small, respectively).

Genomic differentiation and nucleotide diversity (π) between selection treatments

F_{ST} between control and discordant selection treatments was highly variable but included substantial differentiation across

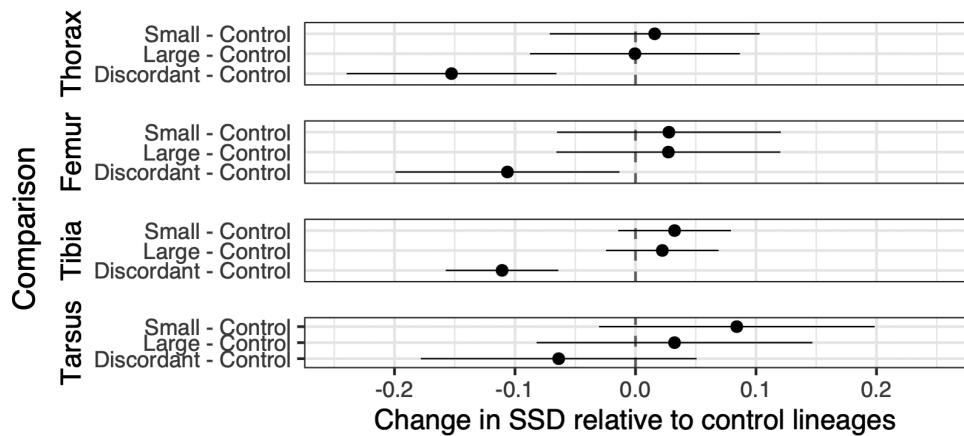


Figure 3. Substantial changes in sexual size dimorphism only occur under sexually discordant selection on size. Contrasts represent a proportional change in sexual size dimorphism for each artificially selected treatment, in comparison to controls, by trait. Modeled using \log_2 transformed length measures (μm), facilitating comparisons of proportional changes. Error bars are 95% CIs for contrasts.

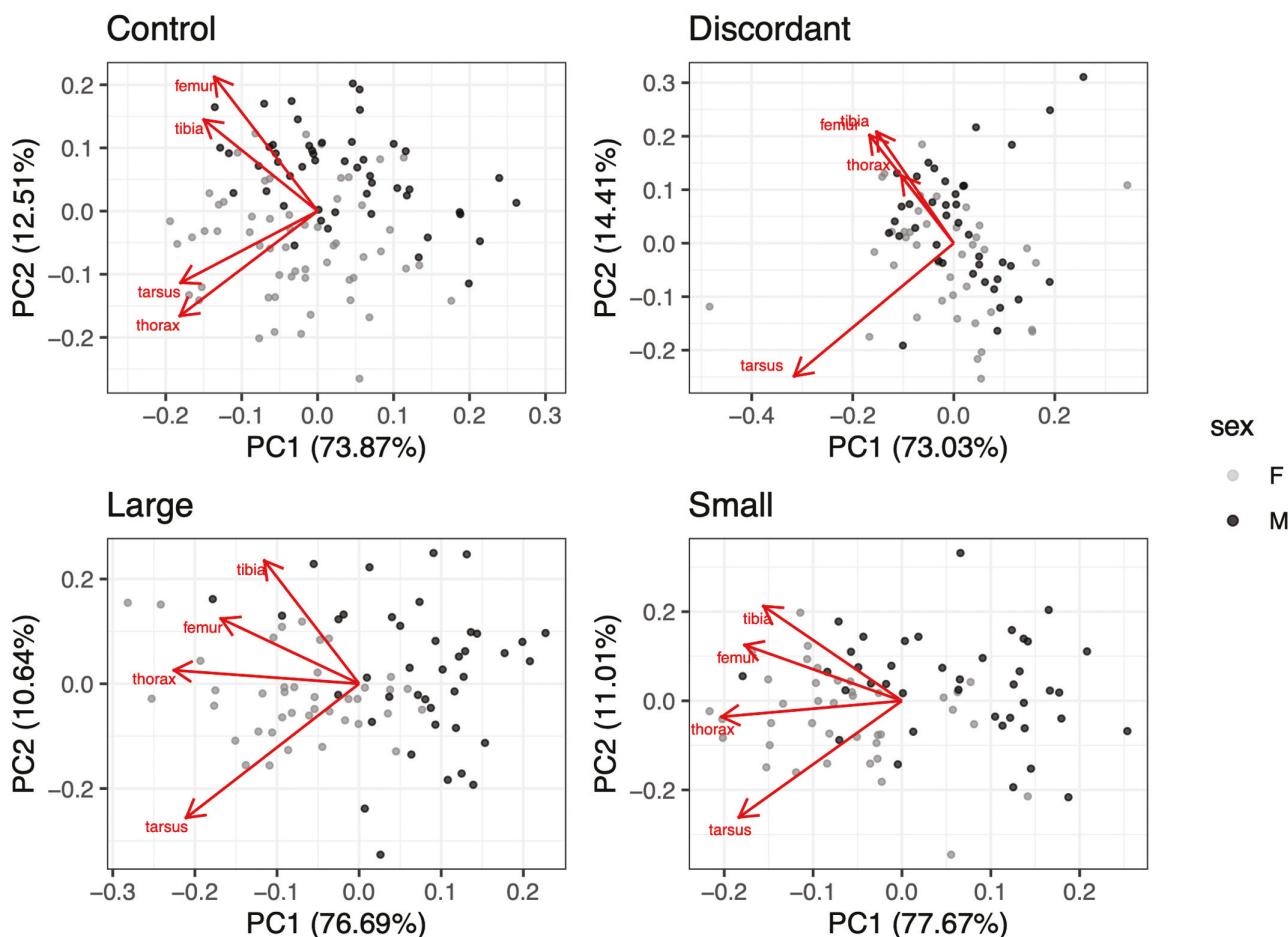


Figure 4. Sexually discordant selection alters patterns of multivariate allometry across sex. Represented as biplots, magnitudes and direction of the loadings for traits are superimposed onto PC1 and PC2. \log_2 transformed length measures were used.

the genome and many regions at or near fixation (Figure 5A). Notably, average F_{ST} is quite high across the chromosomes, consistent with a substantial impact of genetic drift on allele frequencies. When F_{ST} is calculated between discordant and either of the concordant (small or large) treatments, we similarly see elevated F_{ST} and multiple regions at or near fixation (Supplementary Figures S5 and S6). In our discordant selection versus controls, the highest mean F_{ST} is found on chromosome 2L (mean F_{ST} per chromosome, 2L = 0.215, 2R = 0.177, 3L = 0.144, 3R = 0.171, X = 0.205). In the Large versus Small F_{ST} comparison, again, there were many regions showing high differentiation (Figure 5B). The chromosome with the highest mean F_{ST} in the large versus small comparison was the X chromosome (2L = 0.277, 2R = 0.312, 3L = 0.207, 3R = 0.183, X = 0.354). For comparison, we also provide plots of the

CMH statistic (Supplementary Figures S7 and S8). To confirm the effects for SNPs of interest, we modeled allele frequency changes with logistic regression between treatments, examining odds ratio VS. F_{ST} , and verifying high F_{ST} correlated to large odds ratios (Supplementary Figures S9, S10, and S11). After manual curation, several genes with known sex-specific size effects were identified, including *dMyc* (*myc*), *Hairless* (*H*), *Insulin-like receptor* (*InR*), *Regulator of cyclin A1* (*Rca1*), and *stunted* (*sun*). Of our candidate discordant genes, 11/295 (excluding inter-genic SNPs and lncRNA) have both a known size phenotype as well as a sex-limited phenotype. Many other genes with known effects on aspects of the body or trait size were also observed for both the discordant and concordant comparisons (Supplementary Files 1, 2, 3 and 4). For the concordant selective lineages, we examined

Table 1. Loadings for PC1, by treatment.

Trait	Control (0.74)	Small (0.78)	Large (0.77)	Sex-discordant (0.73)
$\log_2(\text{femur})$	0.417	0.490	0.455	0.415
$\log_2(\text{tibia})$	0.460	0.430	0.312	0.380
$\log_2(\text{tarsus})$	0.553	0.508	0.569	0.786
$\log_2(\text{thorax})$	0.556	0.562	0.609	0.258

Note. Loadings for PC1 (eigenvector 1) from variance-covariance matrices for each treatment. The values next to treatment names correspond to the proportion of variation accounted for by PC1. Only the lineages artificially selected discordantly (between the sexes) show substantial changes in multivariate allometry.

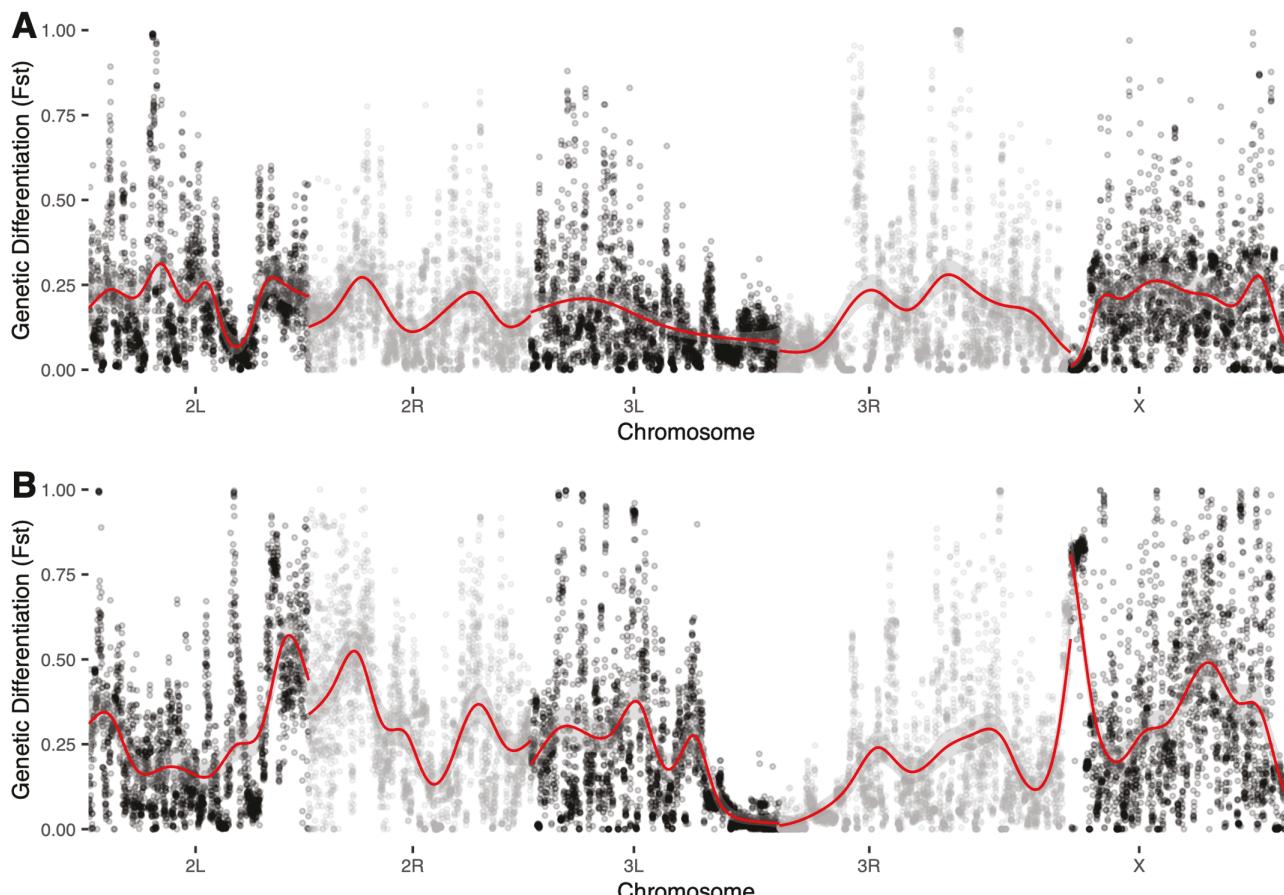


Figure 5. Genomic divergence among artificially selected treatments. Genome-wide F_{ST} (10,000 bp windows). Chromosomal trends for F_{ST} (binomial, gamm) in red. (A) Discordant selection compared to control treatments. (B) Large compared to small treatments.

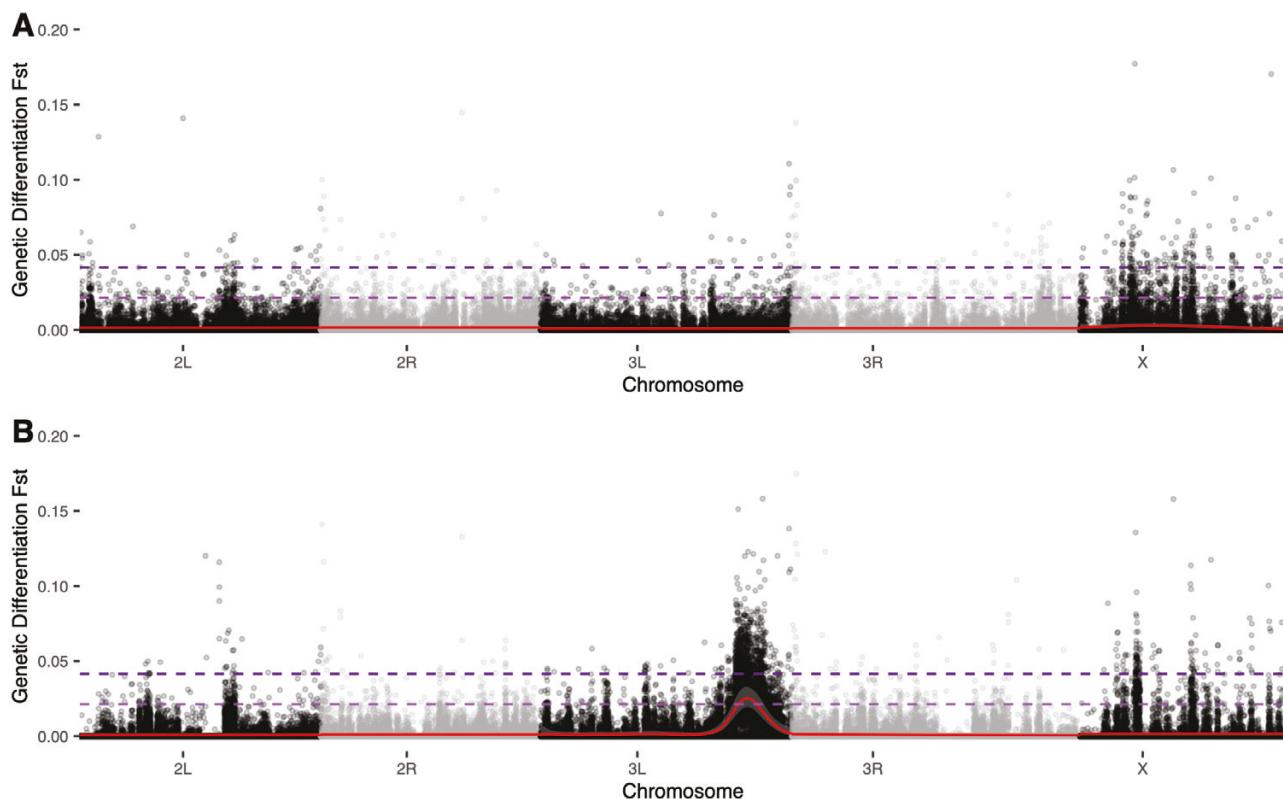


Figure 6. Within-lineage, within-generation, genome-wide F_{ST} (10,000 bp windows) between males and females. Chromosomal trends for F_{ST} (binomial, gamm) in red. Maximum simulated value (neutral evolution, SLiM v4.2) in dark purple, 95th quantile of simulated values in light purple. (A) Male vs. Female for control treatment (replicate lineage 1), showing a trend $\cong 0$, across the genome. (B) Male vs. Female F_{ST} values for sexually discordant treatment (replicate 1) showing a region on chromosome 3L with elevated F_{ST} . Within the region of interest, the maximum simulated value was 0.02, and the 95th quantile of simulated values in that region was 0.001.

genes in our list overlapping with those identified from generation 100 of selection [Turner et al. \(2011\)](#). The only gene that overlapped between the analysis of generation 100 and 378 was *Nop1-like* (*Nop17l*). Modulation of the expression of *Nop17l* in developing wing tissues reduces the size of the wing ([Bennett et al., 2006](#)). If we examine overlapping genes excluding the logistic regression analysis, 22 additional genes or SNPs overlap between F100 and F378 concordant treatments ([Supplementary Files 5 and 6](#)).

The candidate genomic regions identified above were then examined for enrichment of GO terms. Permutated GO enrichment analysis with Gowinda did not return significant terms, except for our large treatment, which only returned the term astral microtubule organization (GO:0030953) as an enriched GO term.

Intersexual genetic differentiation in one sexually discordant lineage may suggest the maintenance of intra-locus sexual conflict

Given the long-term nature of discordant selection in this experiment, these lineages may be useful in identifying signatures of IASC. As autosomes spend equal time in males and females, it is expected intersexual F_{ST} should be close to zero in most circumstances, absent strong sexual conflict ([Cheng & Kirkpatrick, 2016](#); [Kasimatis et al., 2019](#)). For all concordant selection treatments (Control, Large and Small), we found mean chromosomal F_{ST} to be near zero (mean of C1 = 0.0015, C2 = 0.0019, [Figure 6A](#); S12; L1 = 0.0012, L2 = 0.0015, [Supplementary Figures S13 and S14](#);

S14; S1 = 0.0023, S2 = 0.0014, [Supplementary Figures S15 and S16](#)). For the discordant selection treatment replicate 2, we also found a mean chromosomal F_{ST} to be near zero (E2 = 0.0011, [Supplementary Figure S17](#)). For replicate 1 of our discordant selection treatment, however, all chromosomal mean F_{ST} is near zero (E1 = 0.0019), except for an elevated section of ~3.4 Mb on chromosome 3L between position 18,100,000 and 21,600,000, where mean F_{ST} rises to 0.005, with elevated SNPs showing a distinct peak of F_{ST} nearing 0.1, with a couple of windows reaching F_{ST} of 0.25 ([Figure 6B](#)). We did not identify any common inversions on chromosome 3L (*In(3L)P*, 3L133in, 3L165in, 3L096, 3L105, and 3L058), nor novel structural variants using DELLY ([Rausch et al., 2012](#)) contributing to this elevated region of F_{ST} . We identified SNPs in this E1 lineage on chromosome 3L that were 3 standard deviations above the mean between sex F_{ST} ([Supplementary File 1](#)).

We confirmed that this region of elevated intersex F_{ST} in discordant replicate 1 was indeed an outlier using several types of simulations. SLiM simulations, as well as a custom simulation designed to account for various sources of sampling variation ([Supplementary Figures S18–S26](#)), as well as chromosome-wide simulations accounting for aspects of variation in recombination rates and demographic effects of the populations. All male-female comparisons follow a similar pattern to the simulations, with the exception of discordant replicate 1, which appears extreme relative to all simulations ([Supplementary Figure S19](#)). We further modeled all SNPs on chromosome 3L to look for clusters of significant

SNPs within our elevated F_{ST} region ([Supplementary Figures S27–S34](#)). Our discordant lineage replicates 1 shows a high number of significant SNPs within our elevated region when modeled ([Supplementary Figure S29](#)), while no other lineage has a clear cluster of significant SNPs in this region. We also looked for genes that overlap with previously identified conflict genes in the establishing population of LH_M ([Ruzicka et al., 2020](#)) however we found no overlap. [Ruzicka et al. \(2019\)](#) also used LH_M to explore sexual conflict loci, but their LH_M flies have diverged from the LH_M used in this experiment for an unknown number of generations. If we exclude our modeled SNPs and look for overlap in genes that overlap without identified high F_{ST} we find a single named gene, Formin-like (*FrI*), which does not have any known body size or sex-limited phenotypes.

Evolved changes in sex ratio and fecundity may suggest conflict in the discordant lineages

Crossing all treatment groups in single-pair matings reciprocally to the LH_M “ancestor” resulted in ~1:1 sex ratio in the F₁. The exception to this was Control replicate 1 male crossed to LH_M females (C1: F-M = 0.455, CI = 0.426–0.485; [Figure 7](#)), as well as within-lineage crosses in both discordant treatments (E1: F-M 0.550, CI = 0.517–0.582; E2: F-M 0.550, CI = 0.509–0.589; [Figure 7](#)). The deviation in sex ratio in the control replicate 1 is also observed in F₂ (C1: F-M = 0.457, CI = 0.427–0.487; [Figure 8](#)). The sex ratios of the F₂s from crosses within both discordant lineages show male bias, and the confidence intervals for discordant replicate 1 cross do not overlap the 1:1 expectation (E1: F-M = 0.557, CI = 0.521–0.592; [Figure 8](#)). We also observed a deviation from the expected sex ratio in our large replicate 1 male to LH_M female cross (L1: M-F = 0.459, CI = 0.429–0.489; [Figure 8](#)), and our discordant replicate 1 male to LH_M female cross (E1: M-F = 0.453, CI = 0.426–0.480; [Figure 8](#)). This effect in the discordant cross was in the opposite direction of the discordant replicate 1 pure cross.

Next, we explored whether the direction of the cross had an effect on the sex ratio by adding cross-direction as a predictor in the model. In the F₁ generation, the overall direction of the cross had very modest impacts on sex ratios ($\chi^2 = 1.13$, $df = 1$, $p = 0.29$). There was some evidence of deviation from the expected sex ratio when the direction was accounted for

when control 1 sired the cross (C1: M-F = 0.455, CI = 0.426–0.485). In the F₂ generation, the direction of the F₀ cross (sire vs. dam) had a modest impact on sex ratios ($\chi^2 = 5.67$, $df = 1$, $p = 0.017$). Specifically, when the treatment lineage was the sire in the initial cross for discordant replicate 1 and large replicate 1 (E1: M-F = 0.453, CI = 0.426–0.480; L1: M-F = 0.459, CI = 0.429–0.489). The control replicate 1 cross appeared to deviate from the expected sex ratio regardless of which parent served as sire or dam.

While it was not a planned experiment, while counting flies for sex ratio crosses, we observed a possible difference in fecundity between treatments. The preliminary results from the analysis from the F₁, when the dam was from either discordant lineages, showed reduced fecundity ([Supplementary Figure S35](#)). In the F₂ generation, there did not appear to be a direction of cross-effect, but discordant treatments had the lowest fecundity ([Supplementary Figure S36](#)).

Discussion

Sexual dimorphism evolves frequently despite r_{MF} generally being high within species, in particular for morphological traits ([Lande, 1980](#); [Poissant et al., 2010](#)). In the presence of sex-specific optimal phenotypes, this r_{MF} has the potential to generate intra-locus sexual conflict through the “load” on the opposite sex ([Fairbairn et al., 2007](#)). Species that evolve changes in SD must overcome any hurdles due to a high r_{MF} and ensuing genomic conflict. In the long-term, this conflict may reach an equilibrium, with sex-biased alleles creating as near an optimum phenotype for each sex as possible, or be resolved entirely. However, the reintroduction of sex-discordant selection should disrupt this equilibrium and generate additional genomic conflict. Examining the response to sexually discordant selection for body size across the genome after selection provides an opportunity to identify genomic regions undergoing conflict. In this study, we utilize long-term artificially selected lineages ([Stewart & Rice, 2018](#)), selected for body size either in a sexually concordant or discordant manner. As discussed in detail below, in addition to observing trait-specific changes in patterns of SSD, we see potential evidence for the maintenance of polymorphisms consistent with unresolved conflict.

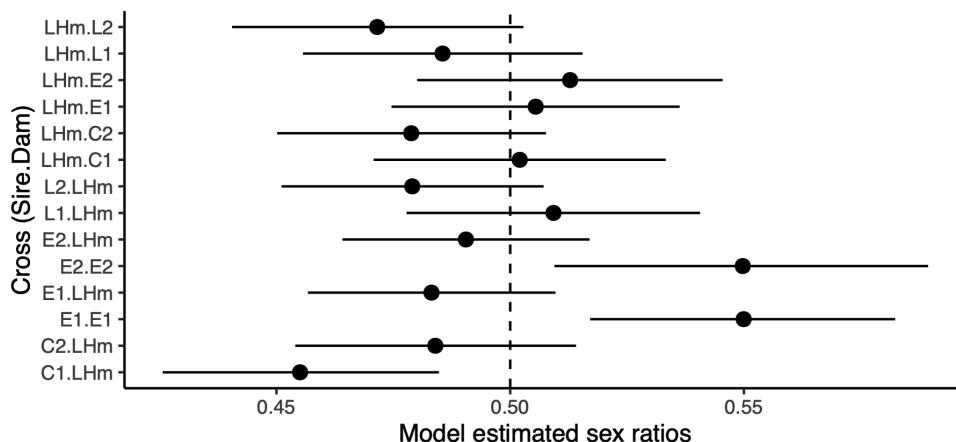


Figure 7. F₁ offspring sex ratios from all treatments crossed to the founder population as well as both discordant lineages crossed “pure.” The cross label has Sire on the left and Dam on the right of the cross identifier. Dashed line marks expected 1:1 sex ratio. LHM = LHm population. L = large selection; C = control; E = discordant selection. Numbers following population labels are replicate lineages.

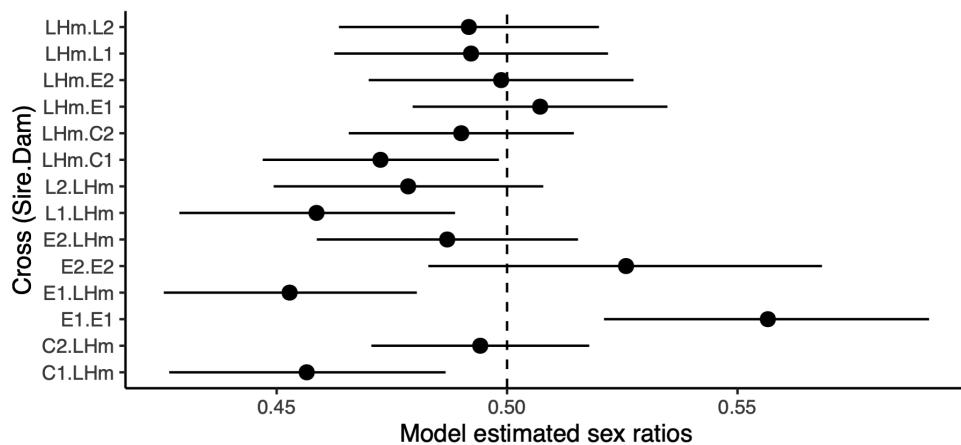


Figure 8. F_2 offspring sex ratios from all treatments crossed to the founder population as well as both discordant lineages crossed “pure.” The cross label has Sire on the left and Dam on the right of the cross identifier. The dashed line marks the expected 50/50 sex ratio. Label codes as in Figure 7.

Concordant selection lineages maintained ancestral patterns of SSD despite selection on size, while discordant selection lineages responded in a trait-specific manner

The lineages under concordant artificial selection for size responded in the expected directions for all traits (Figure 2) and changed size proportionally, maintaining SD for all traits (Figure 3). This proportional response is consistent with relatively high r_{MF} for morphological traits, as previously observed (Cowley & Atchley, 1988; Reeve & Fairbairn, 1996). If r_{MF} had been low in the starting population, we might have observed variation in responses between the sexes, with one sex responding to selection faster than the other and the overall SSD changing from the “baseline,” as has been shown when r_{MF} is intentionally degraded (Delph et al., 2011). Within concordant treatments, SSD was largely maintained despite substantial changes in size. For the discordant selection treatment, SD changed trait specifically (Figures 2 and 3). Of the four traits measured for this study, one (thorax length) showed a substantial degree of female-biased dimorphism in control lineages (difference of ~ 0.16 , in $\log_2 \mu\text{m}$, or $\bar{F}/\bar{M} \approx 1.12$), while leg traits (femur, tibia, and basitarsus lengths) were closer to sexually monomorphic (differences of 0.029, 0.023, 0.074, or ratios of 1.02, 1.02, 1.05 respectively). Under discordant artificial selection, all traits saw a reduction in female-biased dimorphism, with the magnitude of change varying by trait (Figure 3). The dimorphism for thorax length reduced to the greatest degree (difference of ~ 0.012 , in $\log_2 \mu\text{m}$, or $\bar{F}/\bar{M} \approx 1.01$), compared to dimorphism for leg lengths (-0.076 , -0.087 , 0.011 , or ratios of 0.95, 0.94, 1.01). This is perhaps most clearly illustrated in the changes in patterns of multivariate allometry observed in the sexually discordant lineages (Figure 4). Traits with higher levels of SD tend to have lower r_{MF} than nondimorphic traits (Poissant et al., 2010), which may partially account for this. However, this may also reflect the manner in which artificial selection was applied (via a sieve), likely results in a complex composite “size measure” for selection on cross-sectional area, of which a trait like thorax length may contribute substantially more than any individual leg measure. Alternatively, the genetic correlations between the sexes may differ across traits. This would explain why traits like femur and tibia reversed SD when selected indirectly; however, tarsus and thorax (thorax being the most directly selected) did not fully reverse dimorphism.

Within-lineage, across-sex comparisons provide possible evidence of intra-locus sexual conflict in the genome of one discordant lineage

To identify regions under possible intra-locus sexual conflict, we compared male-female F_{ST} within each lineage and conducted several follow-up analyses to identify possible SNPs showing subtle, intersexual, within-generation distortions in allele frequency. There may be small genomic regions or possibly single SNPs showing signs of different allele frequencies between the sexes, which may be evidence of unresolved conflict in the genome. Neither control lineages nor the sexually concordant lineages showed any evidence of elevated F_{ST} (Figure 6A; Supplementary Figures S12–S17). However, one discordantly selected lineage (E1; Figure 6B) has a region on chromosome 3L with F_{ST} reaching as high as >0.2 . Other comparisons of this nature have yielded smaller male-female F_{ST} estimates, which led us to evaluate whether it was an artifact of our pipeline (Cheng & Kirkpatrick, 2016; Dutoit et al., 2018; Flanagan & Jones, 2017; Lucotte et al., 2016; however see Sylvestre et al., 2023) or the demographics of the populations. Modifying SNP calling methods, F_{ST} estimation procedures, and filtering methods, all yielded the same spike in F_{ST} in the E1 lineage. This elevated region was extreme for this lineage in comparison to simulations accounting for a number of sources of sampling variation and overall allele frequencies as well as chromosome-wide simulations of the population. If it was a spurious peak of elevated F_{ST} as a result of a technical aspect of the pipeline or mis-mapping of reads from a sex chromosome, other replicate lineages would likely also show elevated regions. We consider the most likely explanation to be due to the sustained nature of sexually discordant selection in our experiment. Compared with most other studies examining M-F differences in allele frequencies to identify potential loci under conflict, the selection applied to the discordant lineages in this experiment is both strong and sustained across hundreds of generations. Kasimatis et al. (2019) suggest that for a locus to show this degree of asymmetry, an exceedingly high magnitude of selection is required. In their model, this would require nearly 40% mortality each generation to maintain the distortion in M-F allele frequencies at a locus. The nature of artificial selection as applied in the current experiments means well over 40% (~80% of individuals with the current selection design) of flies in each generation (those not at phenotypic extremes) are discarded, essentially

“dead” in the selective sense. Alternatively, Kasimatis et al. also simulated subsamples of 50 males and 50 females with antagonistic selection and found that sampling substantially increased variability in F_{ST} values due to the fact that genomes under antagonistic selection tend to have increased numbers of intermediate-frequency alleles. However, our simulations suggest this is not a likely explanation for the observed results ([Supplementary Figures S18–S26](#)). Why this region in discordant treatment replicate E1 shows this distortion and the contribution of intra-locus conflict requires further exploration.

Given the strange peak in this one discordant replicate, we extracted all SNPs three standard deviations above mean F_{ST} on chromosome 3L, which also were significant in the logistic regression (which better accounts for sampling variance). We filtered for SNPs that appear in both discordant replicates and do not occur in any concordant lineage to look for possible replicated mechanisms of divergence in phenotype. We found fewer candidate genes than in any other comparison ([Supplementary File 1](#)), but multiple genes were present in the WNT pathway, including *frizzled2* (*fz2*). We also found candidate SNPs in the MAPK pathway. Interestingly, none of the male-female discordant treatments SNPs showed evidence for genetic differentiation between artificial selection treatments. Many candidate genes have mutational phenotypes listed on flybase as “abnormal size,” suggestive that the response to selection is highly polygenic with alleles of small effect “inch-ing” toward a phenotype rather than “sprinting.”

Finally, we explored genes found by [Ruzicka et al. \(2019\)](#), which were identified as potentially being in sexual conflict. These genes were considered in conflict based on sex-specific fitness (competitive mating success in males and competitive fecundity in females). An important caveat in our comparison is that although our lineages are from the same initial lab population (LH_M), our treatment groups have gone through nearly 400 generations of strong selection from the ancestral population. It is also unknown how many generations of divergence occurred in the ancestral LH_M populations prior to both experiments. In our comparison between genes identified in [Ruzicka et al. \(2019\)](#) and those identified in our study, only the gene Formin-like (*frl*) overlaps. *Frl* is upstream of our region of interest and does not have any previously identified body size or sex-specific phenotypes. We may not see overlap because [Ruzicka et al. \(2019\)](#) identified conflict alleles based on specific proxies of fitness, whereas ours may reflect more indirect effects due to discordant selection on size. However, we do observe some evidence that adult sex ratios are male-skewed in the discordant lineages and possibly maintained even in crosses to the LH_M “ancestor,” as well as potential reductions in fecundity ([Figures 7 and 8](#); [Supplementary Figures S35 and S36](#)).

Discordant size selection likely has a polygenic basis, involving a number of genes that influence body size in a sex-specific manner

Given the number of generations of artificial selection and the modest population sizes of each generation, genetic drift will have an overwhelming impact on any genome scan performed with the evolved lineages. As such, any interpretation of these genome scans should be tempered by this impact. With that being said, the response is broadly consistent with a polygenic basis, with thousands of SNPs showing evidence of genetic differentiation across the genome. In 378 generations of selection, and more than 125 generations after clear evidence of phenotypic reversal of SSD, the opportunity for a

mutation (or ancestrally segregating relatively rare variants) of large phenotypic effect occurring was possible. We find no clear evidence for alleles of large effect that have swept to fixation in parallel across both lineages of discordant selection. The chance of identifying alleles of large effect is attenuated by the impact of drift, making many SNPs fixed randomly in each lineage. Possibly constrained by genetic correlation, the response to discordant selection appears to be a slow and steady crawl towards a response rather than large leaps forward. The lineage-specific increase in discordant replicate 1 that does not appear in replicate 2 could imply a novel mutation that is under conflict in this replicate. This elevated F_{ST} is shown in a large region of chromosome 3L that includes hundreds of SNPs; however, identifying a novel mutation that is responsible requires further investigation. This slow crawl toward a response is supported by the fact that [Stewart and Rice \(2018\)](#) saw a negligible phenotypic response to discordant selection during at least the first 100 generations, with more substantial responses occurring after this. Large-effect alleles with sex-specific impacts on size have been identified in functional studies; however, the viability cost in the opposite sex may make these large-effect alleles prohibitive ([Millington et al., 2021a; Rideout et al., 2015](#)). Although it is tempting to speculate that these small effect alleles are a likely way for SSD to evolve in natural populations, we must also caveat that it may be the case that these alleles of such small effect may be weeded out by viability or fecundity selection in the wild, which are much weaker in this tightly controlled artificial selection experiment. In fact, [Testa and Dworkin \(2016\)](#) demonstrated that mutations with a sex-specific response are rare in two distinct genetic backgrounds. [Carreira et al. \(2009\)](#), however, found that many genes associated with size change reduced SD when a P-element was inserted. This highlights the complexity of the genetic architecture of SSD. Although [Testa and Dworkin \(2016\)](#) found that mutations with sex-specific effects on size were rare, they did identify mutations in the EGFR pathway with sex-specific effects. In our sex-discordant lineages, we found a number of genes with known sex-specific size effects (*dMyc*, *H*, *InR*, *Rca1*, *sun*). Follow-up on these genes to determine their role in response to discordant selection is required.

We also compared our genes of interest to the previously published gene list from [Ruzicka et al. \(2019\)](#) mentioned above. In our comparison, we still find modest overlap (11 named genes and 8 unnamed genes). Two genes, in particular, stand out; *discs large 1* (*dlg1*), which has a known increased size phenotype as well as courtship and oogenesis phenotypes, and *smrtr* (*smr*), which has known decreased size phenotypes as well as a sex-limited reduced fecundity phenotype in females. Although the caveats of this comparison mentioned above still stand, it is interesting that we find genes under potential conflict coming up in these two studies, both using LH_M . Optimistically, this could suggest genes under conflict that are stable in the population and are being captured in these studies that explore sexual conflict, but more work must be done to confirm this. Both of these interesting genes are also located on the X chromosome, which has been suggested to be enriched in conflict alleles ([Rice, 1984](#)).

Concordant selection lineages show a response in growth pathways but do not align with a previous analysis

The concordant selection lineages (small and large) showed SNPs of interest in multiple growth-related pathways.

Interestingly, we found no clear overlap with analyses done nearly 300 generations prior by Turner et al. (2011). This result could be due to a few possibilities: (1) inconsistencies with bioinformatics tools and pipeline choices, (2) improvements in sequencing and software refining our more recent search, and (3) potentially most biologically interesting; this could be due to multiple “soft” selective sweeps generating allelic turn-over phenomena in the genes under active selection for size. To circumvent the first possibility, we reanalyzed the F100 data used by Turner et al. with our pipeline, adjusted for coverage and quality. Importantly (and not surprisingly, given when the experiment was done), the sequence coverage per sample is very modest (coverage ~25x) in comparison to our current study. Using our pipeline, the F100 data had no genes with known size phenotypes overlapping with our large genes of interest (Supplementary File 5). In the gene list overlapping with our small treatment genes of interest, we retrieved Mnt (*mnt*), which has a known increased body size phenotype; we also identified saxophone (*sax*), which has an abnormal size phenotype, and potentially interestingly, we recovered Tousled-like kinase (*tlk*), which has a known decreased body size phenotype (Supplementary File 6). The fact that we recover overlapping size-related genes in the small treatment but not the large treatment is likely due to the stronger selection in this treatment.

Although our candidate genes differ from those listed by Turner et al., both analyses found genes in related general pathways such as ecdysone signaling and the EGFR pathway. This could be evidence that the standing genetic variation present in the starting population (LH_M) could be fueling the clearly polygenic response, and new mutations or stochastic changes in allele frequency are continuously altering which genes are used to respond to the selection over such a large time frame. The pathways both we and Turner et al. identify have been previously implicated as being under selection in naturally occurring clines in *D. melanogaster* and, therefore, may provide interesting insight into natural variation in body size. We identified a substantial number of genes after very conservative filtering (126 in Large and 101 in Small) that appeared to be under selection in our concordant selection lineages. A number of these genes have known size phenotypes; however, since such a large number of genes in the genome impact body size, it is difficult to say if these are directly responsible for response to selection or are mere coincidence (Carreira et al., 2009; Turner et al., 2011).

Our results are both an important exploration of genomic conflict using artificial selection and a follow-up on one of the longest (in generations) artificial selection experiments in animals. Although with such strong and persistent selection, our results may not mirror natural settings, it does demonstrate one possibility of how the genome may respond to discordant selection. We also manage to demonstrate a likely differentially segregating region of the genome in a discordantly selected lineage, possibly a distinct autosomal region of sexual conflict in the genome. Although these works require further experimentation to narrow down and validate alleles of interest, we suggest that we have demonstrated at least one route for the genome to respond to discordant selection and sexual conflict.

Supplementary material

Supplementary material is available online at *Evolution*.

Data availability

All phenotypic data and scripts are available on GitHub (https://github.com/DworkinLab/Audet_etal_Evolution_2024) along with a static copy on DRYAD (<https://doi.org/10.5061/dryad.6t1g1jx6k>). All raw genomic sequence data is deposited in the NCBI SRA under BioProject PRJNA1107500.

Author contributions

Study conceptualization and funding: I.D.; Study design: I.D.; Artificial selection and rearing: A.D.S.; Dissections: J.K.; DNA extraction: T.A. and K.P.; Image analysis: J.K. and I.D.; Analysis: I.D., T.A., and K.P.; Manuscript drafting: T.A. and I.D.; Manuscript editing: T.A., A.D.S., and I.D.; Manuscript revisions: T.A. and I.D.

Funding

This work was funded by an NSERC Discovery grant to I.D. and an Ontario Graduate Scholarship to T.A.

Conflict of interest: Editorial processing of the manuscript was done independently of I.D., who is an Associate Editor of *Evolution*. The other authors declare no conflict of interest.

Acknowledgments

We thank the associate editor Dr. Charles Baer, the handling editor Dr. Tim Connallon, and two anonymous reviewers for feedback that has improved this manuscript.

References

- Alicchio, R., & Palenzona, D. L. (1971). Changes of sexual dimorphism values in *Drosophila Melanogaster*. *Italian Journal of Zoology*, 38, 75–84.
- Arguello, J. R., Laurent, S., & Clark, A. G. (2019). Demographic history of the human commensal *Drosophila melanogaster*. *Genome Biology and Evolution*, 11(3), 844–854. <https://doi.org/10.1093/gbe/evz022>
- Ashburner, M. (1989). *Drosophila. A laboratory handbook*. Cold Spring Harbor Laboratory Press.
- Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bennett, D., Lyulcheva, E., Alphey, L., & Hawcroft, G. (2006). Towards a comprehensive analysis of the protein phosphatase 1 Interactome in *Drosophila*. *Journal of Molecular Biology*, 364(2), 196–212. <https://doi.org/10.1016/j.jmb.2006.08.094>
- Bird, M. A., & Schaffer, H. E. (1972). A study of the genetic basis of the sexual dimorphism for wing length in *Drosophila melanogaster*. *Genetics*, 72(3), 475–487. <https://doi.org/10.1093/genetics/72.3.475>
- Blackith, R. E., & Reymont, R. A. (1971). *Multivariate morphometrics*. Academic Press.
- Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9, 378.
- Bushnell, B. (2021). *BBTools: BBMap short read aligner, and other bioinformatic tools*. SourceForge.

