Complexities of recapitulating polygenic effects in natural populations: replication of genetic effects on wing shape in artificially selected and wild-caught populations of *Drosophila melanogaster*

Katie Pelletier,[1,†] William R. Pitchers,[2,3,†] Anna Mammel,[2,4] Emmalee Northrop-Albrecht,[2,5] Eladio J. Márquez,[6,7]
Rosa A. Moscarella,[6,8] David Houle,[6,*] Ian Dworkin[1,2,*]

[1]Department of Biology, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4L8, Canada
[2]Department of Integrative Biology, Michigan State University, East Lansing, MI 48824, USA
[3]Present address: BiomeBank, 2 Ann Nelson Dr, Thebarton, Adelaide, SA 5031, Australia
[4]Present address: Neurocode USA, 3548 Meridian St, Bellingham, WA 98225, USA
[5]Present address Division of Gastroenterology and Hepatology, Mayo Clinic, 200 First St. SW, Rochester, MN 55905 USA
[6]Department of Biological Science, Florida State University, 319 Stadium Drive, Tallahassee, FL 32306-4295, USA
[7]Present address: Branch Biosciences, 1 Marina Park Dr., Boston, MA 02210, USA
[8]Present address: Department of Biology, University of Massachusetts, 221 Morrill Science Center III, 611 North Pleasant Street, Amherst, MA 01003-9297, USA

*Corresponding author: Email: dhoule@bio.fsu.edu; *Corresponding author: Email: dworkin@mcmaster.ca
[†]Co-first authors.

Abstract

Identifying the genetic architecture of complex traits is important to many geneticists, including those interested in human disease, plant and animal breeding, and evolutionary genetics. Advances in sequencing technology and statistical methods for genome-wide association studies have allowed for the identification of more variants with smaller effect sizes, however, many of these identified polymorphisms fail to be replicated in subsequent studies. In addition to sampling variation, this failure to replicate reflects the complexities introduced by factors including environmental variation, genetic background, and differences in allele frequencies among populations. Using *Drosophila melanogaster* wing shape, we ask if we can replicate allelic effects of polymorphisms first identified in a genome-wide association studies in three genes: *dachsous*, *extra-macrochaete*, and *neuralized*, using artificial selection in the lab, and bulk segregant mapping in natural populations. We demonstrate that multivariate wing shape changes associated with these genes are aligned with major axes of phenotypic and genetic variation in natural populations. Following seven generations of artificial selection along the *dachsous* shape change vector, we observe genetic differentiation of variants in *dachsous* and genomic regions containing other genes in the hippo signaling pathway. This suggests a shared direction of effects within a developmental network. We also performed artificial selection with the *extra-macrochaete* shape change vector, which is not a part of the hippo signaling network, but showed a largely shared direction of effects. The response to selection along the *emc* vector was similar to that of *dachsous*, suggesting that the available genetic diversity of a population, summarized by the genetic (co)variance matrix (**G**), influenced alleles captured by selection. Despite the success with artificial selection, bulk segregant analysis using natural populations did not detect these same variants, likely due to the contribution of environmental variation and low minor allele frequencies, coupled with small effect sizes of the contributing variants.

Keywords: *Drosophila*, wing shape, morphometrics, quantitative genomics, GWAS, artificial selection, **G** matrix, hippo signaling, complex traits

Introduction

Dissecting the genetic architecture underlying complex traits remains challenging, because of the joint contributions of many alleles of small effect, genotype-by-environment interactions, and other factors. Progress in sequencing technology in conjunction with development of genome-wide association studies (GWAS) statistical methodologies has enabled identification of loci contributing to numerous complex traits and diseases. However, such mapping approaches identify only a subset of loci contributing to trait variation (Visscher *et al.* 2017). In part, this reflects the low power to detect rare alleles, and those with small effects (Tam *et al.* 2019). For alleles that are relatively

common in a population, replication rates between GWAS studies are high, even when effect sizes are small (Marigorta *et al.* 2018). However, GWAS studies have failed to replicate the effects observed in many candidate gene studies, in part due to the fact that many alleles identified in these studies are rare in populations, and require very large cohorts to detect (Ioannidis *et al.* 2011; Fritsche *et al.* 2016).

In cases where an association is replicated between studies, the magnitude of the effect can vary substantially between different cohorts or populations (CONVERGE consortium 2015; Marigorta *et al.* 2018). Differences can arise because of genetic background due to epistatic gene-by-gene (GxG) interactions, or due to gene-by-environment (GxE) interactions. The initial estimates of

effect size will be biased upwards if statistical testing in the initial cohort is used to determine which SNPs are chosen for replication studies. It is important to understand which of these causes of differences in effect size are of practical significance when we want to generalize results to different populations or environments.

In this study, we focus on the issue of replication in a multivariate context, where the joint inheritance of multiple features are simultaneously investigated. We will refer to the suite of measured features as a "multivariate trait" for convenience. In this case, what we want to estimate is the vector of effects of each SNP on all measured features. Each SNP may have a unique combination of effects. Univariate effects vary only in magnitude, as we can only infer effects on a single feature. For a multivariate trait, estimated genetic effects vary in magnitude, the sum of effects on all traits, and also in direction, how the total effect is allocated among different features (Melo *et al.* 2019). The ability to study the direction along with the magnitude of genetic effects provides an additional and important way of assessing repeatability. For a univariate trait, there is a 50% chance that the replicate estimate will be in the same direction as the original estimate, even with no true effect. By contrast, the probability of a "replicated" genetic effect sharing a similar direction by chance alone decreases as the number of measured features increases (Stephens 2013; Marquez and Houle 2015).

Studying genetic effects in a multivariate context is beneficial in other ways. First, it has been demonstrated both empirically and via simulations, that genetic mapping for multivariate traits generally increases statistical power over trait by trait analyses (Shriner 2012; Porter and O'Reilly 2017; Fatumo *et al.* 2019; Pitchers *et al.* 2019). Second, some multivariate traits cannot be sensibly reduced to a single measurement. The wing shape we study is a great example of such a multivariate trait. We have good reason to believe that wing shape is important for flight (Ray *et al.* 2016), but we cannot yet say that any feature, such as wing length or width, is more or less important than any other. Natural selection on wing shape may affect any or all combinations of measurements.

Perhaps most importantly, traits are not inherited in isolation, but are the joint outcome of an integrated developmental process that results in extensive genetic correlations that can have important effects on evolution. The main source of such correlations are the patterns of pleiotropic effects generated by mutational effects. Multivariate studies of inheritance allow pleiotropic effects to be estimated in a rigorous and justifiable manner (Melo *et al.* 2019). The multivariate breeder's equation, $\Delta \mathbf{z} = \mathbf{G}\boldsymbol{\beta}$, enables short-term prediction of evolutionary responses. Key to understanding how populations respond to selection in the short term requires an understanding of properties of the genetic (co)variance matrix (**G**), and in particular the axis of greatest genetic variation, $\mathbf{g}_{max}$. Studies demonstrate that the direction of $\mathbf{g}_{max}$ influences evolutionary trajectories (Schluter 1996; McGuigan 2006; Blows and McGuigan 2015). The degree to which genetic effects associated with particular variants align to major axes of genetic (co)variance, expressed through **G**, may provide insights into which alleles are most likely to be "captured" by selection (Pitchers *et al.* 2019).

Due to the polygenic nature of complex traits, including multivariate ones, it is important to consider not only the direction of effect for alleles in a single gene but also correlated effects between genes contributing to the phenotype. Interestingly, initial comparisons of directions of genetic effects among induced mutations in two *Drosophila melanogaster* wing development pathways showed only partially correlated effects on wing shape within

and between pathways (Dworkin and Gibson 2006). However, recent work has demonstrated that despite large differences in magnitude, the direction of genetic effects of variants segregating in populations are sometimes similar to those from validation experiments using RNAi knockdown of those same genes (Pitchers *et al.* 2019). Additionally, Pitchers *et al.* (2019) demonstrated this shared direction of effect could also be shared between a SNP and RNAi knockdown of other genes in the same signaling pathway, such as those involved with hippo signaling, a key pathway involved with wing growth and morphogenesis (Pan *et al.* 2018).

Pitchers *et al.* (2019) identified over 500 polymorphisms contributing to wing shape variation in the *Drosophila* genetic resource panel (DGRP). Among these, the hippo pathway was overrepresented in SNPs associated with wing shape (Pitchers *et al.* 2019). The degree to which identified hippo pathway variants reflect allele specific effects, differences in magnitude of genetic effects, and even the large statistical uncertainty associated with genetic effects of small magnitude are unclear. Given common dominance patterns, and the likely nonlinear genotype–phenotype relationships of most genetic effects, small to moderate changes in gene function may result in modest phenotypic effects (Wright 1934; Green *et al.* 2017; Melo *et al.* 2019). Large effect mutants and many RNAi knockdown studies have moderate to large phenotypic effects that are not reflective of the magnitude of genetic effects of SNPs contributing to phenotypic variance in natural populations.

The expression of genetic effects also depends on genetic and environmental context, with GxG and GxE interactions contributing to phenotypic variation. The context-dependence of genetic effects for a multivariate trait has been demonstrated for *Drosophila* wing shape. Variants in *Epidermal growth factor receptor* (*Egfr*), influencing *Drosophila* wing shape are replicable in both lab reared, and wild-caught cohorts (Palsson and Gibson 2004; Dworkin *et al.* 2005; Palsson *et al.* 2005). However, in replication studies, effect sizes of alleles were diminished in both outbred populations and wild cohorts. In the latter case the same variant explained one-tenth of the phenotypic variance explained in the initial study (Dworkin *et al.* 2005). Interestingly, in a series of experimental crosses among strains, the effects of the SNP were replicable for direction and magnitude in multiple experimental assays and crossing schemes. Despite this, the genetic effect on wing shape from this SNP largely disappeared in one natural population (Palsson *et al.* 2005). A number of reasons have been proposed for the failure to replicate genetic effects including environmental effects, differences between controlled lab and natural environments (Dworkin *et al.* 2005), and genetic background (Greene *et al.* 2009), among others. Because both environment and genetic background likely affect the genotype–phenotype map in a nonlinear fashion (Wright 1934), it is important to test observed associations in other experimental contexts.

A promising approach to confirm the estimated effects of candidate genetic variants is to test whether they respond to artificial selection in the direction of the inferred effect. This approach is particularly relevant to evolutionary questions, but has rarely been used. In this study, we use artificial selection and bulk segregant analysis (BSA), to replicate and validate SNPs associated with three genes, previously identified in a GWAS of *Drosophila* wing shape (Pitchers *et al.* 2019); *dachsous* (*ds*), an atypical cadherin involved with hippo signaling; the transcriptional co-repressor *extra-macrochetae* (*emc*), and the E3 ubiquitin ligase *neuralized* (*neur*), involved with Notch signaling. Using the vectors of shape change based on RNAi knockdowns of each gene, we demonstrate that the direction of shape change for these genetic effects is

aligned with major axes of natural phenotypic and genetic variation. Using artificial selection based on the direction of shape change defined by RNAi knockdown, we were able to replicate the effects observed for *ds*, but not *emc*, likely due to the available genetic diversity in the population. We then asked if these effects could be replicated in a natural population using a bulk segregant approach, observing little evidence for replication in these samples. We discuss our results in the context of the replicability of genetic effects and the shared direction of genetic effects due to shared developmental processes.

Methods
Source populations and phenotypic analysis
Drosophila *strains*
Phenotype data for the DGRP was collected for 184 strains as part of a GWAS study as described in Pitchers *et al* (2019). Genotype data for these strains was obtained from freeze 2 of the DGRP (Huang *et al.* 2014). For replication using artificial selection, 27 DGRP strains were used: DGRP-149, 324, 383, 486, 563, 714, 761, 787, 796, 801, 819, 821, 822, 832, 843, 849, 850, 853, 859, 861, 879, 887, 897, 900, 907, 911, 913. These strains were selected to increase genetic variation at the *ds* locus (Supplementary Fig. 1; Table 1). Reciprocal pairwise crosses between the selected DGRP strains were used to create heterozygotes and these heterozygous genotypes were successively pooled for four subsequent generations, allowing for recombination. After pooling, the synthetic outbred population was maintained for approximately 47 subsequent generations (allowing for recombination) before the start of artificial selection experiments.

For the replication in wild-caught populations using BSA, individuals were collected via sweep-netting from orchards and vineyards in Michigan and after species identification, stored in 70% ethanol. In 2013 and 2014, cohorts were collected from Fenn Valley Winery (FVW13 and FVW14, respectively, GPS coordinates: 42.578919, −86.144936). Additionally in 2014, cohorts were collected from Country Mill Orchard (CMO, GPS coordinates: 42.635270, −84.796706), and Phillip's Hill Orchard (PHO, GPS coordinates: 43.117981, −84.624235). For all collected cohorts, except for the FVW14 collection, only males were used in this study given difficulties distinguishing *Drosophila melanogaster* and *D. simulans* females morphologically. For the genomic analysis of the FVW14 wild-caught population (below) we utilized both males and females as the number of individuals was insufficient otherwise. For the collection where females were included in the study, there is no evidence of contamination with *D. simulans* as all dissected wings were classified as *D. melanogaster* using linear discriminant analysis (LDA). LDA was trained using male wings from the collected *D. melanogaster* data set and males from *D. simulans*. There was 100% agreement between the classification of females within each species with our phenotypic classification, indicating that it is unlikely that *D. simulans* females were included in our samples (Supplementary Fig. 2).

Morphometric data
Landmark and semi-landmark data were captured from black and white TIFF images using the pipeline described in Houle *et al.* (2003). First, two landmark locations, the humeral break and alula notch, were digitized using tpsDig2 (version 2.16). Wings (Van der Linde 2004–2014, v3.72) software was used to fit nine cubic B-splines, and manually correct errors. All shape data was subjected to Procrustes superimposition (registration), removing the

effects of location, isometric scaling, and minimizing effects of rotation, via an iterative least squares approach (Rohlf and Slice 1990). Generalized Procrustes superimposition (registration) and extraction of 14 landmarks and 34 semi-landmarks was done using CPR v1.11 (Márquez 2012–2014, Fig. 1). Sliding of semi-landmarks utilized minimization of Procrustes distance as the objective function. Superimposition results in the loss of four possible dimensions of variation while semi-landmarks are constrained to vary along one "axis", restraining these points to approximately a single dimension of variation each. This results in a total of ~58 available dimensions of shape variation, that can be summarized using the first 58 principal components (PCs). Allometry was adjusted for in the analysis by fitting a model for landmark coordinates onto centroid size, and using the residuals from this model (Klingenberg 2022). By accounting for the allometric component of shape, shape variation associated with size variation can be accounted for (Supplementary Fig. 3). For most analyses, "allometry corrected" shape data were used, with the exception of shape models fit using the Geomorph package in R, where Procrustes landmarks were used and centroid size was included as a predictor in the model.

Generation of shape vectors for artificial selection and BSA
A panel of shape change vectors was estimated using the progesterone-inducible Geneswitch GAL4, under the regulation of an ubiquitous *tubulin* driver, to drive the expression of RNAi for genes of interest (*ds*, *emc*, *neur*), as previously described in Pitchers *et al.* (2019). GAL4 expression was induced throughout larval development by adding mifepristone, an analog of progesterone, to the larval food. Knockdown was varied by assaying phenotypes at mifepristone concentrations of 0.3, 0.9, and 2.7 μM, plus a control without mifepristone. Wing shape change associated with knockdown of the gene of interest was estimated using multivariate regression of shape on concentration of mifepristone. Shape change vectors estimated from the RNAi experiments for *ds*, *emc*, and *neur*, were used in this experiment (Fig. 1b; Supplementary Fig. 4). The magnitude ("length") of the vector measures how much shape change occurs per unit change in mifepristone. In general, vectors of greater magnitudes enable better estimate of direction of effect for shape change. As reported in Pitchers *et al.* (2019), the magnitude (l^2-norm) of vectors for RNAi knockdown of these genes are 5.5 for *ds*, 2.8 for *neur*, and 0.44 for *emc*.

Shape data collected as part of a previous study (Pitchers *et al.* 2019) was used to assess the relationships between shape change vectors from the RNAi titration and $\mathbf{g_{max}}$, the first eigenvector of the **G** matrix estimated from DGRP line means. The effects of sex, centroid size and their interaction were removed using a linear model and these residuals were used to calculate shape score by projecting the data (see Supplementary Fig. 5) onto the shape change vector estimated in each knockdown experiment. To assess major axes of genetic variation among DGRP strains, principal component analysis (PCA) was performed on allometry adjusted model residuals (Supplementary Fig. 5b). PCA was done in a similar manner for individuals from the wild-caught cohorts. Correlations between the first three eigenvectors ("genetic PCs" including $\mathbf{g_{max}}$), the first three PCs from the wild-caught cohorts and the shape scores for *ds*, *emc*, and *neur* were calculated (Fig. 1a; Supplementary Fig. 5). From this, *ds*, *emc*, and *neur* shape change vectors were selected for further experiments given high correlation with directions of natural genetic variation (Fig. 1; Supplementary Fig. 5). Note, as described below, while *ds* and *emc* were used for artificial selection, due to the similar response

Table 1. Variants from Pitchers *et al.* (2019) in *ds* artificial selection experiment. Estimated effect sizes for SNPs are estimated from a GWAS in the DGRP using LASSO regularized coefficients. Average frequency is given with replicate lineage frequencies in brackets. Estimated effect is the l^2-*norm* of shape differences associated with the variant.

Variant	Estimated effect	DGRP MAF	Estimated MAF in synthetic outcross	Average allele frequency "up" selection	Average allele frequency "down" selection	Average allele frequency "control" selection
2L:655894	0.072	0.44	0.067	0 (0, 0, 0)	0 (0, 0, 0)	0.003 (0, 0, 0.0105)
2L:702560[a]	0.159	0.056	0.6	0.995 (1, 0.98, 1)	0.446 (0.32, 0.35, 0.67)	0.705 (0.69, 0.56, 0.87)
2L:702798	0.101	0.089	0.1	0.007 (0, 0.0217, 0)	0 (0, 0, 0)	0.005 (0, 0, 0.139)
2L:718623	0.225	0.033	0	0 (0, 0, 0)	0 (0, 0, 0)	0 (0, 0, 0)
2L:718627	0.11	0.033	0	0 (0, 0, 0)	0 (0, 0, 0)	0 (0, 0, 0)

[a] This is a complex polymorphism with linked SNPs and INDELs, in Pitchers *et al.* (2019) a SNP in this region was found to be linked. However, the variant calling pipeline used in this work recognized an INDEL in this region which was used for counting.

between them, we used *ds* and substituted *neur* (for *emc*) for the BSA.

Artificial selection of synthetic outbred population

The synthetic outbred population resulting from pooling DGRP lines was used as the parent population for artificial selection. Both the *ds* and *emc* artificial selection experiment were carried out with three independent replicates of each "up" and "down" selection regimes, along with unselected control lineages. Each generation, wings of live flies were imaged using the "wingmachine" system and shape data collected (Houle *et al.* 2003, Van der Linde 2004–2014, v3.72). Shape scores were calculated by projecting the data onto the *ds* or *emc* shape change vector as described above, and the 40 individuals each with highest or lowest shape scores, were selected to found the next generation (Supplementary Fig. 5a). For the control lineages, 40 individuals were randomly selected for the next generation within each replicate lineage. Following seven generations of selection, 75 individuals from each lineage were selected for pooled sequencing, described below. The response to selection was evaluated both by computing Procrustes distance (PD) between average shape of wings between generations one and seven, and using shape scores (projections) with a linear mixed effect model allowing for the fixed effect factors of treatment and sex, continuous predictors of centroid size and generation, with third-order interactions among these effects. The effect of generation was allowed to vary by replicate lineages (lmer(ds ~ (CS + Sex + line + gen0)^3 + (1 + gen0|line:rep)). Realized heritabilities were estimated separately for up and down selection lineages, from the slope of the regression of cumulative selection differentials on cumulative selection response, averaging over sex and with a random effect of replicate lineage.

Wild populations

For the BSA, wings for wild-caught individuals were dissected and mounted in 70% glycerol in PBS. Images of wings were captured using an Olympus DP30B camera mounted on an Olympus BX51 microscope (Olympus software V.3,1,1208) at 20× magnification. When possible, both left and right wings were dissected, imaged and averaged to calculate an individual's mean shape. For some individuals a wing was damaged so only one wing could be used. Shape was captured as described above. The total number of individuals phenotyped from each cohort can be found in Supplementary table 1.
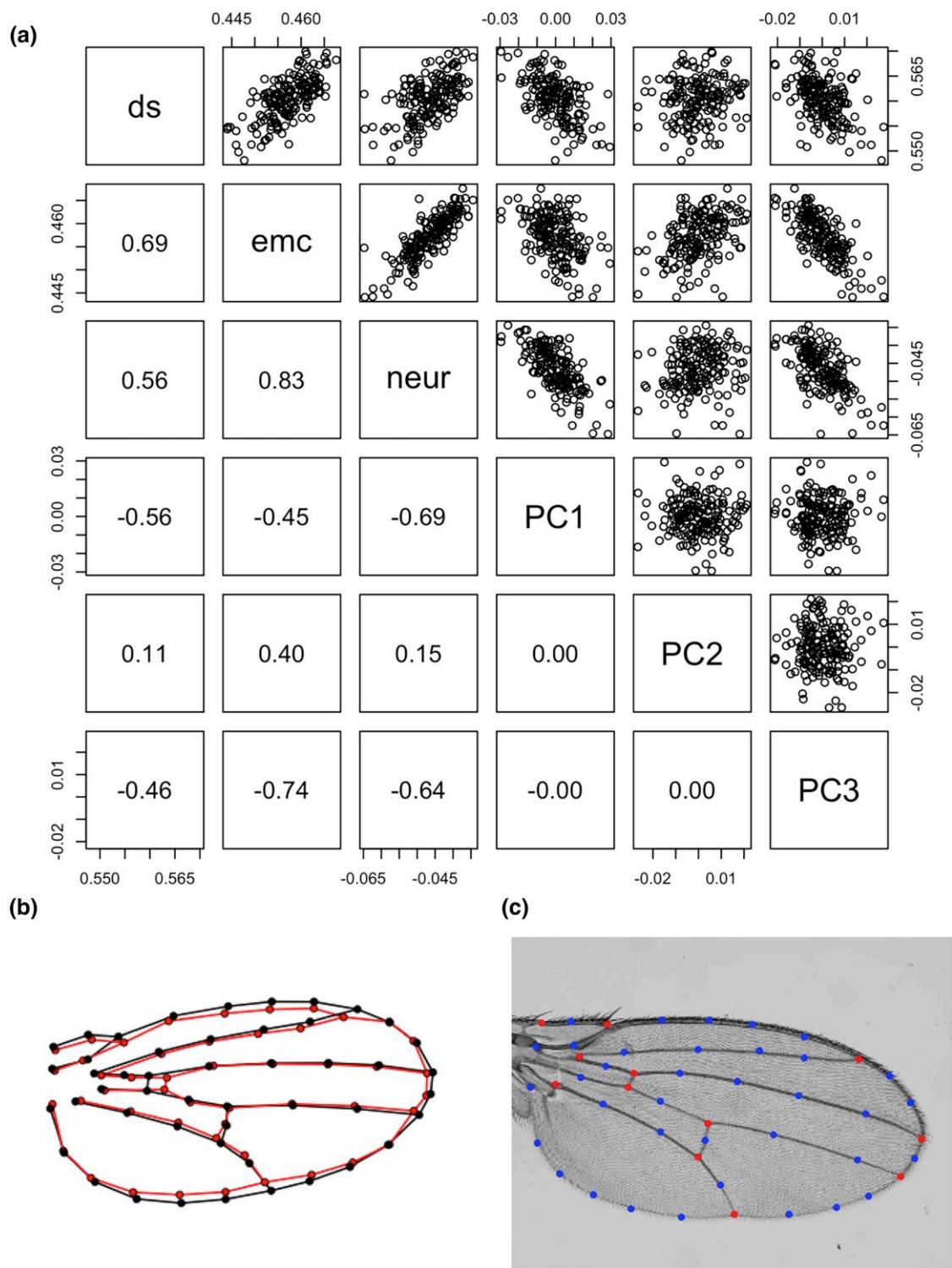
To remove allometric effects in the data, shape was regressed onto centroid size and the model residuals were used for all subsequent morphometric analysis. Only data from males was used to compare shape in wild populations, although, including

females from the FVW14 population and regressing shape onto centroid size and sex gave equivalent results (Supplementary Fig. 6). To test for shape differences between collection cohorts, the effect of centroid size and collection cohort on shape were modeled (procD.lm(shape ~ CS + pop_year)) using the procD.lm function in Geomorph v 3.1.3. (Adams and Otárola-Castillo 2013) and distances between populations were calculated using the pairwise function. To select individuals for sequencing, a "shape score" was calculated using the method described above. Shape data was projected onto the vector of shape change defined by the *ds* or *neur* knockdowns. The *emc* projection vector was not used for BSA due to the high similarity with *ds* shape change (Fig. 1), and the similarity of the selection response. Its inclusion would result in selection of largely the same cohorts of individuals for sequencing for both *ds* and *emc*. As an alternative, we utilized the *neur* shape vector as it was largely uncorrelated with that of *emc* and *ds*, but strongly correlated with natural variation in shape. The 75 most extreme individuals on the shape score distribution, within each wild-caught cohort, were selected for pooled sequencing. Allele frequencies within each population was estimated by sequencing 75 random individuals within each cohort. The difference vector between mean shapes of selected pools (within each population) was used to calculate PD between pools and the correlation of this shape change vector with the selection vector used. An estimate of genetic distances between populations was calculated using allele frequencies (mapping pipeline described below) in the pools of the 75 randomly selected individuals using Bray's distance with the vegdist() function from the vegan package (v2.6-2) in R.

Sequencing and genomic analysis

DNA extractions from pools of selected individuals was performed using a Qiagen DNeasy DNA extraction kit. Library preparation and Illumina sequencing were performed at the research technology support facility at Michigan State University. All library samples were prepared using the Rubicon ThruPLEX DNA Library Preparation kit, without a procedure for automatic size selection of samples. Paired end libraries (150 bp) were sequenced using Illumina HiSeq 2500, with each sample (either one pool of 75 individuals in the BSA or one pooled replicate lineage in the artificial selection) being run on two lanes.

Reads were trimmed with Trimmomatic (v0.36) to remove adapter contamination and checked for quality using FastQC prior to alignment (Bolger *et al.* 2014). Trimmed reads were aligned to the *D. melanogaster* genome (v6.23) using BWA-MEM (v0.7.8) (Li and Durbin 2010). Sequencing replicates of the same biological samples were merged using SAMtools (v1.11). PCR duplicates were removed using Picard with the MarkDuplicates tool

Fig. 1. Projections of data onto RNAi shape change vectors are correlated with major axes of shape variation among DGRP strains. a) Shape change vectors from RNAi titration experiments for *ds*, *emc*, and *neur* were used, and DGRP line means were projected onto these vectors to calculate shape scores. Eigenvectors for the PCA were estimated based on the same DGRP line means. Vector correlations between shape change vectors from RNAi knockdown: *ds–emc*: 0.65, *ds–neur*: 0.03, *emc–neur*: 0.30. b) Effect of *ds* shape change estimated from RNAi knockdown, effects not magnified. c) Landmarks (red) and semi-landmarks (blue) used in geomorphic morphometric analysis on a *Drosophila* wing. PCs 1–3 account for 22%, 20%, and 9% of the overall, among DGRP shape variance.

(v 2.10.3) and reads with a mapping quality score less than 20 were removed using SAMtools (Li *et al.* 2009). A local realignment around indels was performed using GATK using the IndelRealigner tool (v3.4.46). For artificial selection experiments, reads were merged for all up, down, and control selection lines as replicates lineages were independent. For wild cohorts, pools were not merged between populations. mpileup files were created using SAMtools and used for subsequent genomic analysis. Highly

repetitive regions of the *Drosophila* genome were identified and subsequently masked in mpileup files using RepeatMasker (v4.1.1) with default settings. INDELs and regions within 5 bp of an indel were identified and masked using popoolation2 scripts. Population genetic statistics were calculated using PoPoolation (v1.2.2) and PoPoolation2 (v1.201) (Kofler, Orozco-terWengel, *et al.* 2011; Kofler, Pandey, *et al.* 2011).

For the BSA in the wild-caught cohorts, a modified Cochran–Mantel–Haenszel (CMH) test was used to measure significantly differentiated sites between pools of individuals. Sampling effects were accounted for using the ACER package (v.1.0) in R, assuming $N_e = 10^6$ with 0 generations of differentiation between selected pools (Spitzer *et al.* 2020). To adjust for multiple testing, the *P*-value was corrected using a Benjamini–Hochberg correction (Benjamini and Hochberg 1995) with an adjusted alpha of 0.05. For each significant site from the CMH test, using an adjusted *P*-value cut-off of 0.05, we identified the nearest gene using BEDtools (v2.19.1) (Quinlan and Hall 2010). In addition, to account for sampling variation, we sampled genomic coverage to 75× for all samples, dropping sites that did not meet this threshold and repeating the CMH test. We confirmed that there was no association between genetic and shape differentiation between populations, and that the populations do not show strong phenotypic differentiation based on either overall shape variation, or shape scores used to identify selected individuals for BSA (Supplementary Figs. 3 and 7). There was some variation among populations in overall wing size (Supplementary Fig. 8), however we (assuming common allometry) adjusted for allometric effects on shape.

For artificial selection experiments, F_{ST} was calculated in 5,000 bp windows. We chose this window size as it is expected that blocks of LD in the synthetic outbred population will be much larger in comparison to that of the wild-caught samples (King, Macdonald, *et al.* 2012; King, Merkes, *et al.* 2012; Marriage *et al.* 2014). This statistic was used to compare the "up" selected pools to the "down" selected pools to help identify regions of differentiation between selected populations.

For the artificial selection comparisons, genes in regions of high F_{ST} were identified by finding overlaps between outlier windows and annotated *Drosophila* genes using GenomicRanges (v1.46.1) in Bioconductor. High F_{ST} was defined as F_{ST} values greater than three standard deviations above the mean. GO terms associated with identified genes were annotated using TopGO package (v2.34.0) (Alexa *et al.* 2006) in Bioconductor. GO enrichment was then performed to identify those terms overrepresented in the identified list using TopGO and a Fisher's exact test. Over representation of two GO terms in outlier windows (hippo signaling, GO:0035329; negative regulation of hippo signaling GO:0035331) were tested using a permutation test that randomly sampled genomic windows from the total windows for which F_{ST} was calculated and the permutation was run 1,000 times. The distribution of the ratio of observed to expected genes annotated with the term of interest within randomly sampled regions was compared to the number observed in the data.

Verification of *ds* indel in DGRP

Sanger sequencing was performed on individuals from a cross between DGRP lines predicted to have the polymorphism (DGRP-195, 28, 96, 48, 59, 801) and those without (DGRP-129, 301, 69, 385, 75, 83, 491, 34, 774) crossed to a line carrying a deletion in the region of interest (BDSC 24960) to account for potential residual heterozygosity in otherwise inbred strains. DNA was prepared by incubating flies in DNA extraction buffer (1 mM EDTA, 25 mM NaCl, 10 mM Tris–HCl pH 7.5) for 10 min, followed by storage at −2 °C. PCR application of the region of interest (Forward primer:
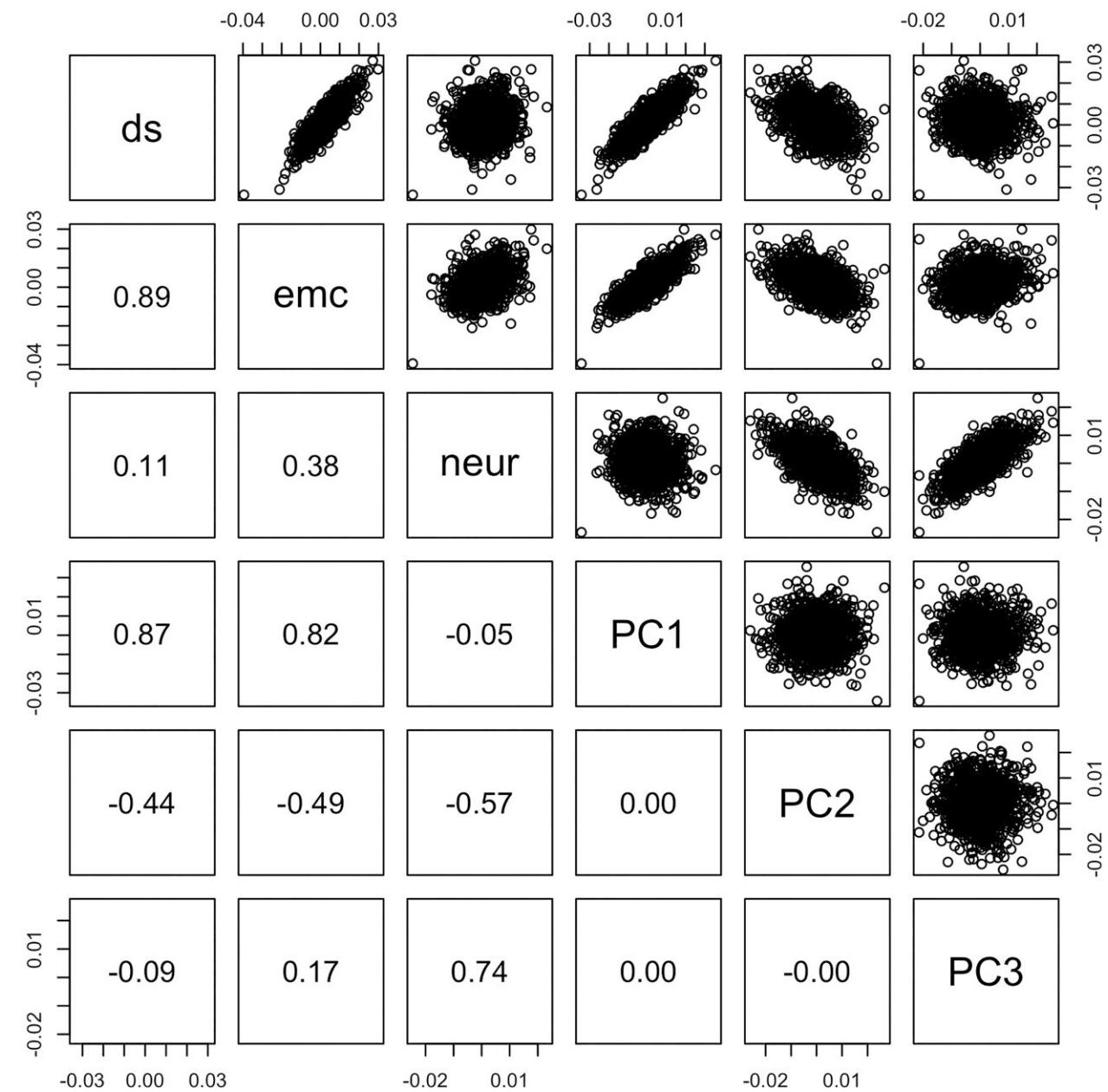
ggagtacaagttgctcgaac: Reverse primer: cagatcgtgttccctttagc) using Taq DNA polymerase (Gene DirectX) (PCR mix: 1 μL DNA, 1 μL forward primer, 1 μL reverse primer, 0.5 μL 10 mM dNTPs, 2 μL 10× PCR buffer, 0.1 μL taq, H20 to 20 μL). PCR conditions were as follows: 5 min 95 °C (30 s 95 °C, 30 s 55 °C, 30 s 72 °C)× 30. Reactions were checked on a gel and cleaned with the GenepHlow Gel/PCR Kit (Geneaid). Sanger sequencing reactions were performed by the Mobix Lab at McMaster University. All alignments were created using ClustalOmega (Madeira *et al.* 2022).

Results

Dachsous shape change is aligned with major axes of genetic and phenotypic variation in natural populations

To assess the relationship between shape change vectors and axes of natural variation described in the DGRP, mean shape vectors were calculated for each DGRP strain, then used in a PCA to summarize axes of variation among strains. Mean shape vectors for each strain of DGRP were projected onto shape change vectors for *ds*, *emc*, and *neur*, defined from the RNAi knockdowns (see Supplementary Fig. 5, which visually explains the procedure), generating gene specific "shape scores". Correlations between shape scores for individual DGRP projected onto the shape change vectors (Fig. 1; Supplementary Fig. 5), and with PC1 generated from the DGRP ($PC1_{DGRP}$) strains was estimated ($PC1_{DGRP}$-*ds*: $r = -0.56$; $PC1_{DGRP}$-*emc*: $r = -0.45$; Fig. 1). The correlation of the DGRP data, projected onto each of the *ds* and *emc* shape change vectors was also correlated (Fig. 1, *ds*–*emc*: $r = 0.69$). This is likely due to the correlation between gene-specific shape change vectors themselves ($r = 0.65$), based on RNAi titration experiments. Projections of the DGRP data onto the vector defining the *neur* shape change is aligned with PC1 ($PC1_{DGRP}$-*neur*: $r = -0.69$) and PC3 ($PC1_{DGRP}$-*neur*: $r = -0.64$), indicating this as an important axis of shape variation in this population (Fig. 1), that is moderately similar to projections onto *ds* (*ds*–*neur*: $r = 0.56$) and very similar to *emc* (*neur*–*emc*: $r = 0.83$) shape change vectors. Interestingly, the strength of the correlation for the DGRP strains projected onto these vectors, differs from the magnitude of correlations for the RNAi titration vector of *neur* with that of *ds* ($r = 0.034$) or *emc* (0.3). Because of these observed correlations, and previous associations observed (Pitchers *et al.* 2019) *ds*, *emc*, and *neur* were selected as focal genes for subsequent studies.

We also examined the relationship between direction of phenotypic effects with the wild-caught cohorts. For these samples, phenotypic variance for shape is due to the joint contribution of genetic and environmental effects. To illustrate the difference in shape variance in wild populations and the DGRP, we calculated correlations between the first three eigenvectors for shape in the DGRP, the combined wild cohorts as well as the CMO cohort alone. We observed low correlations between the DGRP eigenvectors and those estimated from wild populations (Supplementary Table 2). As observed with the DGRP, there is a substantial correlation between projections of shapes of individuals onto the *ds* shape change vector and PC1 (defined by phenotypic variation among wild-caught files, PC_{wild}) in most of the sampled cohorts (*ds*-$PC1_{wild}$ PHO: $r = 0.78$; CMO: $r = 0.87$; FVW13: $r = -0.22$; FVW14: $r = 0.95$, Fig. 2, Supplementary Fig. 9). In cohorts where the *ds* shape change vector was not correlated with PC1, specifically the FVW13 collection, this vector is correlated with PC2 (*ds*-$PC2_{wild}$ PHO: $r = 0.12$; CMO $r = -0.44$; FVW13: $r = -0.63$; FVW14: $r = 0.19$; Fig. 2; Supplementary Fig. 5). The pattern for the wing shape from wild-caught individuals projected onto the *emc* shape change vector was generally similar to that observed for *ds* (Fig. 2). We also observe a correlation between
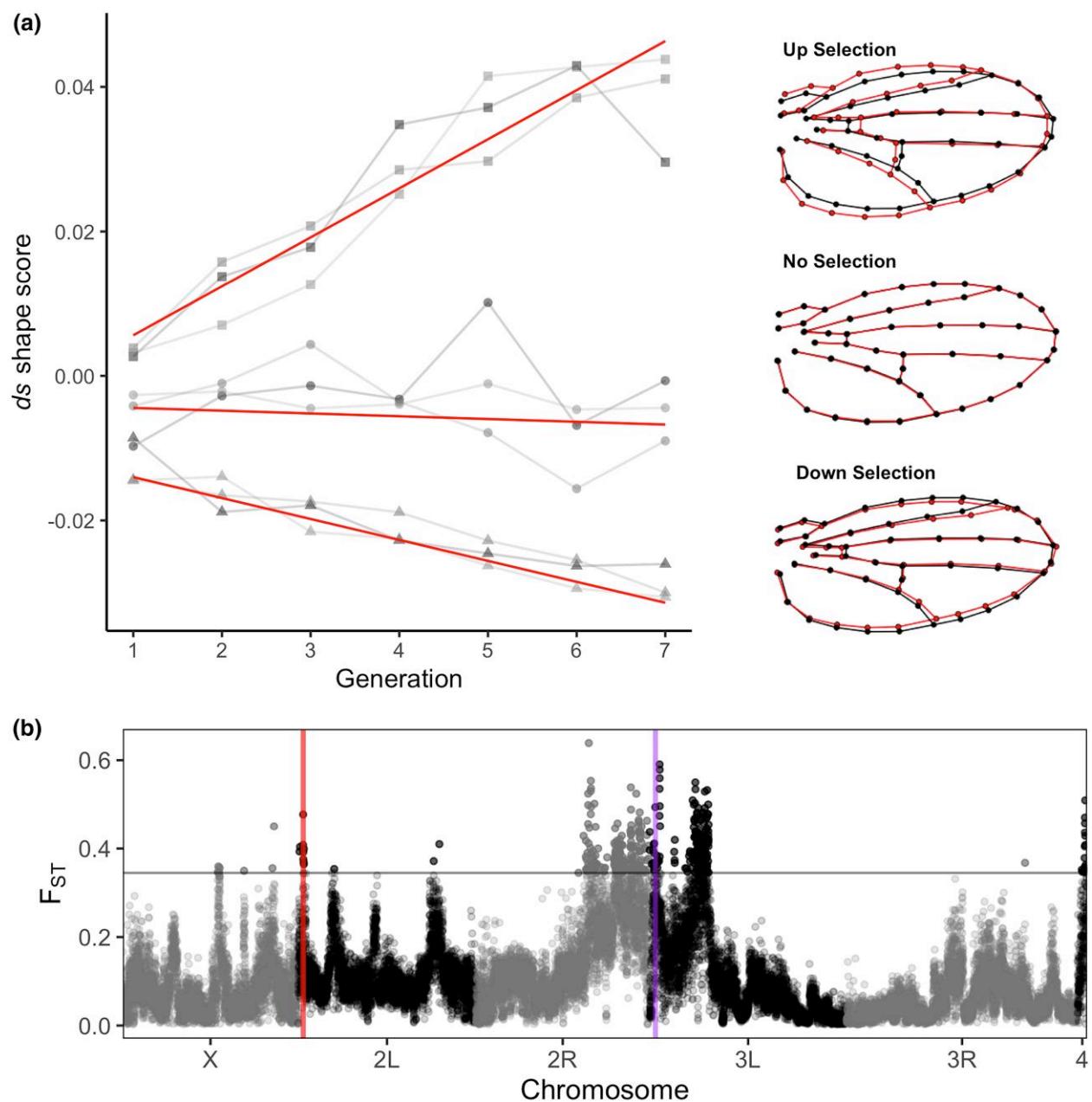
Fig. 2. Projections of data onto RNAi shape change vectors are correlated with major axes of shape variation in wild-caught *Drosophila*. Correlations between projection of shape data from CMO population onto *ds*, *emc*, and *neur* RNAi shape change vectors, and the first three eigenvectors from the PCA, calculated from shape data from all samples in the CMO population. PCs 1–3 account for 24%, 18%, and 9% of overall shape variance in the CMO population.

neur shape change and PC1 in most cohorts (*neur*-PC1$_{wild}$ PHO: $r =$ 0.51; CMO: $r = -0.051$; FVW13: $r = -0.95$; FVW14: $r = -0.084$; Fig. 2; Supplementary Fig. 7). As with the *ds* shape change vector, in some cohorts such as the CMO the stronger correlation is between the *neur* shape change vector and PC2 (PHO: $r = 0.22$; CMO: $r = -0.57$; FVW13: $r = -0.059$; FVW14: $r = 0.85$; Fig. 2; Supplementary Fig. 7). Interestingly, in the CMO cohort, the correlations between the projection of shape data onto the *ds* and *neur* shape change vectors is low (*ds-neur*: $r = 0.11$, Fig. 2).

Multiple loci linked to hippo signaling—including *ds*—respond to artificial selection for *ds* and *emc* shape changes

To examine if variants in *ds* are contributing to shape variation, and independently replicate the findings of the earlier GWAS

(Pitchers *et al.* 2019), we performed an artificial selection experiment for wing shape along the *ds* shape change vector, and examined the genomic response to selection. By the final generation of selection, we observed a substantial shape change in both the "up" (females: PD = 0.039, males: PD = 0.044) and "down" directions (females: PD = 0.022, males: PD = 0.022), compared to the base population at the start of the experiment. In comparison, the shape change among unselected control lineages was much smaller (females: PD = 0.005, males: PD = 0.005) (Fig. 3; Supplementary Fig. 10). The direction of phenotypic shape change after seven generations of selection was in a similar direction to the *ds* shape change vector (defined by RNAi knockdown) for both the up (females: $r =$ 0.90, males: $r = 0.90$) and down (females: $r = -0.82$, males: $r = -0.77$) selection lineages. Realized heritabilities, averaged over sex and replicate were moderate (Supplementary Fig. 11, up =

Fig. 3. Artificial selection along *ds* shape change vector influences allele frequencies of variants at *ds*. a) Phenotypic response to selection based on *ds* shape change vector. Only data from females is plotted for ease of visualization. Each replicate of up (squares), control (dots), and down (triangles) selection lineages are plotted (greys). Estimated response to selection shown along straight red lines. Wing plots represent the effect of selection on shape change between generation one and seven (red, effects not magnified). b) Genomic differentiation (F_{ST}) between up and down selection treatments measured in 5,000 bp windows. Red and purple line represent genomic locations of *ds* and *emc* respectively. Grey line represents three standard deviations from genome wide mean F_{ST}.

0.38, 95% CI: 0.25–0.50; down = 0.28, 95% CI: 0.24–0.50). Hippo signaling, including the effects of *ds*, is often associated with changes in size (Pan 2007). However, we do not observe a significant change of wing size in our selection lineages in either sex (Supplementary Fig. 12). It is possible that with more generations of selection we would have observed a clear change in size, as there is a trend indicating such divergence (Supplementary Fig. 12).

Genome-wide patterns of F_{ST} were examined between up and down *ds* selection lineages. We observed strong genetic differentiation linked with the *ds* locus (Fig. 3; Supplementary Fig. 13), along with several other regions in the genome. One of the SNPs in the intron of *ds* (2L:702560), identified in Pitchers *et al.* (2019) through

GWAS, showed the expected pattern of response to selection, with opposing sign in up and down selection lineages, with the SNP going to high frequency in all three up selection lineages (Table 1). It should be noted that this SNP is near a complex polymorphism including an insertion of 18 bp that may result in inaccurate genotyping at this locus (Supplementary Fig. 14). Gene ontology analysis for genes in regions of the genome with an F_{ST} greater than 0.345 (three standard deviations from mean F_{ST}), show enrichment for hippo signaling loci (Supplementary Table 3). The top 20 enriched terms are all related to cell signaling and development. Of note is the inclusion of the terms for "negative regulation of hippo signaling" (GO: 0035331), and "hippo

signaling" (GO: 0035329) in this list (Supplementary Table 3; Supplementary Fig. 13). Using a permutation test we confirmed these results, selecting random sets of genomic intervals equal in size to the number of observed outlier windows, and measured the ratio of genes annotated to the expected number of genes in these regions. The observed value for the terms for hippo signaling (ratio = 4.76) and negative regulation of hippo signaling (ratio = 9.23) were in the upper 99.5% percentile in comparison to the distributions under permutation (Supplementary Fig. 15).

For the artificial selection experiment based on the *emc* shape change vector we observed phenotypic differentiation under artificial selection in both up (females: PD = 0.043, males: PD = 0.040), and down directions (females: PD = 0.021, males: PD = 0.020), with little change in control lineages (females: PD = 0.009, males: PD = 0.008) (Fig. 4). The direction of phenotypic change is correlated with the *emc* (RNAi knockdown) shape change vector in both up (females: $r = 0.75$, males: $r = 0.69$) and down (females: $r = -0.69$, males: $r = -0.75$) directions. Realized heritabilities, averaged over sex and replicate were calculated for both up and down lineages (Supplementary Fig. 16, up = 0.38, 95% CI: 0.29–0.47; down = 0.28, 95% CI: 0.21–0.35). Genetic differentiation linked to the *emc* locus was modest following selection, but we again observed striking genetic differentiation linked to *ds* (Fig. 4; Supplementary Fig. 13). Notably, as seen in Supplementary Fig. 1, the site frequency spectrum (SFS) suggests modest allelic variation at the *emc* locus in the synthetic outbred population. Using a three standard deviation cut-off for F_{ST}, we did observe enrichment for various developmental GO terms, but not of hippo signaling terms (Supplementary Table 4; Supplementary Fig. 13).

BSA in wild-caught cohorts does not recapitulate effects of the GWAS or artificial selection

Having demonstrated that variants in (or linked to) *ds* respond to artificial selection for wing shape along the *ds* shape change vector, we next wanted to determine whether we could recapitulate these findings with wild-caught individuals. In addition to determining whether we can replicate effects in wild cohorts, it provides the opportunity to identify causal SNPs because of low LD generally observed in wild-caught *Drosophila*. Wild-caught populations introduce considerably more environmental variation for shape along with a different site frequency spectrum for variants contributing to shape variation (and *ds* like shape changes specifically). In particular, it is known that several of the variants that the original GWAS detected in *ds* have low minor allele frequency (MAF) (Pitchers *et al.* 2019) (Table 2). The SNP at 2L:702560 does appear to be at intermediate frequency but it occurs both directly before and after an indel, making alignment and variant calling in this region challenging (Supplementary Fig. 14). We have included the frequencies (Table 2), but these results should be interpreted with caution due to the technical complexities of mapping and variant calling close to indels.
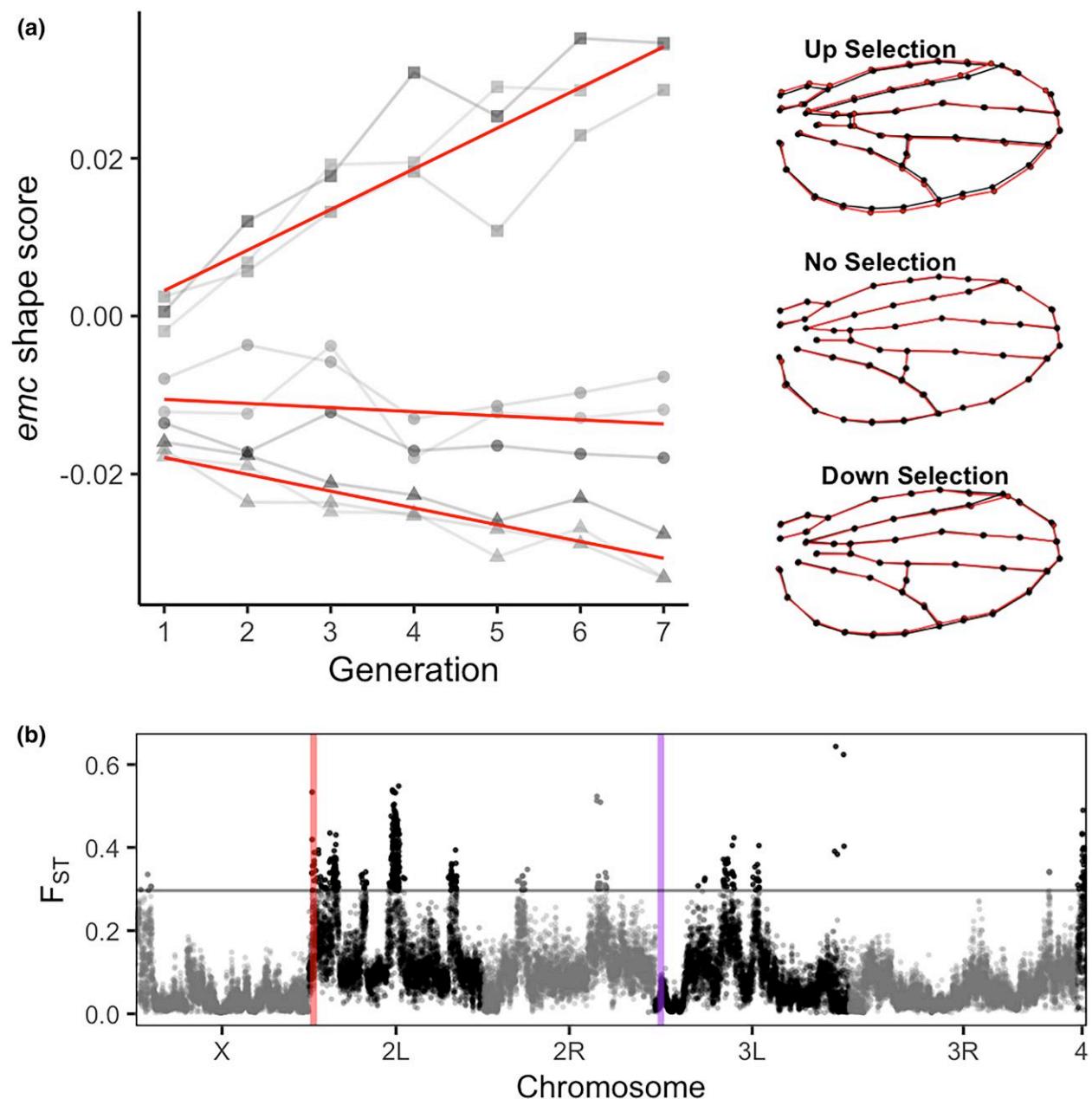
As we sampled multiple cohorts of wild-caught flies in different locations and years in Michigan (USA), we wanted to confirm that any phenotypic differentiation among these samples was modest and would not impact genomic analysis for the BSA. We observe modest, statistically significant wing shape differences among cohorts from a Procrustes ANOVA, utilizing permutations of the residuals for the relevant "null" model (Supplementary Table 5; R^2 = 0.16, $F = 351$, $Z_{RRPP} = 18.3$, $P = 0.001$) (Collyer and Adams 2018). This appears to be due to differences in wing shape between the PHO population and other populations based on pairwise PD (Supplementary Table 5; Supplementary Fig. 3). In a joint PCA

including all populations, there is very modest separation between populations using allometry adjusted shape (Supplementary Fig. 3). Most relevant to the BSA approach we used, when we project all wild-caught individuals onto the *ds* and *neur* vectors, there is no clear separation among sampling locales (Supplementary Fig. 7). There is some variation in wing size between populations (Supplementary Fig. 8), but this is unlikely to influence downstream analysis as we use size adjusted estimates. There is little evidence of genetic differentiation between populations with the two collections from FVM separating more on a Principal Co-ordinate Analysis (PCoA) (Supplementary Fig. 17) than other sampling locales. There is also no relationship between genetic and phenotypic distances between samples (Supplementary Fig. 18). These results suggests that the multiple sampling locales should not influence downstream genomic analysis as individuals used for generating pools were compared within each population, and we observe little evidence for substantial differences among populations.

Because there is a single bout of phenotypic selection distinguishing pools for the BSA, changes in shape and allele frequencies are expected to be modest. We observe shape differences between the two pools within each population (PD = CMO: 0.033; PHO: 0.036; FVW13: 0.040; FVW14: 0.041; Supplementary Fig. 19). Correlations of the shape difference vectors of the pools (i.e. difference between the two pools created from the extremes along the *ds* shape change axis), and the direction of the *ds* shape change vector used for selection, is high (CMO: 0.94, PHO: 0.79, FVW13: 0.92, FVW14: 0.90).

BSA genome scans show little evidence of genetic differentiation linked to the *ds* gene (Fig. 5). Across the genome, 15 sites were detected as significantly differentiated between "up" and "down" selected pools based on a CMH test with FDR cut-off of 5% (Fig. 5; Table 3). The genes nearest to these sites are not associated with hippo signaling pathways or implicated in the development of the *Drosophila* wing (Table 3). Because PHO had somewhat distinct shape variation from the other populations and had a lower correlation of the difference vector between selected pools and *ds* shape change vector, we repeated the CMH test with this population left out. We observe significant differentiation at 174 sites between "up" and "down" pools (Supplementary Table 6; Supplementary Fig. 20). We identified the nearest genes to these sites and GO analysis indicated enrichment for wing development terms, in particular related to Wnt signaling, but not hippo signaling terms (Supplementary Table 6). Importantly, we do not observe differentiation linked to *ds* or any other hippo loci. To ensure that the results we obtained were not due to uneven coverage between samples, we down-sampled genomic coverage to 75× for each sample, dropping sites that did not meet this threshold. Significant differences were detected at 19 sites (Supplementary Fig. 21; Supplementary Tables 7 and 8), but none of these overlapped with those identified using all the genomic data. Two of the significant sites are located in the *dumpy* gene, a gene known to have a role in wing morphogenesis during pupation (Etournay *et al.* 2015). F_{ST} between selected and random pools within each cohort are generally low (Supplementary Fig. 22).

In addition to the BSA selection based upon the *ds* shape change, we also selected pools of individuals based on the *neur* shape change vector. We did not use *emc* shape change in this experiment due to the high similarity between the *ds* and *emc* shape change vectors ($r = 0.65$), and the similar response to selection reported above. We selected the *neur* shape change vector as it is not aligned with *ds*, but does align with directions of natural variation, in wild populations (Figs. 1 and 2; Supplementary Fig. 9). Additionally,
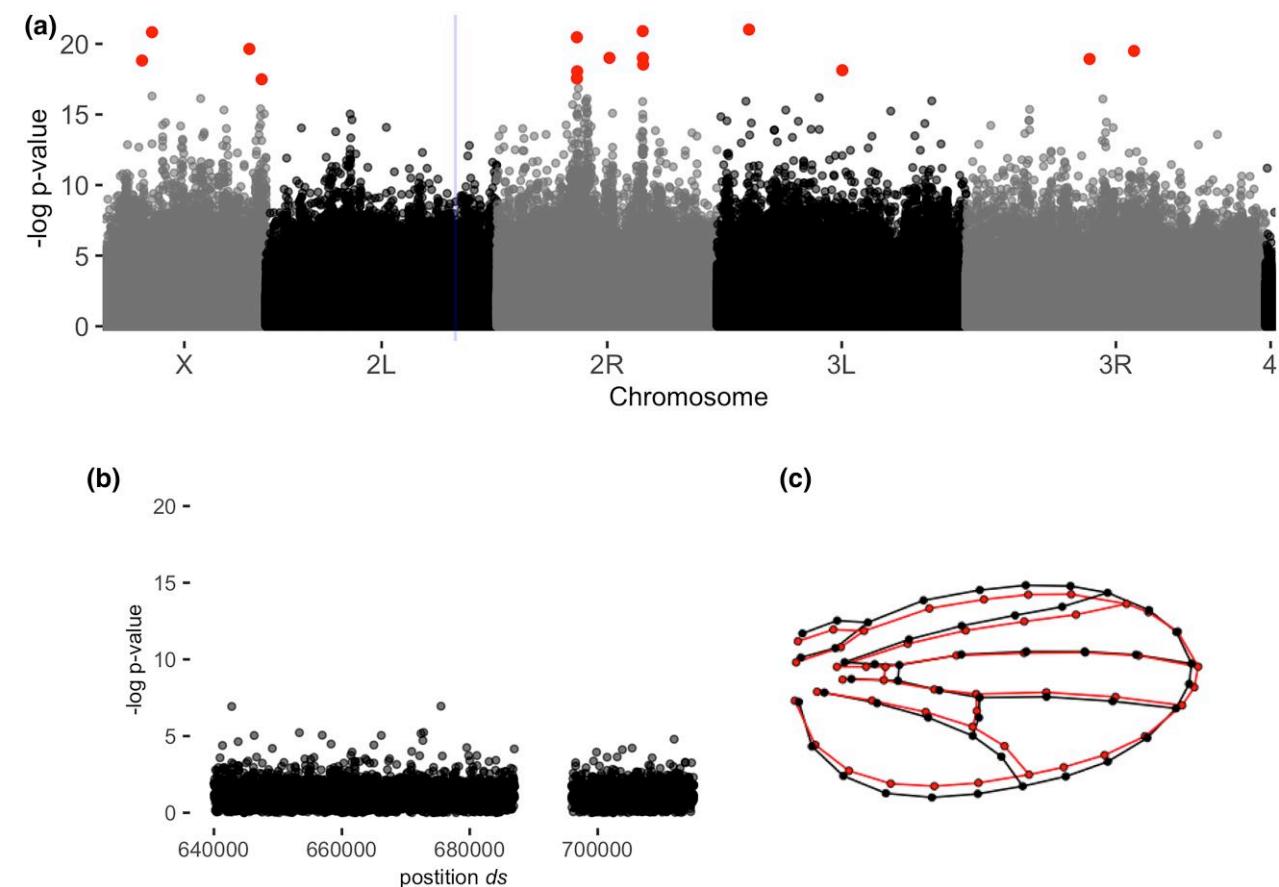
Fig. 4. Artificial selection along *emc* shape change vector has modest influence on allele frequencies at *emc*, but a greater impact at the *ds* locus. a) Phenotypic response to selection based on the *emc* shape change vector. Only data from females is plotted for ease of visualization. Each replicate of up (squares), control (dots), and down (triangles) selection lineages are plotted in greys. Estimated response to selection shown along red lines. Shape change between generation 1 and 7 is indicated on the right. Shape effects have been magnified 5×. b) Genomic differentiation between up and down selection lineages (F_{ST}) measured in 5,000 bp sliding windows. Red and purple vertical lines represent genomic locations of *ds* and *emc*, respectively. Grey line represents three standard deviations from genome wide mean F_{ST}.

there is little relationship between the *ds* and *neur* shape change axis (*r* = 0.12, Supplementary Figs. 7 and 8) in the wild-caught cohorts. We observe shape changes between pools of individuals (PD = CMO: 0.027; PHO: 0.028; FVW14: 0.041; FVW13: 0.038, Supplementary Fig. 23). There is little evidence of genetic differentiation between *neur* selected pools (Supplementary Fig. 24). Only four sites were identified as being significantly differentiated between pools and none of these sites are associated with wing development (Supplementary Table 9). When population differentiation between pools within populations is measured using F_{ST}, genetic differentiation remains low across the genome (Supplementary Fig. 25)

Discussion

The primary goal of this study was to determine whether we could recapitulate genetic effects initially observed through a traditional GWAS using an "inverted" approach: artificially selecting on phenotypes and observing changes in allele frequencies. We observed that shape changes associated with the *ds*, *emc*, and *neur* genes were associated with major axes of genetic variation among a panel of wild type strains (DGRP) reared in the lab, and axes of phenotypic variation among wild-caught individuals (Figs. 1 and 2). After observing a strong response to artificial selection along two shape change vectors (*ds* and *emc*), we examined patterns of

Fig. 5. Genetic differentiation between pools selected based on *ds* shape change among the wild-caught cohorts. a) Genome-wide scan for differentiated loci between pools selected based on *ds* shape change vector using the CMH test implemented in ACER. Points in red indicate sites with significant differentiation. Position of *ds* gene in blue b) Genomic differentiation at *ds* between pools selected based on *ds* shape change vector. No sites are significantly differentiated in *ds*. The large gap in sites is due to a masked region in the genome due to repetitive sequence and poor (syntenic) mapping scores. c) Shape difference between selected pools of individuals from one representative (CMO) population, with the mean shape of pools represented in black and red.

genomic differentiation and observed substantial changes in allele frequency for markers linked with *ds* itself (Fig. 3), and markers linked to numerous genes associated with hippo signaling (Supplementary Figs. 13 and 15).

In contrast, our BSA experiments, using pools of wild-caught individuals chosen to be phenotypically divergent on the same shape vectors, did not detect differences in the loci identified in the artificial selection experiments (Fig. 5; Supplementary Fig. 24). As we discuss in detail below, these seemingly contradictory results are in fact not that surprising.

Following artificial selection based on *ds* shape change we observe allele frequency changes not only at *ds* but also linked to a number of other hippo signaling loci (Fig. 3; Supplementary Table 3). The previous GWAS study identified a number of loci associated with wing shape variation in the DGRP, however, this approach cannot predict which alleles are causative (Pitchers *et al.* 2019). In our synthetic outbred population, we maximized variation among haplotype blocks containing many of the candidate SNPs in *ds*, increasing our ability to detect frequency changes at and near the implicated variants. Although LD blocks in the outcrossing population from this study remain large, *ds* variants exist on multiple distinct haplotypes, allowing for an examination of allele frequency changes for each. Of particular interest is SNP 2L:702560, previously identified though GWAS (Pitchers *et al.* 2019) as influencing wing shape variation. It was driven to near

fixation in each of the artificial selection lineages (Table 1). Although this polymorphism is annotated as a SNP, this region may contain a complex polymorphism (Supplementary Fig. 14), making it difficult to accurately assess genotypic calls. Because of this, the predicted allele frequency in the founding population and allele frequencies in this region may be inaccurate. Previous studies demonstrate the importance of alleles at intermediate frequency in founding populations to those contributing to responses to selection over short timescales (Kelly and Hughes 2019). If this polymorphism is at a more intermediate frequency in the founding population, it would be more likely to be captured by selection during these experiments. Additionally, haplotype blocks in the initial population are large, and may contain many potential functional variants. However, based on the results of both the current and previous studies, these *ds* variants associated with 2L:702560 are good candidates for functional validation in future work.

When selecting on the *emc* shape change vector, which is similar to that of the *ds* shape change, we observe only a modest allele frequency change at *emc*, and a more robust response at *ds* (Fig. 4). In hindsight, this is not particularly surprising and there are multiple contributing factors. Given the increased genetic diversity at *ds* compared to *emc* in the founding population, alleles in *ds* may have provided a more accessible genetic target, as selection can only act upon the diversity available in the population. Additionally, if our

Table 2. *ds* Variants from Pitchers *et al.* (2019) in wild-caught cohorts used in the present study. Estimated effect sizes for SNPs are estimated from the DGRP GWAS with LASSO regularized coefficients. MAF in wild cohorts was estimated from sequenced pools of 75 random individuals.

Variant	Estimated effect (Pitchers 2019)	DGRP MAF	Estimated MAF CMO	Estimated MAF FVW13	Estimated MAF FVW14	Estimated MAF PHO
2L:655894	0.072	0.445	0	0	0	0
2L:702560[a]	0.159	0.056	0.375	0.473	0.485	0.336
2L:702798	0.101	0.089	0.077	0.101	0.044	0.034
2L:718623	0.225	0.033	0.051	0.021	0.044	0.100
2L:718627	0.11	0.033	0.055	0.020	0.046	0.099

[a] This is a complex polymorphism with linked SNPs and INDELs, in Pitchers *et al* (2019) a SNP in this region was found to be linked. However, the variant calling pipeline used in this work recognized an INDEL in this region which was used for counting.

Table 3. Significantly differentiated variants for *ds* shape change from the wild-caught cohorts (BSA).

Location	CMH P-value (FDR corrected)	Gene	FlyBase ID	Distance from ORF (bp)
2R:17491270	0.026	NT5E-2	FBgn0050104	0
2R:17498059	0.034	CG30103	FBgn0050103	2,061
2R:17515133	0.022	CG4853	FBgn0034230	0
2R:20537878	0.013	CG13423	FBgn0034513	0
2R:23601278	0.005	CG10332	FBgn0260455	0
2R:23601278	0.005	IM18	FBgn0067903	0
2R:23613785	0.013	Eglp4	FBgn0034885	0
2R:23613785	0.013	Eglp2	FBgn0034883	0
2R:23646252	0.016	retn	FBgn0004795	0
3L:12831924	0.005	CG10960	FBgn0036316	0
3L: 20999119	0.022	skd	FBgn0003415	0
3R: 21523866	0.013	CG7956	FBgn0038890	0
3R: 2559549	0.011	Pzl	FBgn0267430	0
X: 14891220	0.013	Flo2	FBgn0264078	0
X: 14891220	0.013	CG9514	FBgn0030592	0
X:16039731	0.017	Muc14a	FBgn0052580	0
X: 793052	0.011	CG16989	FBgn0025621	95
X: 9448676	0.034	mgl	FBgn0261260	0

estimated direction of effects and selection for *emc* (based on RNAi knockdown) was not well aligned with the actual direction of *emc* SNP effects, this could result in weaker selection on variants at the *emc* locus. It is worthwhile pointing out the small magnitude of the *emc* shape change vector (0.44) relative to *ds* (5.5). However, previous work has indicated that there is a relationship between this estimated *emc* shape change vector (from RNAi) and the effect of SNPs in *emc* on shape change (Pitchers *et al.* 2019).

In addition to a response on allele frequency associated with *ds*, our results suggest a response on segregating variation at other hippo signaling loci in the *ds* artificial selection experiment. Earlier work has suggested that the direction of effects within signaling pathways are inconsistent for alleles of small effect (Dworkin and Gibson 2006). However, allelic effect sizes in the 2006 study were heterogeneous and may result in direction and magnitude being confounded. In contrast, in both the current and the Pitchers *et al.* (2019) studies, we estimated the direction of genetic effects by titrating gene knockdown. The strength of this approach is highlighted in the result that segregating variation at multiple hippo loci was selected on (Supplementary Figs. 13 and 15). Our finding is consistent with models for the architecture of complex traits that predict that many alleles of small effect will contribute to trait variation with many genes within developmental pathways (Boyle *et al.* 2017; Wray *et al.* 2018). This pathway response has also been demonstrated in human adaptation to pathogen resistance (Daub *et al.* 2013) and high altitude (Gouy *et al.* 2017). These results are consistent with the expectation that polymorphisms in the same developmental pathway would show correlated phenotypic effects and therefore

correlated genomic responses to selection. However, this may not be reflective of all wild-caught populations. In this study, we generated a population that had high diversity at *ds*, while these variants are at much lower frequency in natural populations (Table 2). The amount of selectable variation a variant provides, depends on both effect size, a, and variant frequencies, p, as $V_A = 2p(1 - p)a^2$. When allele frequencies are near 0 or 1, even variants with large effects will have only a small contribution to short-term selection response. Therefore, the outcrossed population we created here is an ideal situation to validate the existence of the measured effects. It is unlikely to be typical of natural populations where functional variants may be rare.

Given the clear and robust response observed in the artificial selection experiment, it may seem surprising that we do not observe allele frequency changes in the BSA using the wild cohorts. Indeed, previous work has demonstrated that variants in *Egfr*, could be replicated in wild-caught samples (Dworkin *et al.* 2005; Palsson *et al.* 2005) and were also found in genome wide associations (Pitchers *et al.* 2019). However, there are many explanations for why we did not detect these allele frequency changes in our experiment. First, the addition of environmental variation to the system introduces additional complications. In the aforementioned example with *Egfr*, the genetic effect of the SNP in wild-caught cohorts was ~10% of the magnitude estimated in lab-reared flies. As discussed previously, the *ds* variants implicated in the previous GWAS study are at low frequency in the natural cohorts (Table 2). Given that natural populations of *Drosophila* are generally large and wing shape is likely under weak selection (Gilchrist and Partridge 2001), mutation-drift-selection balance may maintain

most variation, resulting in low minor allele frequencies at these sites. Because allelic contribution to wing shape is expected to be both rare in wild populations and of small phenotypic effect, we do not expect large allele frequency changes given only one "generation" of selection. Using the approach of ACER (Spitzer *et al.* 2020) to account for sampling effects, we observe few differentiated sites, and none in the *ds* gene, indicating that BSA may not be well suited to identify modest allele frequency changes, thus, not particularly effective for polygenic traits. Although our approach was tailored to look for variants that had consistent direction of frequency changes across the four collection cohorts, it is possible that different loci were contributing variation within each cohort. We attempted to address this question by examining allele frequency changes between selected pools within each cohort (Supplementary Figs. 22 and 24) but could not identify specific loci contributing to differences within any one population. Previous successful BSA studies identified smaller numbers of contributing loci with few polymorphisms contributing to the trait of interest. For example, in *Drosophila*, a number of melanin synthesis genes contributing to variance in pigmentation between populations were identified using a BSA (Bastide *et al.* 2013). Pigmentation may represent a relatively "simpler" genetic architecture (fewer variants of individually larger genetic effect, smaller impact of environmental variation, smaller mutational target size) and if so, this may have enabled the success of the BSA approach with such systems. In the case of wing shape, we know that many alleles of small effect contribute to variation in the trait (Pitchers *et al.* 2019).

Our approach for the BSA was to perform the same phenotypic selection within each of four distinct "populations". It is important to recognize that there was heterogeneity among our populations, not only in allele frequencies, but in environmental variance and potentially GxE, even though all were caught in locales in lower Michigan. We detected small degrees of phenotypic and genetic differences between cohorts, however these effects are neither correlated with one another, nor related to the *ds* and *neur* shape scores used for selecting individuals (Supplementary Figs. 3, 7, and 18). The population from the PHO was phenotypically distinct from the other populations. When we performed the BSA without this population, we observed a larger set of variants associated with shape (Supplementary Fig. 20), albeit still not showing any effects at *ds* or *neur* genes themselves. One possibility is that the increased number of sites when the PHO sample is removed from analysis represents an unknown statistical artefact we have not identified. However, a more likely explanation is that there are some large unknown environmental influences (E), or that the genetic effects show a degree of GxE (with a specific environment in PHO) that contributed to shape variation along the *ds* direction in this population. Such obfuscating effects have been observed before with the previously discussed *Egfr* example, where the SNP effect identified and validated in multiple contexts (Dworkin *et al.* 2005; Palsson and Gibson 2004; Palsson *et al.* 2005) could not be detected in one natural population, despite being at intermediate frequencies in each sample (Palsson *et al.* 2005). Importantly, we did detect differentiation at sites associated with developmental processes in the wild cohorts, suggesting that the failure to detect variation linked to *ds* or other hippo signaling loci (Table 3; Supplementary Tables 6, 7) is not due simply to a lack of power.

The response to selection at *ds* and other hippo signaling loci in the artificial selection experiment based on *ds* shape change indicates that this is an important source of variation for wing shape. Coupled with the alignment of phenotypic effects of perturbations in genes in this pathway with directions of **G** and **P**, this finding

may seem to suggest a developmental bias in available variation. However, we caution against such interpretations based solely on the findings in this study. The structure of the **G** matrix strongly influenced our findings as we artificially created a population to maximize genetic diversity at *ds*. When another effect is aligned with *ds* shape change, as in the case of *emc* shape change, we observed the same response at the hippo signaling loci and not at *emc*. Only the genetic diversity in the starting population was available to be selected on so this influenced selection towards the "spiked in" *ds* variants, even if the inferred phenotypic effects of *emc* variants are very similar. Alternatively, the inferred *emc* direction of effects (via RNAi knockdown) may be sufficiently "distant" from true effects of *emc* variants. If this was the case, we were ineffectively selecting for *emc* shape changes. In other cases where single genes are implicated in divergence between multiple populations, such as *mc1r* in mice (Steiner *et al.* 2007) or *pitx1* in stickleback (Chan *et al.* 2010), other factors such as low pleiotropy, developmental and mutational constraints, and history of selection in the population are used to explain why these genes are so often implicated in evolutionary change (Gompel and Prud'homme 2009; Stern and Orgogozo 2008; Martin and Orgogozo 2013). In our case, it is not *ds* itself that is special but rather the orientation of the **G** matrix to align $\mathbf{g_{max}}$ with the direction of effect for *ds* that shapes our results. Selection acts on variants aligned with the vector of selection (Reddiex and Chenoweth 2021). By varying the orientation of $\mathbf{g_{max}}$ in the parental population, we would be able to address questions about the repeatability of hippo overrepresentation and if this can be explained by more than just the orientation of **G.**

Despite the need for skepticism about the potential for developmental bias influencing directions of variation, the correlated response of sites linked to multiple other hippo signaling genes is intriguing. Coupling of more traditional mapping approaches like GWAS with short-term artificial selection provides an additional route to validation and replication of genetic effects. It also suggests that using multivariate data to address the distribution of genetic effects will pay long-term dividends to our understanding of both inheritance and the evolution of multivariate traits.

Data availability

All code and processed data needed to complete the analysis is available on GitHub at: https://github.com/DworkinLab/WingShapeBSA/. A static version of the repository is available on figshare (https://doi.org/10.6084/m9.figshare.22141154.v1).

All raw sequence data available as part of the NCBI Short Read Archive, BioProject PRJNA936488 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA936488/), with individual sequence accessions SAMN33354503–SAMN33354634.

Supplemental material available at GENETICS online.

Acknowledgements

We thank the Hosken lab for *D. simulans* wing shape data. Yun Bo Xi aided with the sequencing of *ds* haplotype. We thank Dr. Brian Golding for computational resources. We thank Dr. Catherine Peichel and two anonymous reviewers for helpful feedback on the manuscript.

Funding

Conflicts of interest

The authors declare no conflict of interest.

Literature cited
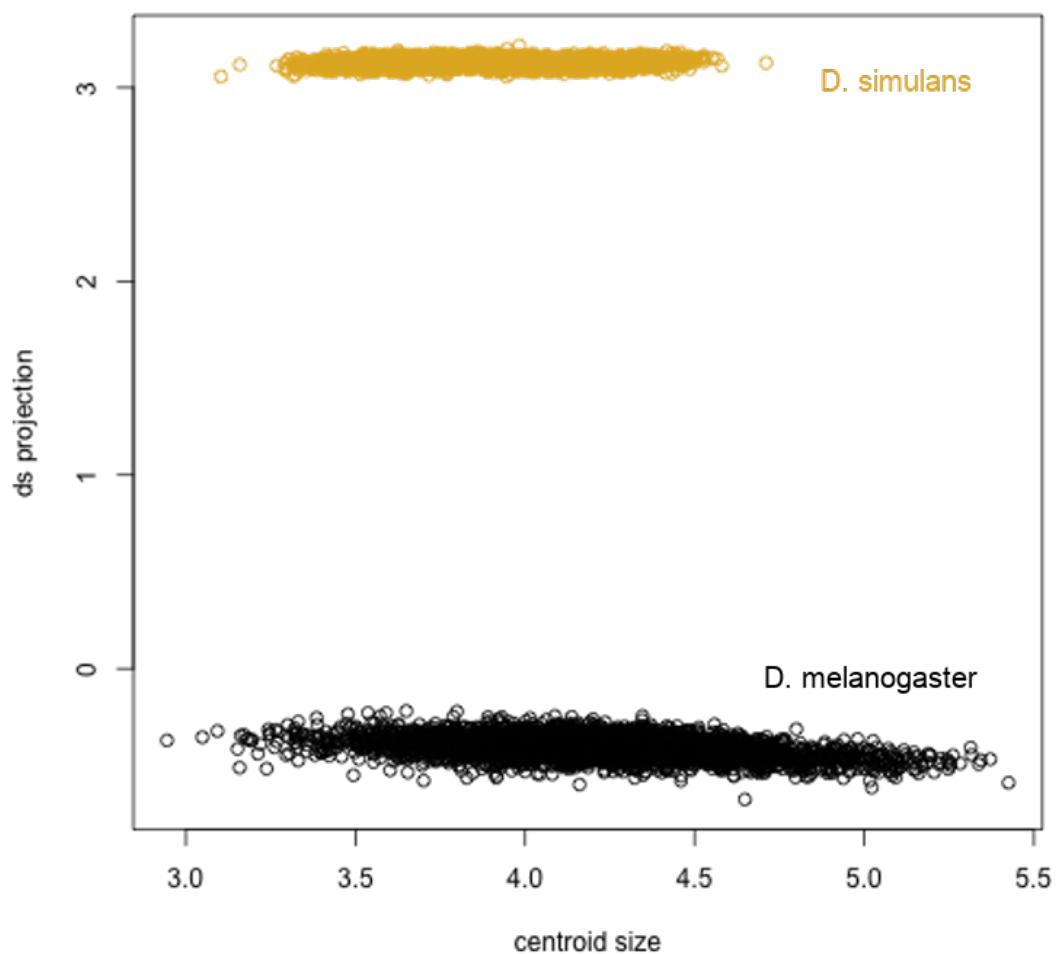
Adams DC, Otárola-Castillo E. Geomorph: an R package for the collection and analysis of geometric morphometric shape data. Methods Ecol Evol. 2013;4(4):393–399. doi:10.1111/2041-210X.12035.

Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics. 2006;22(13):1600–1607. doi:10.1093/bioinformatics/btl140.

Bastide H, Betancourt A, Nolte V, Tobler R, Stöbe P, Futschik A, Schlötterer C. A genome-wide, fine-scale map of natural pigmentation variation in *Drosophila melanogaster*. PLoS Genet. 2013;9(6):e1003534. doi:10.1371/journal.pgen.1003534.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol. 1995;57(1):289–300. doi:10.1111/j.2517-6161.1995.tb02031.x.

Blows MW, McGuigan K. The distribution of genetic variance across phenotypic space and the response to selection. Mol Ecol. 2015;24(9):2056–2072. doi:10.1111/mec.13023.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–2120. doi:10.1093/bioinformatics/btu170.

Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. Cell. 2017;169(7):1177–1186. doi:10.1016/j.cell.2017.05.038.

Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, *et al.* Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science. 2010;327(5963):302–305. doi:10.1126/science.1182213.

Collyer ML, Adams DC. RRPP: an R package for fitting linear models to high-dimensional data using residual randomization. Methods Ecol Evol. 2018;9(7):1772–1779. doi:10.1111/2041-210X.13029.

CONVERGE Consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature. 2015;523(7562):588–591. doi:10.1038/nature14659.

Daub JT, Hofer T, Cutivet E, Dupanloup I, Quintana-Murci L, Robinson-Rechavi M, Excoffier L. Evidence for polygenic adaptation to pathogens in the human genome. Mol Biol Evol. 2013;30(7):1544–1558. doi:10.1093/molbev/mst080.

Dworkin I, Gibson G. Epidermal growth factor receptor and transforming growth factor-β signaling contributes to variation for wing shape in *Drosophila melanogaster*. GENETICS. 2006;173(3):1417–1431. doi:10.1534/genetics.105.053868.

Dworkin I, Palsson A, Gibson G. Replication of an *Egfr*–wing shape association in a wild-caught cohort of *Drosophila melanogaster*. GENETICS. 2005;169(4):2115–2125. doi:10.1534/genetics.104.035766.

Etournay R, Popović M, Merkel M, Nandi A, Blasse C, Aigouy B, Brandl H, Myers G, Salbreux G, Jülicher F, *et al.* Interplay of cell dynamics and epithelial tension during morphogenesis of the *Drosophila* pupal wing. eLife. 2015;4:e07090. doi:10.7554/eLife.07090.

Fatumo S, Carstensen T, Nashiru O, Gurdasani D, Sandhu M, Kaleebu P. Complimentary methods for multivariate genome-wide association study identify new susceptibility genes for blood cell traits. Front Genet. 2019;10:334. doi:10.3389/fgene.2019.00334.

Fritsche LG, Igl W, Bailey JNC, Grassmann F, Sengupta S, Bragg-Gresham JL, Burdon KP, Hebbring SJ, Wen C, Gorski M, *et al.* A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. Nat Genet. 2016;48(2):134–143. doi:10.1038/ng.3448.

Gilchrist AS, Partridge L. The contrasting genetic architecture of wing size and shape in *Drosophila melanogaster*. Heredity (Edinb). 2001;86(2):144–152. doi:10.1046/j.1365-2540.2001.00779.x.

Gompel N, Prud'homme B. The causes of repeated genetic evolution. Dev Biol. 2009;332(1):36–47. doi:10.1016/j.ydbio.2009.04.040.

Gouy A, Daub JT, Excoffier L. Detecting gene subnetworks under selection in biological pathways. Nucleic Acids Res. 2017;45(16):e149. doi:10.1093/nar/gkx626.

Green RM, Fish JL, Young NM, Smith FJ, Roberts B, Dolan K, Choi I, Leach CL, Gordon P, Cheverud JM, *et al.* Developmental nonlinearity drives phenotypic robustness. Nat Commun. 2017;8(1):1970. doi:10.1038/s41467-017-02037-7.

Greene CS, Penrod NM, Williams SM, Moore JH. Failure to replicate a genetic association may provide important clues about genetic architecture. PLoS ONE. 2009;4(6):e5639. doi:10.1371/journal.pone.0005639.

Houle D, Mezey J, Galpern P, Carter A. Automated measurement of *Drosophila* wings. BMC Evol Biol. 2003;3(1):25. doi:10.1186/1471-2148-3-25.

Huang W, Massouras A, Inoue Y, Peiffer J, Ramia M, Tarone AM, Turlapati L, Zichner T, Zhu D, Lyman RF, *et al.* Natural variation in genome architecture among 205 *Drosophila melanogaster* genetic reference panel lines. Genome Res. 2014;24(7):1193–1208. doi:10.1101/gr.171546.113.

Ioannidis JPA, Tarone R, McLaughlin JK. The false-positive to false-negative ratio in epidemiologic studies. Epidemiology. 2011;22(4):450–456. doi:10.1097/EDE.0b013e31821b506e.

Kelly JK, Hughes KA. Pervasive linked selection and intermediate-frequency alleles are implicated in an evolve-and-resequencing experiment of *Drosophila simulans*. GENETICS. 2019;211(3):943–961. doi:10.1534/genetics.118.301824.

King EG, Merkes CM, McNeil CL, Hoofer SR, Sen S, Broman KW, Long AD, Macdonald SJ. Genetic dissection of a model complex trait using the *Drosophila* synthetic population resource. Genome Res. 2012;22(8):1558–1566. doi:10.1101/gr.134031.111.

King EG, Macdonald SJ, Long AD. Properties and power of the *Drosophila* synthetic population resource for the routine dissection of complex traits. Genetics. 2012;191(3):935–949. doi:10.1534/genetics.112.138537.

Klingenberg CP. Methods for studying allometry in geometric morphometrics: a comparison of performance. Evol Ecol. 2022;36(4):439–470. doi:10.1007/s10682-022-10170-z.

Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS ONE. 2011;6(1):e15925. doi:10.1371/journal.pone.0015925.

Kofler R, Pandey RV, Schlötterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics. 2011;27(24):3435–3436. doi:10.1093/bioinformatics/btr589.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–2079. doi:10.1093/bioinformatics/btp352.

Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. 2010;26(5):589–595. doi:10.1093/bioinformatics/btp698.

Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A, Lopez R. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. 2022;50(W1):W276–W279. doi:10.1093/nar/gkac240.

Marigorta UM, Rodríguez JA, Gibson G, Navarro A. Replicability and prediction: lessons and challenges from GWAS. Trends Genet. 2018;34(7):504–517. doi:10.1016/j.tig.2018.03.005.

Márquez EJ. CPR: using Drosophila wing shape data. 2012–2014. Available at: https://www.bio.fsu.edu/ dhoule/wings.html

Marquez EJ, Houle D. Dimensionality and the statistical power of multivariate genome-wide association studies. bioRxiv 016592. doi:10.1101/016592, 17 March 2015, preprint: not peer reviewed.

Marriage TN, King EG, Long AD, Macdonald SJ. Fine-mapping nicotine resistance loci in *Drosophila* using a multiparent advanced generation inter-cross population. GENETICS. 2014;198(1):45–57. doi:10.1534/genetics.114.162107.

Martin A, Orgogozo V. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution. 2013;67(5):1235–1250. doi:10.1111/evo.12081.

McGuigan K. Studying phenotypic evolution using multivariate quantitative genetics. Mol Ecol. 2006;15(4):883–896. doi:10.1111/j.1365-294X.2006.02809.x.

Melo D, Marroig G, Wolf JB. Genomic perspective on multivariate variation, pleiotropy, and evolution. J Hered. 2019;110(4):479–493. doi:10.1093/jhered/esz011.

Palsson A, Dodgson J, Dworkin I, Gibson G. Tests for the replication of an association between *Egfr* and natural variation in *Drosophila melanogaster* wing morphology. BMC Genet. 2005;6(44):44. doi:10.1186/1471-2156-6-44.

Palsson A, Gibson G. Association between nucleotide variation in *Egfr* and wing shape in *Drosophila melanogaster*. GENETICS. 2004;167(3):1187–1198. doi:10.1534/genetics.103.021766.

Pan D. Hippo signaling in organ size control. Genes Dev. 2007;21(8):886–897. doi:10.1101/gad.1536007.

Pan Y, Alégot H, Rauskolb C, Irvine KD. The dynamics of hippo signaling during *Drosophila* wing development. Development. 2018;145(20):dev165712. doi:10.1242/dev.165712.

Pitchers W, Nye J, Márquez EJ, Kowalski A, Dworkin I, Houle D. A multivariate genome-wide association study of wing shape in *Drosophila melanogaster*. GENETICS. 2019;211(4):1429–1447. doi:10.1534/genetics.118.301342.

Porter HF, O'Reilly PF. Multivariate simulation framework reveals performance of multi-trait GWAS methods. Sci Rep. 2017;7(1):38837. doi:10.1038/srep38837.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–842. doi:10.1093/bioinformatics/btq033.

Ray RP, Nakata T, Henningsson P, Bomphrey RJ. Enhanced flight performance by genetic manipulation of wing shape in *Drosophila*. Nat Commun. 2016;7(1):10851. doi:10.1038/ncomms10851.

Reddiex AJ, Chenoweth SF. Integrating genomics and multivariate evolutionary quantitative genetics: a case study of constraints on sexual selection in *Drosophila serrata*. Proc R Soc B Biol Sci. 2021;288(1960):20211785. doi:10.1098/rspb.2021.1785.

Rohlf FJ, Slice D. Extensions of the procrustes method for the optimal superimposition of landmarks. Syst Zool. 1990;39(1):40. doi:10.2307/2992207.

Schluter D. Adaptive radiation along genetic lines of least resistance. Evolution. 1996;50(5):1766. doi:10.2307/2410734.

Shriner D. Moving toward system genetics through multiple trait analysis in genome-wide association studies. Front Genet. 2012;3(1):1. doi:10.3389/fgene.2012.00001.

Spitzer K, Pelizzola M, Futschik A. Modifying the Chi-square and the CMH test for population genetic inference: adapting to overdispersion. Ann Appl Stat. 2020;14:202–220. doi:10.1214/19-AOAS1301.

Steiner CC, Weber JN, Hoekstra HE. Adaptive variation in beach mice produced by two interacting pigmentation genes. PLoS Biol. 2007;5(9):e219. doi:10.1371/journal.pbio.0050219.

Stephens M. A unified framework for association analysis with multiple related phenotypes. PLoS ONE. 2013;8(7):e65245. doi:10.1371/journal.pone.0065245.

Stern DL, Orgogozo V. The loci of evolution: how predictable is genetic evolution? Evolution. 2008;62(9):2155–2177. doi:10.1111/j.1558-5646.2008.00450.x.

Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019;20(8):467–484. doi:10.1038/s41576-019-0127-1.

Van der Linde K. Wings: automated capture of Drosophila wing shape. Tallahassee, FL. 2004–2014. Available at: https://www.bio.fsu.edu/ dhoule/wings.html

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 Years of GWAS discovery: biology, function, and translation. Am J Hum Genet. 2017;101(1):5–22. doi:10.1016/j.ajhg.2017.06.005.

Wray NR, Wijmenga C, Sullivan PF, Yang J, Visscher PM. Common disease is more complex than implied by the core gene omnigenic model. Cell. 2018;173(7):1573–1580. doi:10.1016/j.cell.2018.05.051.

Wright S. Physiological and evolutionary theories of dominance. Am Nat. 1934;68(714):24–53. doi:10.1086/280521.
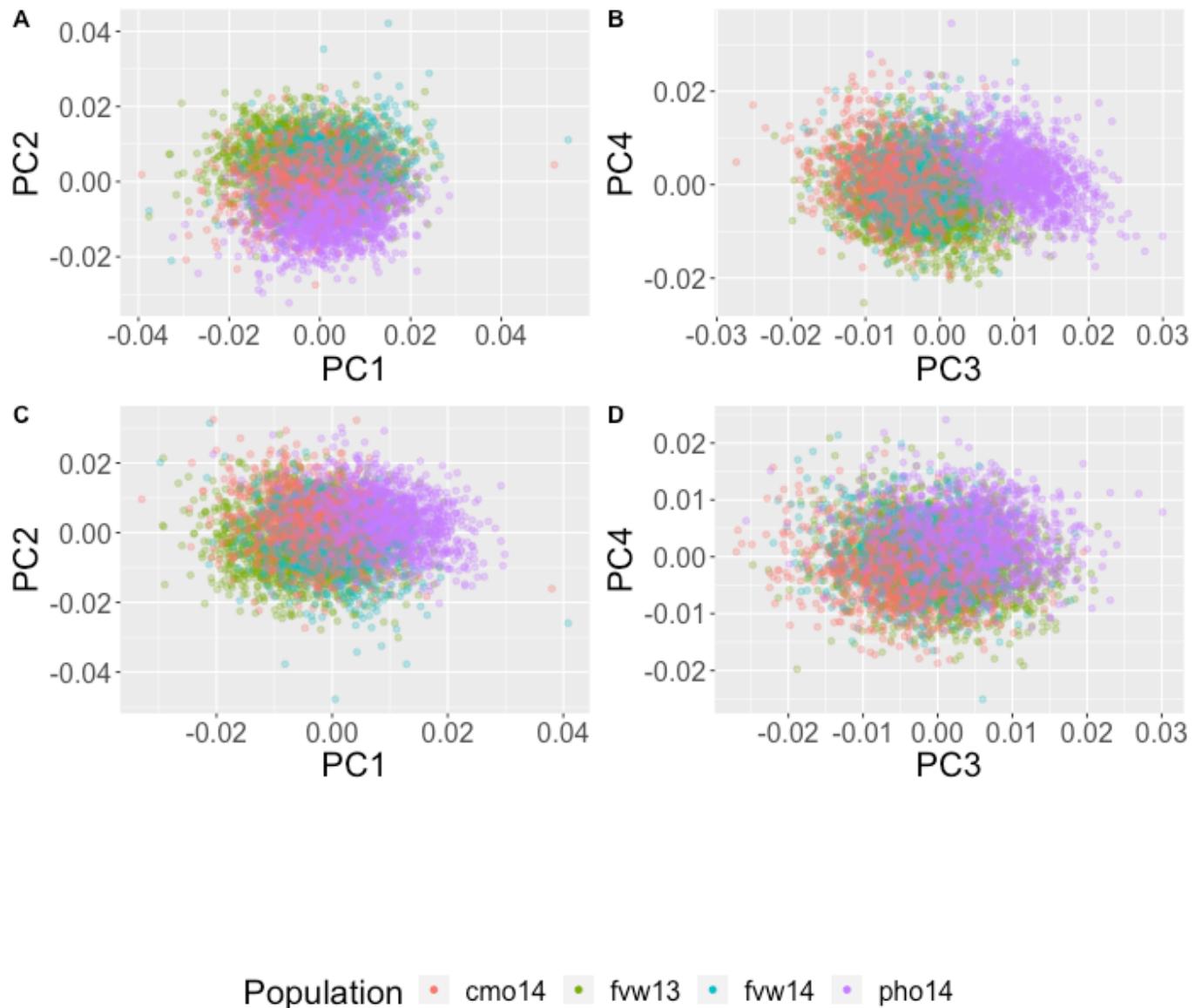
Editor: C. Peichel

Supplemental Figure 1. Allele Frequency spectra demonstrate more variation at *ds* compared with *emc* in the synthetic outbred population used for artificial selection. Figures show estimated alternate allele frequencies at *ds* and *emc*. Alternate allele frequencies are estimated using parental strain genotype data and assuming an equal contribution from each parent to the founding population. Note the different y axis scale between *ds* and *emc*.

Supplemental Figure 2. Projection of FVW14 wings onto *ds* shape change vector shows clear distinction between *D. melanogaster* and *D. simulans.* Clear separation between the FVW14 samples (black) and *D. simulans* data (gold) indicates that *D. melanogaster* females were accurately identified.
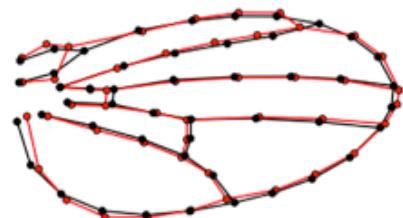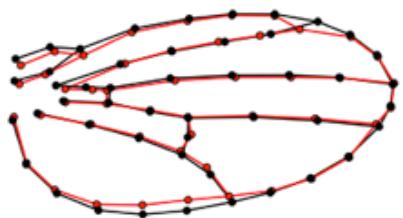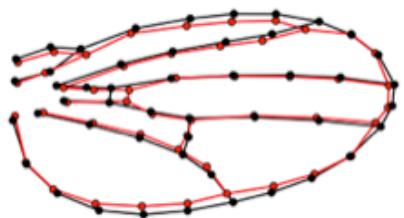
Supplemental Figure 3. Principal component analysis of shape variation within and among populations for wild collected *Drosophila*. First four axes from the PCA for shape variation are shown. (A) and (B) PCA includes all shape variation. (C) and (D) use the 'allometry corrected' landmarks, (residuals from a model regressing landmarks onto centroid size of wings).
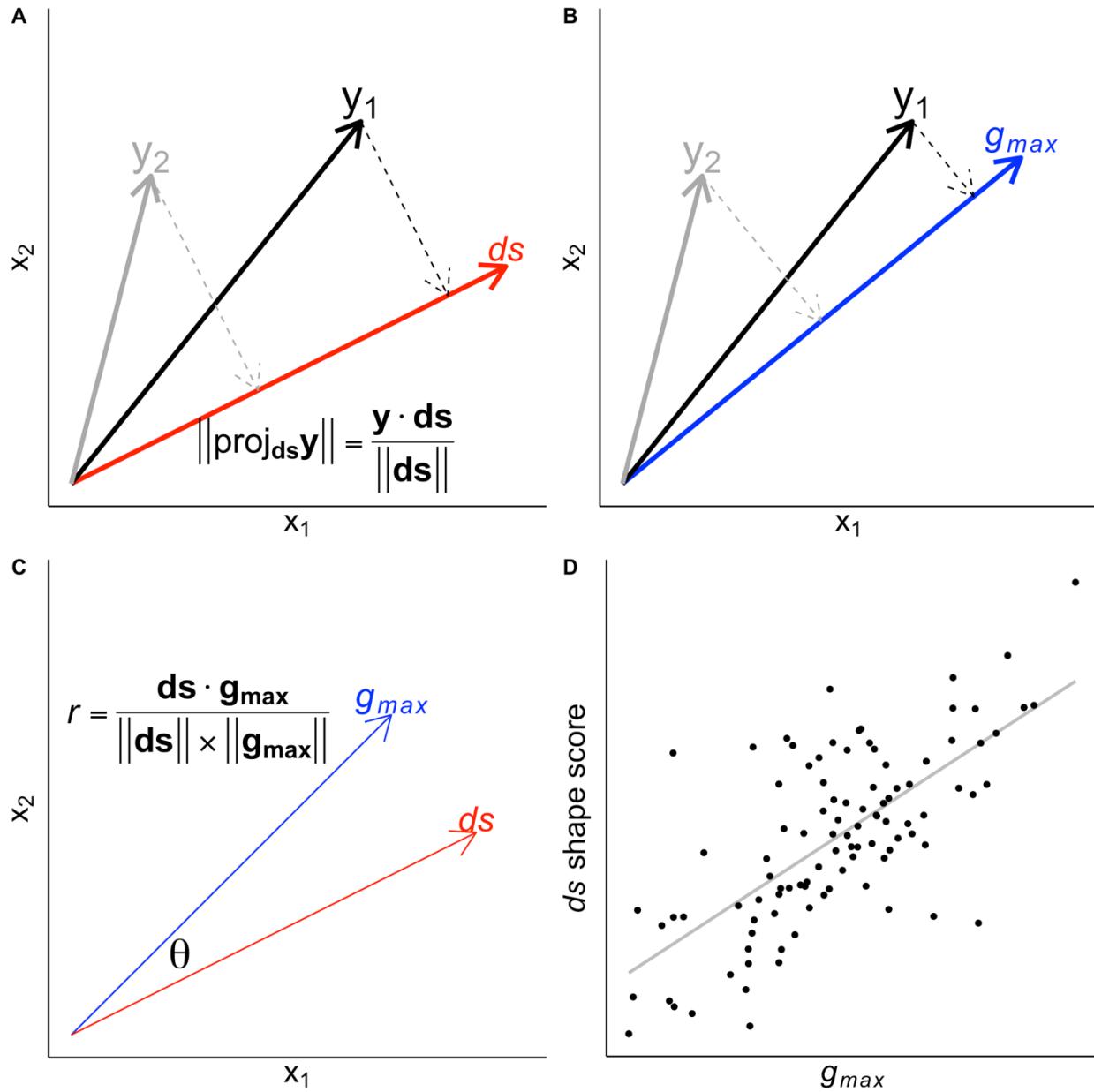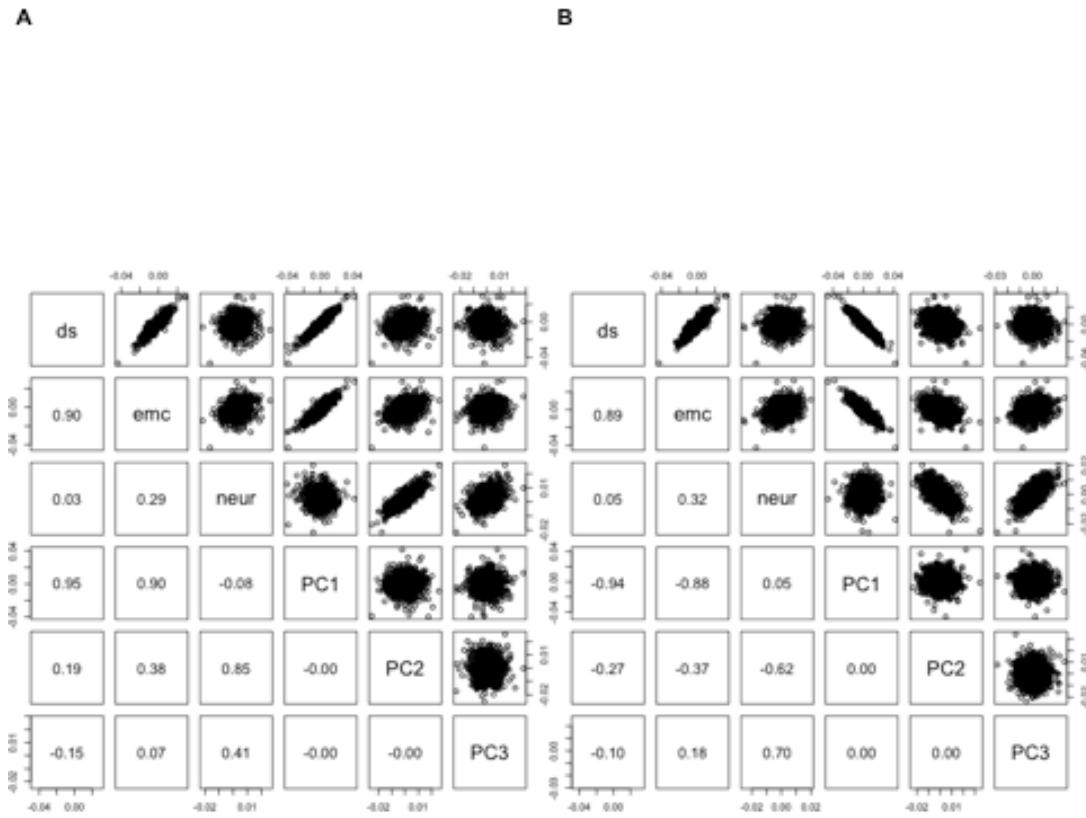
ds 1x   **emc 10x**   **neur 2x**
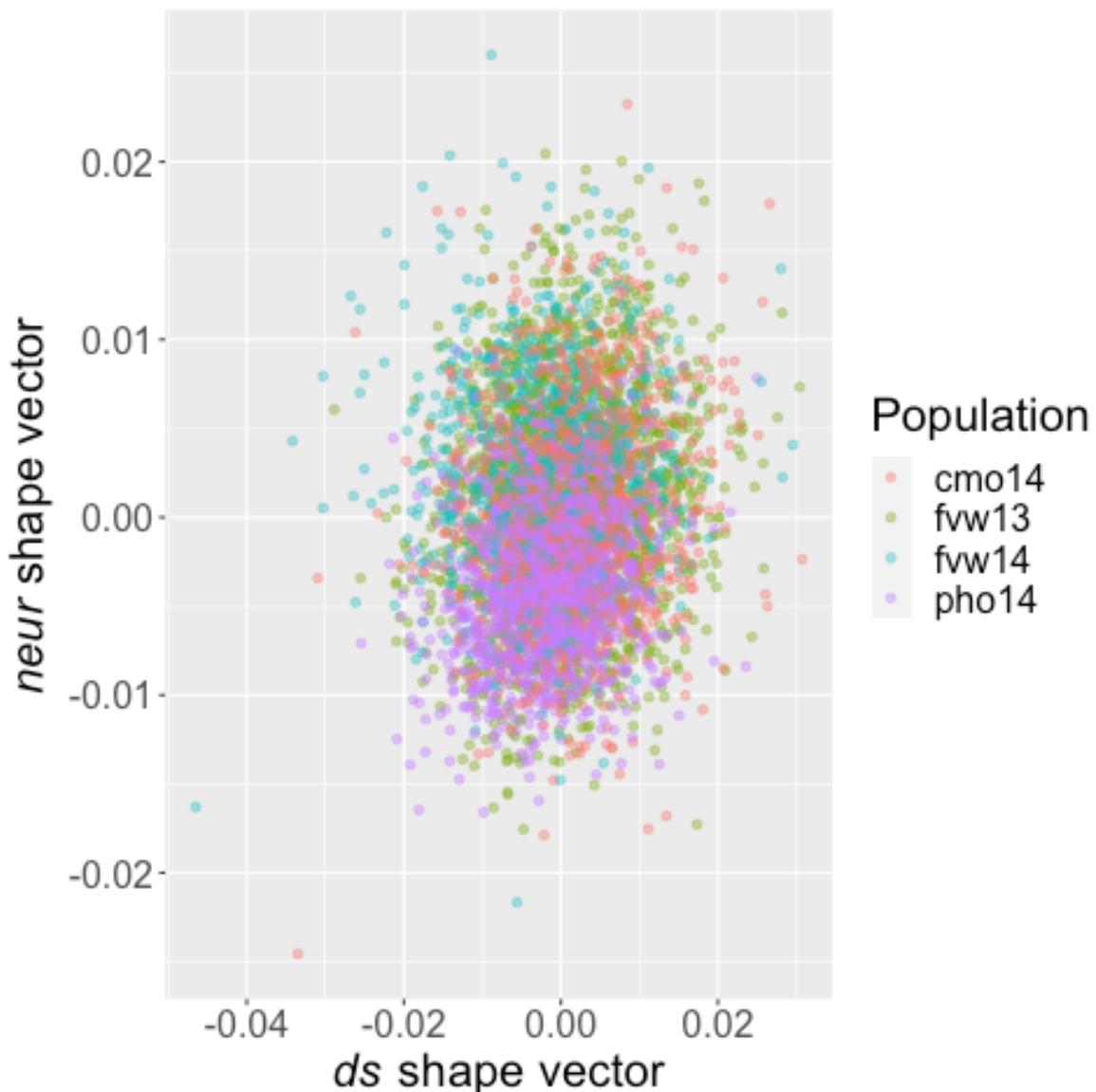
Supplemental Figure 4. Shape change effects due to RNAi knockdown of *ds*, *emc* and *neur.* Scaling of effects provided to aid in visualization of shape changes. Different magnifications are provided to account for the disparate magnitudes for estimated shape change vectors: *ds* = 5.5, *emc* = 0.44, *neur* = 2.8. The vectors from these analyses were used for projections in this study.
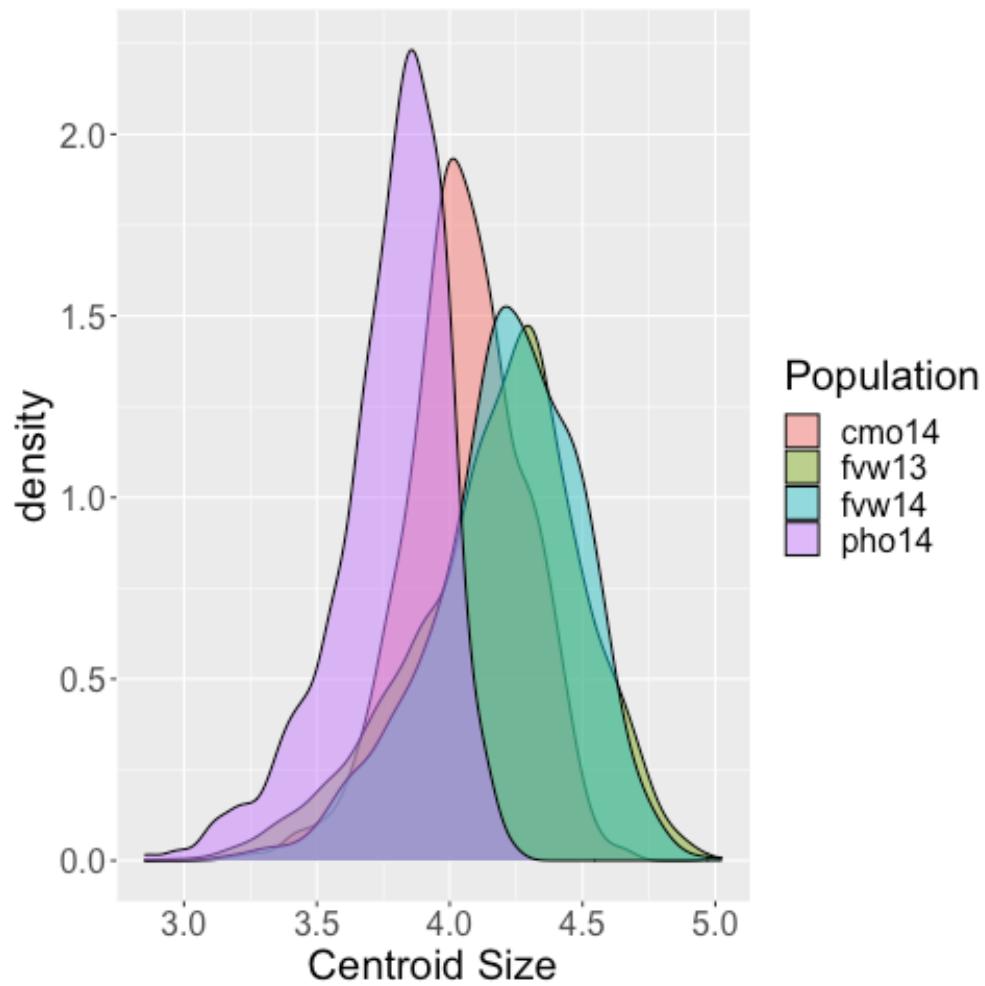
Supplemental Figure 5. Illustration of projections onto shape vectors to generate shape scores used in this study. The *ds* shape change vector is used for demonstration. (A) Calculating shape score using projection. y_1 and y_2 represent vectors of landmarks for two representative individuals. Dotted arrows represent projection of y_1 and y_2 onto the shape change vector defined by *ds* RNAi to generate the *ds* shape score. (B) Similarly, y_1 and y_2 are projected onto the g_{max} (PC1) vector, representing the direction of maximum genetic variation in the genetic variance-covariance matrix, G. (C) Comparing g_{max} and shape change vectors, using the correlation (r) of vectors directly, or via the angle θ, between vectors. (D) Hypothetical relationship between *ds* shape scores and g_{max}, indicating a relationship between directions of *ds* induced shape change and the direction of genetic variation in shape variation.

A

B
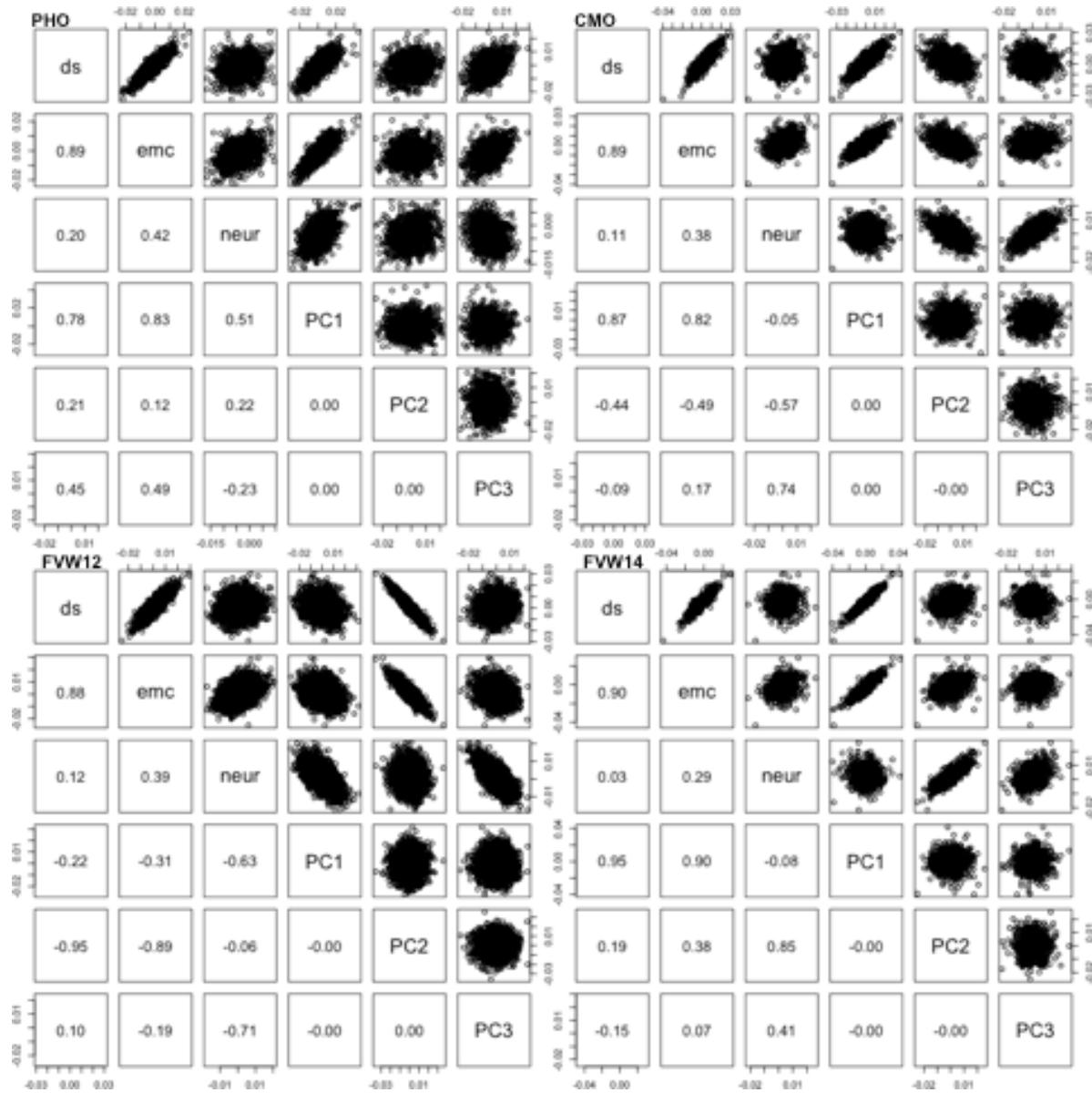
Supplemental Figure 6. Including wild-caught females in the analyses does not change interpretation for FVW14 data. Correlation between projection of shape data from FVW14 population onto *ds, emc* and *neur* shape change vectors and the first three PCs are calculated from shape data from all wings in the FVW14 population. (A) male only data, (B) females and male data. Inclusion of female data does not change the conclusions drawn from the relationships between directions shape change vectors and PCs. Note that the "flipped" directions of the PCs between (A) and (B) represents the arbitrary sign of the set of eigenvectors for the PCA (i.e. the set of eigenvectors can equivalently be multiplied by -1).

Supplemental Figure 7. Shape variation in field collected samples by projecting individual shape data onto *ds* and *neur* shape change vectors. Projections were performed used size "adjusted" landmark data onto *ds* and *neur* shape change vectors. The correlation between the ds and neur shape scores is r = 0.12. In linear models with shape scores (*ds* and *neur* respectively) regressed onto population and wing size, the partial R^2 for population effects is 0.040 (*ds*) and 0.18 (*neur*).
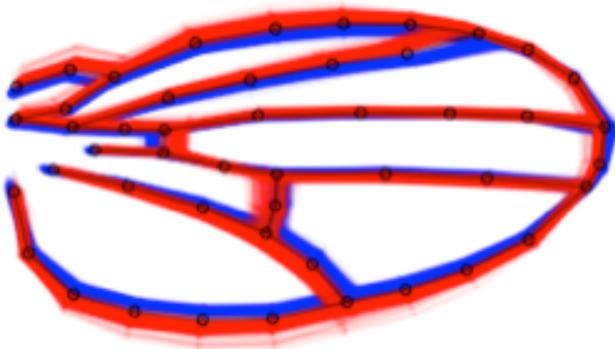
Supplemental Figure 8. Distribution of wing sizes (males only) in wild caught cohorts. When size is linearly regressed onto population, model $R^2 = 0.27$.
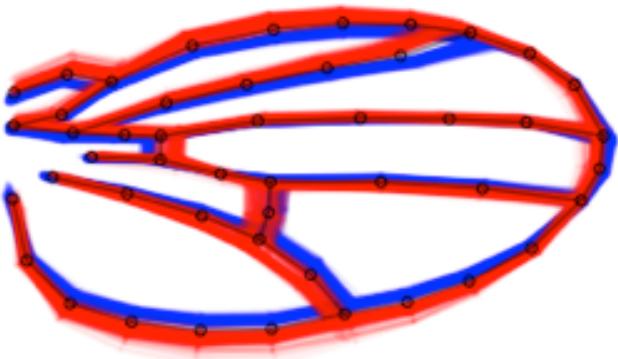
Supplemental Figure 9. **Projections of data onto RNAi shape change vectors are correlated with major axes of shape variation in wild-caught *Drosophila* from each population**. Correlation between projection of shape data from wild cohorts onto *ds, emc* and *neur* shape change vectors and the first three PCs are calculated from shape data from all wings in each cohort independently. Note that the "flipped" directions of the PCs for FVW12 represents the arbitrary sign of the set of eigenvectors for the PCA (i.e. the set of eigenvectors can equivalently be multiplied by -1). Note that the eigenvectors representing PC1 and PC2 in FVW12 have "swapped" because of similarities in variance accounted for.
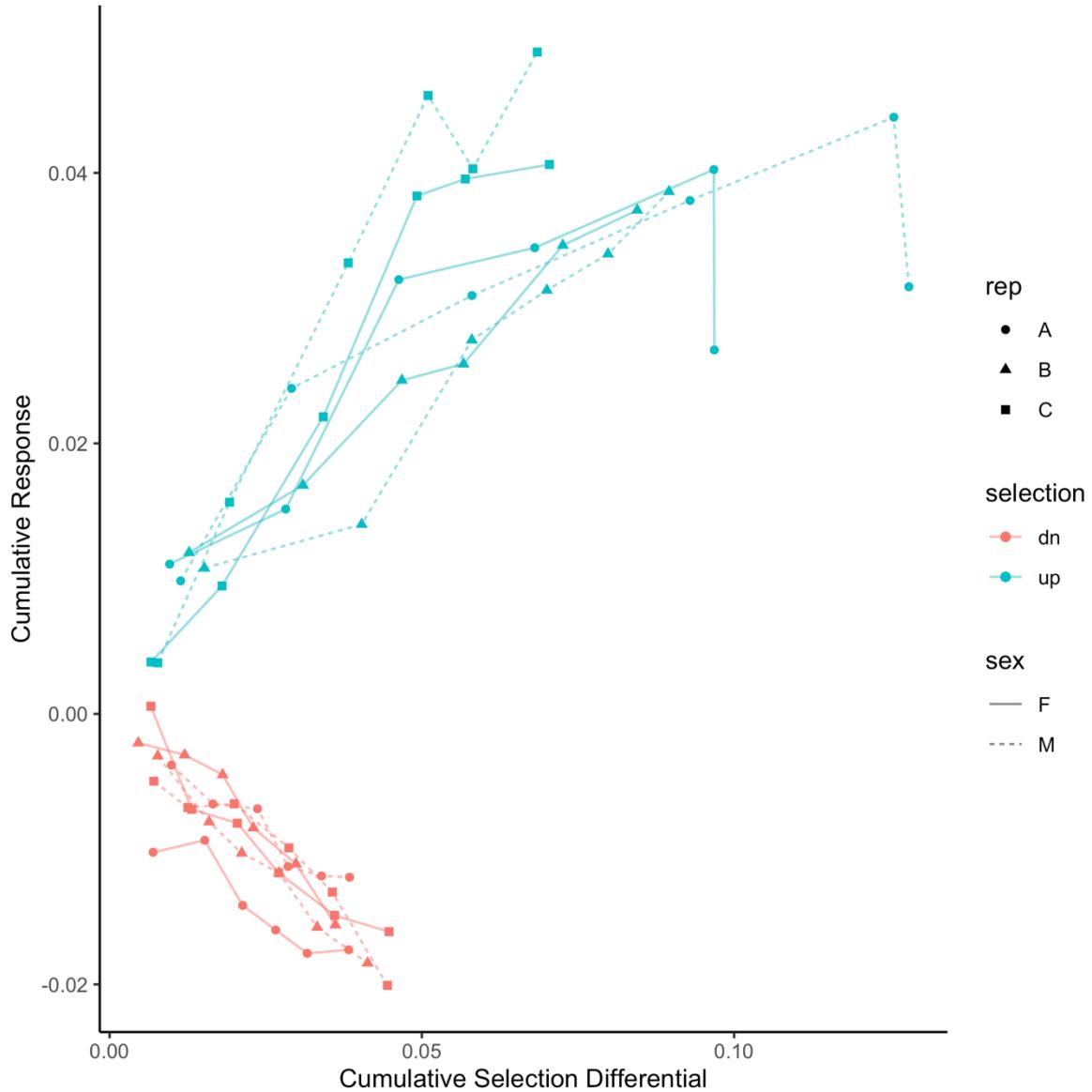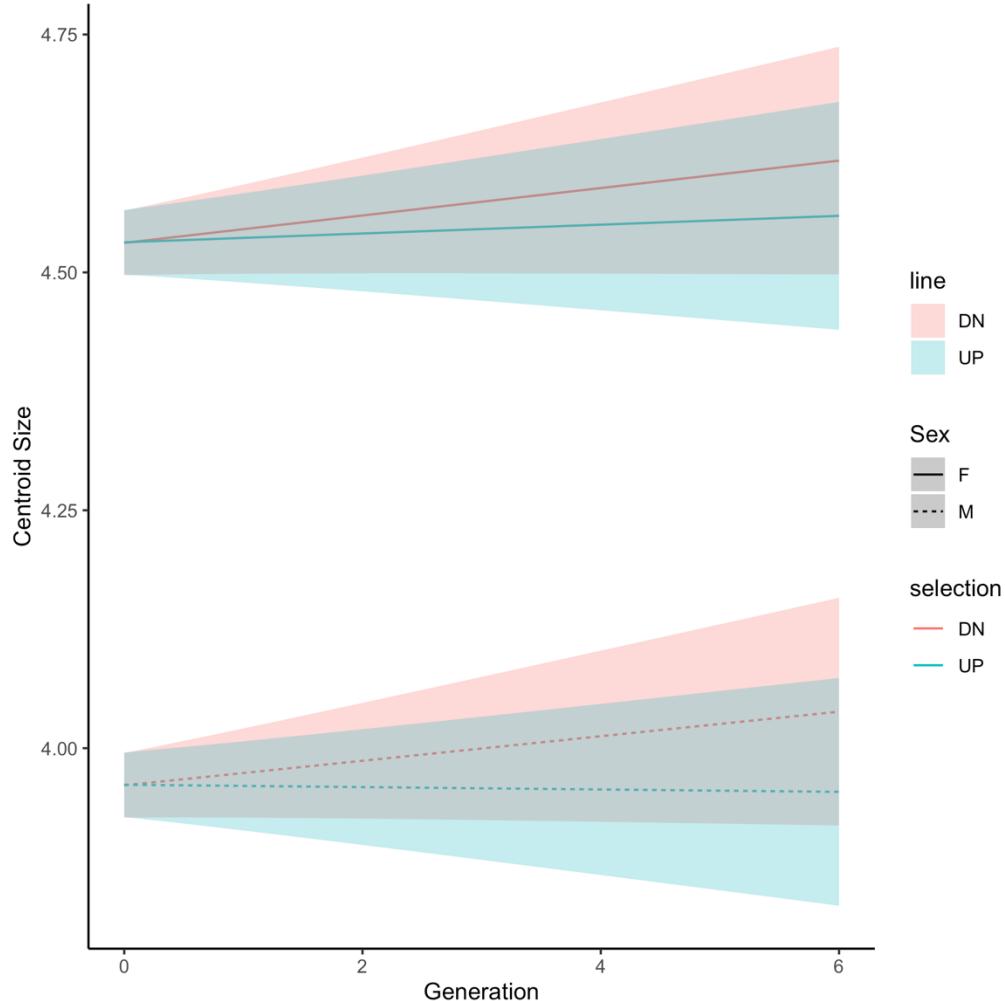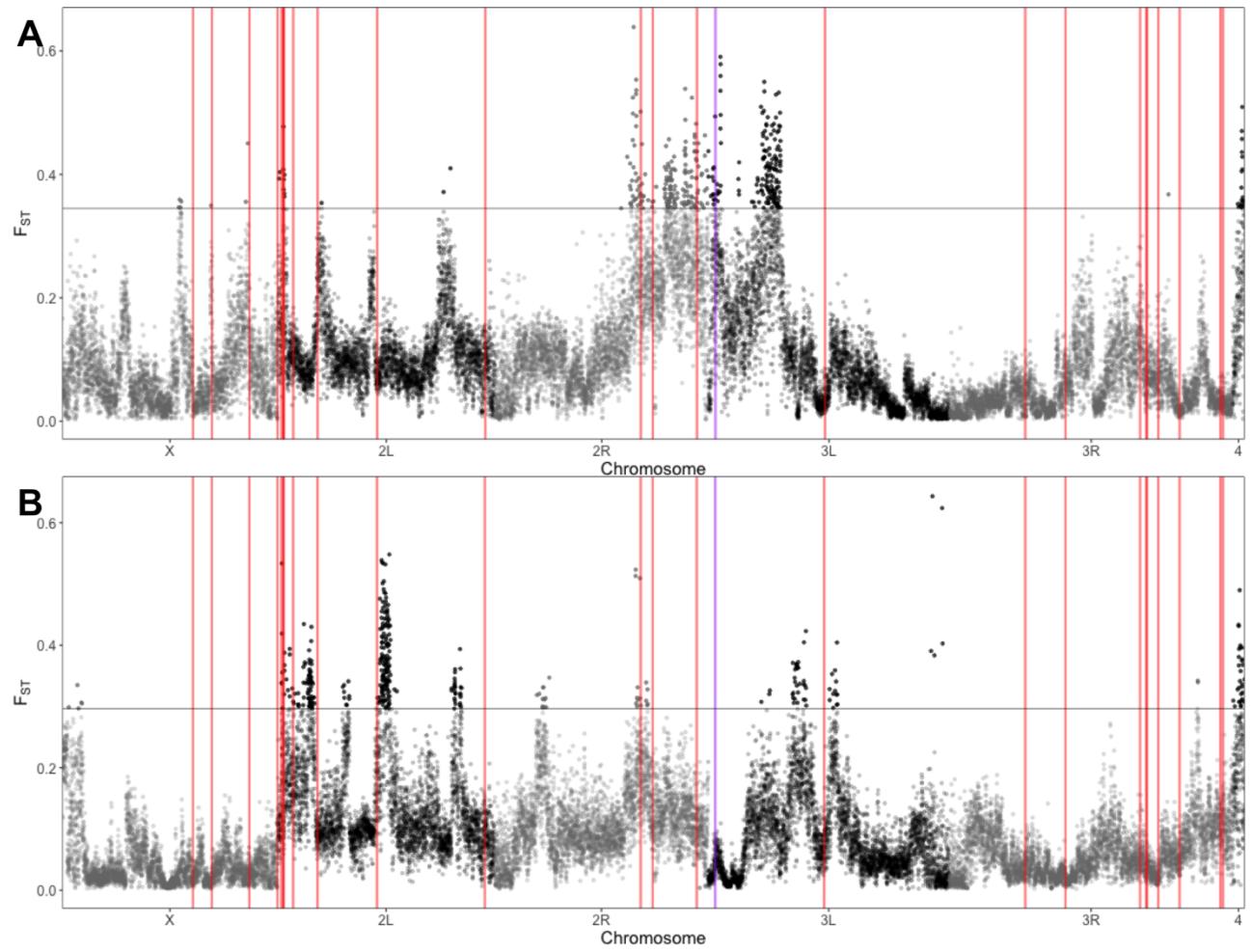
Supplemental Figure 10. Variation in wing shape among individuals from artificial selection along the *ds* shape change vector. All wings from generation 7 of up (red) and down (blue) selection lineages plotted, with females in (A) and males in (B). Black wireframe is mean shape from the experiment.

Supplemental Figure 11. Response to artificial selection along the *ds* shape change vector. Regression of cumulative selection differential onto cumulative response was used to estimate realized heritability in both treatments ("up" & "down") independently.
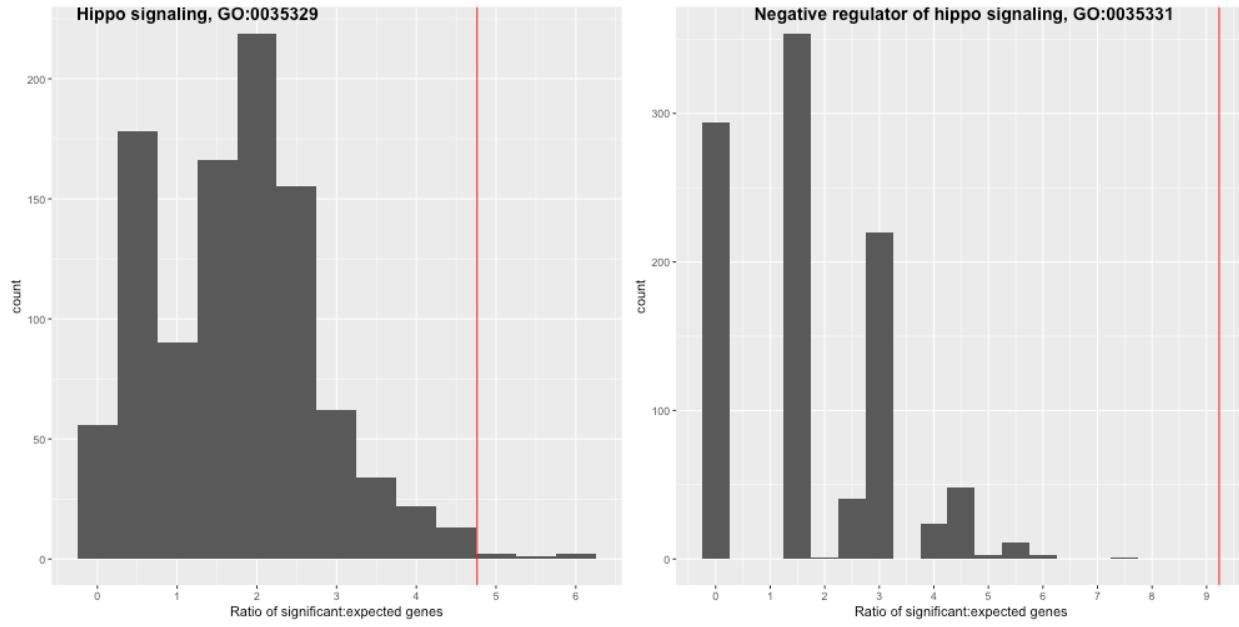
Supplemental Figure 12. Modest (and not significant) changes in wing size following artificial selection based on *ds* shape change. Mean size estimated for up and down *ds* selection lineages estimated from a linear mixed model with replicate lineage fitted as a random effect. Shaded regions represent 95% confidence bands for the correlated response to selection.
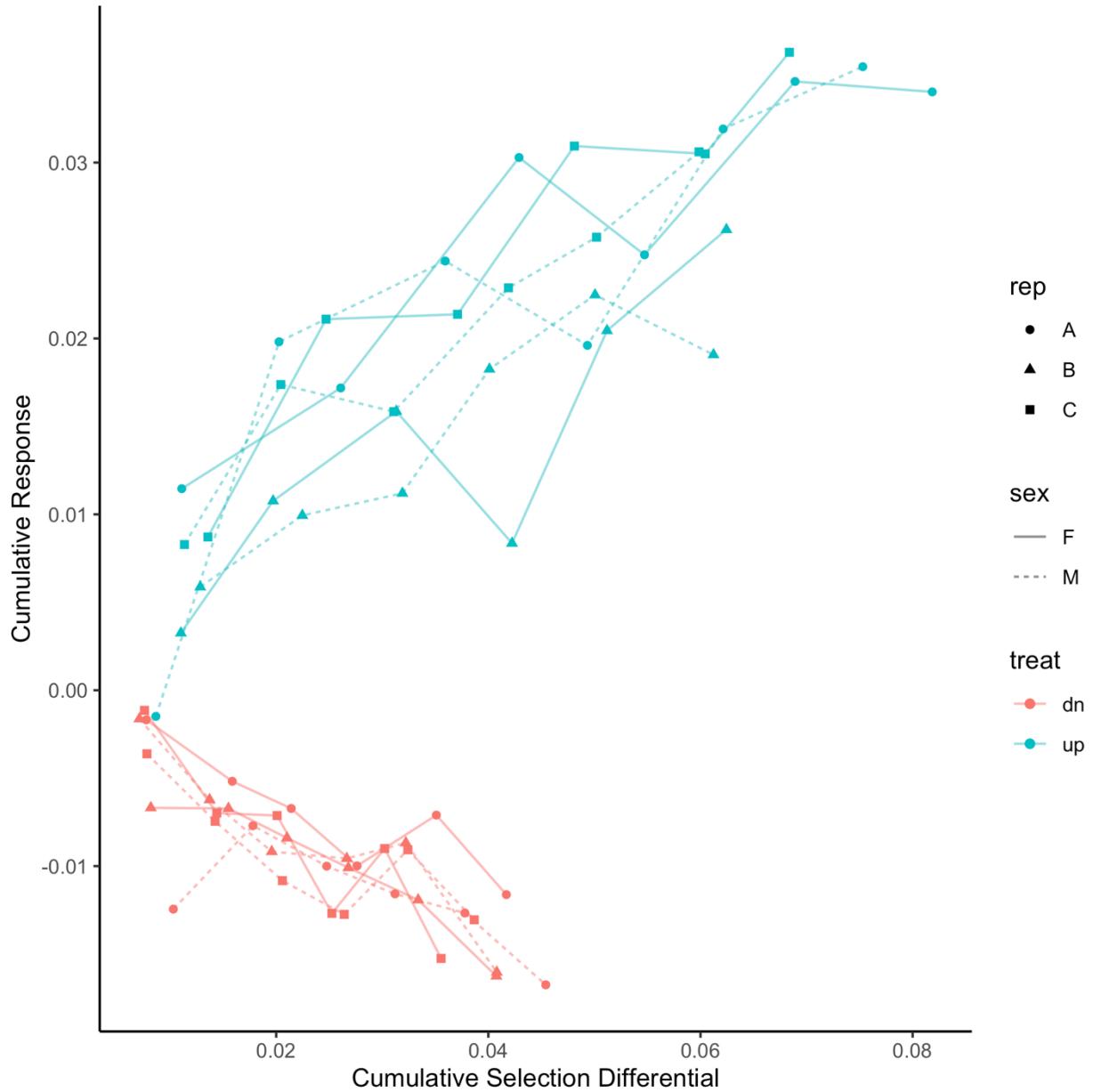
Supplemental Figure 13. Genetic differentiation between artificial selection pools (Figure 3B), with core hippo signaling loci (red) and *emc* (purple) marked. Genomic differentiation between up and down selection lineages (F_{ST}) measured in 5000bp windows for the artificial selection along *ds* (A) and *emc* shape change vectors (B). Horizonal grey lines represents 3sd from mean F_{ST}.

```
195    GGATTATATGGGGTTTACTAGAAGTGTATATGACCCATCTTAGTTGGCATATTTGTGCTT    108
28     GGATNATATGGGGTTTACTAGAAGTGTATATGACCCATCTTAGTTGGCATATTTGTGCTT    106
96     GGATNATATGGGGTTTACTAGAAGTGTATATGACCCATCTTAGTTGGCATATTTGTGCTT    106
48     GGATNATATGGGGTTTACTAGAAGTGTATATGACCCATCTTAGTTGGCATATTTGTGCTT    106
59     GGATTATATGGGGTTTACTAGAAGTGTATATGACCCATCTTAGTTGGCATATTTGTGCTT    109
801    GGATTATATGGGGTTTACTAGAAGTGTATATGACCCATCTTAGTTGGCATATTTGTGCTT    108
ref    ggattatatggggtgtacaagaagtgtatatgactcatcttag-----------------    163
129    GGATTATATGGGGTGTACTAGAAGTGTATATGACTTATCTTAG-----------------    102
301    GGATTATATGGGGTGTACTAGAAGTGTATATGACTTATCTTAG-----------------    100
69     GGATTATATGGGGTGTACTAGAAGTGTATATGACTTATCTTAG-----------------    102
385    GGATTATATGGGGTGTACTAGAAGTGTATATGACTTATCTTAG-----------------    100
75     GGATTATATGGGGTGTACTAGAAGTGTATATGACTCATCTTAG-----------------    101
83     GGATTATATGGGGTGTACTAGAAGTGTATATGACTTATCTTAG-----------------    100
491    GGATTATATGGGGTGTACTAGAAGTGTATATGACTCATCTTAG-----------------    101
34     GGATTATATGGGGTGTACAAGAAGTGTATATGACTCATCTTAG-----------------    99
774    GGATTATATGGGGTGTACAAGAAGTGTATATGACTCATCTTAG-----------------    102
       ****  *********  ***  ***************   *******

195    AGGTGCATATTTCTAACAAAAATCCGTAACTATTTTGAGTGGCCAAGAAAATTTCAAATT    168
28     AGGTGCATATTTCTAACAAAAATCCGTAACTATTTTGAGTGGCCAAGAAAATTTCAAATT    166
96     AGGTGCATATTTCTAACAAAAATCCGTAACTATTTTGAGTGGCCAAGAAAATTTCAAATT    166
48     AGGTGCATATTTCTAACAAAAATCCGTAACTATTTTGAGTGGCCAAGAAAATTTCAAATT    166
59     AGGTGCATATTTCTAACAAAAATCCGTAACTATTTTGAGTGGCCAAGAAAATTTCAAATT    169
801    AGGTGCATATTTCTAACAAAAATCCGTAACTATTTTGAGTGGCCAAGAAAATTTCAAATT    168
ref    -ttggcatatttctaacaaaaatccgtaactattttgagtgtccaagaaaatttcaaa-t    221
129    -TTGGCATATTTCTAACAAAAATCCGTAACTATTTTGAGTGGTAGAAA-ATTTCAAAT-T    159
301    -TTGGCATATTTCTAACAAAAATCCGTAACTATTTTGAGTGGTAGAAA-ATTTCAAAT-T    157
69     -TTGGCATATTTCTAACAAAAATCCGTAACTATTTTGAGTGGTAGAAA-ATTTCAAAT-T    159
385    -TTGGCATATTTCTAACAAAAATCCGTAACTATTTTGAGTGGTAGAAA-ATTTCAATT-T    157
75     -TTGGCATATTTCTAACAAAAATCCGTAACTATTTTGAGTGGTAGAAA-ATTTCAAAT-T    158
83     -TTGGCATATTTCTAACAAAAATCCGTAACTATTTTGAGTGGTAGAAA-ATTTCAAAT-T    157
491    -TTGGCATATTTCTAACAAAAATCCGTAACTATTTTGAGTGGTAGAAA-ATTTCAAAT-T    158
34     -TTGGCATATTTCTAACAAAAATCCGTAACTATTTTGAGTGTCCAAGAAAATTTCAAA-T    157
774    -TTGGCATATTTCTAACAAAAATCCGTAACTATTTTGAGTGTCCAAGAAAATTTCAAA-T    160
         *************************************   *  *  * **   *    *

195    TTTTTTTGCGTGTGCTTGGATGTTGTAATTTTGCATTGTTTGGCTAATCTACGTAAATCT    228
28     TTTTTTTGCGTGTGCTTGGATGTTGTAATTTTGCATTGTTTGGCTAATCTACGTAAATCT    226
96     TTTTTTTGCGTGTGCTTGGATGTTGTAATTTTGCATTGTTTGGCTAATCTACGTAAATCT    226
48     TTTTTTTGCGTGTGCTTGGATGTTGTAATTTTGCATTGTTTGGCTAATCTACGTAAATCT    226
59     TTTTTTTGCGTGTGCTTGGATGTTGTAATTTTGCATTGTTTGGCTAATCTACGTAAATCT    229
801    TTTTTTTGCGTGTGCTTGGATGTTGTAATTTTGCATTGTTTGGCTAATCTACGTAAATCT    228
ref    tttttttgcgtgtgcttggatgttgtaatttcgcattgtttggctaatctacgtaaatct    281
129    TTTTTTTGCGTGTGCTTGGATGTTGTAATTTCGCATTGTTTGGCTAATCTACGTAAATCT    219
301    TTTTTTTGCGTGTGCTTGGATGTTGTAATTTCGCATTGTTTGGCTAATCTACGTAAATCT    217
69     TTTTTTTGCGTGTGCTTGGATGTTGTAATTTCGCATTGTTTGGCTAATCTACGTAAATCT    219
385    TTTTTTTGCGTGTGCTTGGATGTTGTAATTTCGCATTGTTTGGCTAATCTACGTAAATCT    217
75     TTTTTTTGCGTGTGCTTGGATGTTGTAATTTCGCATTGTTTGGCTAATCTACGTAAATCT    218
83     TTTTTTTGCGTGTGCTTGGATGTTGTAATTTCGCATTGTTTGGCTAATCTACGTAAATCT    217
491    TTTTTTTGCGTGTGCTTGGATGTTGTAATTTCGCATTGTTTGGCTAATCTACGTAAATCT    218
34     TTTTTTTGCGTGTGCTTGGATGTTGTAATTTCGCATTGTTTGGCTAATCTACGTAAATCT    217
774    TTTTTTTGCGTGTGCTTGGATGTTGTAATTTCGCATTGTTTGGCTAATCTACGTAAATCT    220
       *******************************  ***************************
```
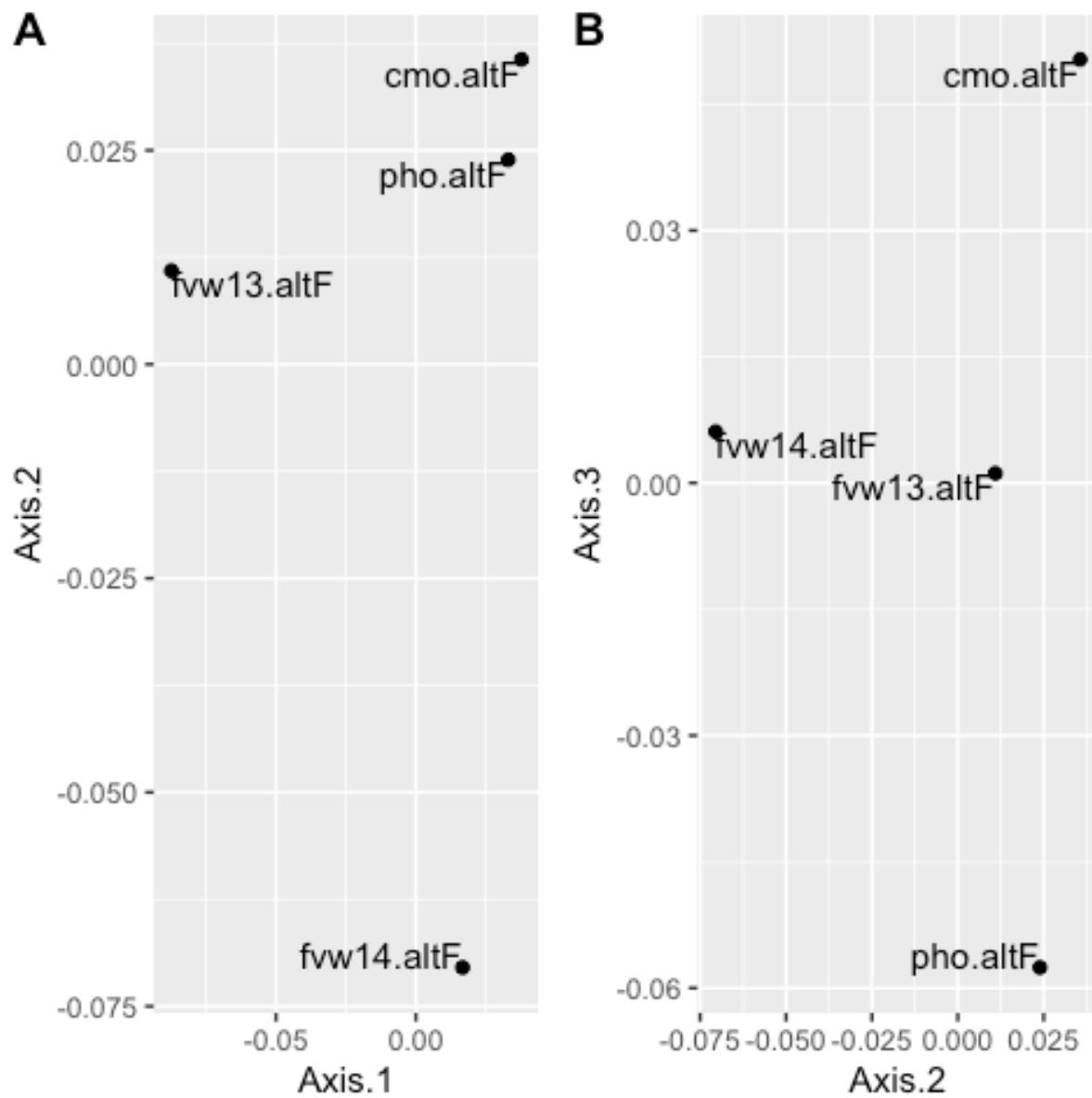
Supplemental Figure 14. Alignment of sanger sequencing of the region containing the *ds* polymorphism from several DGRP lines (line numbers indicated at line starts) to reference sequence (from *Drosophila melanogaster* genome). DGRP lines are indicated with DGRP 195, 28, 96, 48, 59 and 801 predicted to have the insertion, and, DGRP 129, 301, 69, 385, 75, 83, 491, 34 and 774 without. In addition to the insertion in these lines, several other associated SNPs are found linked to the insertion.
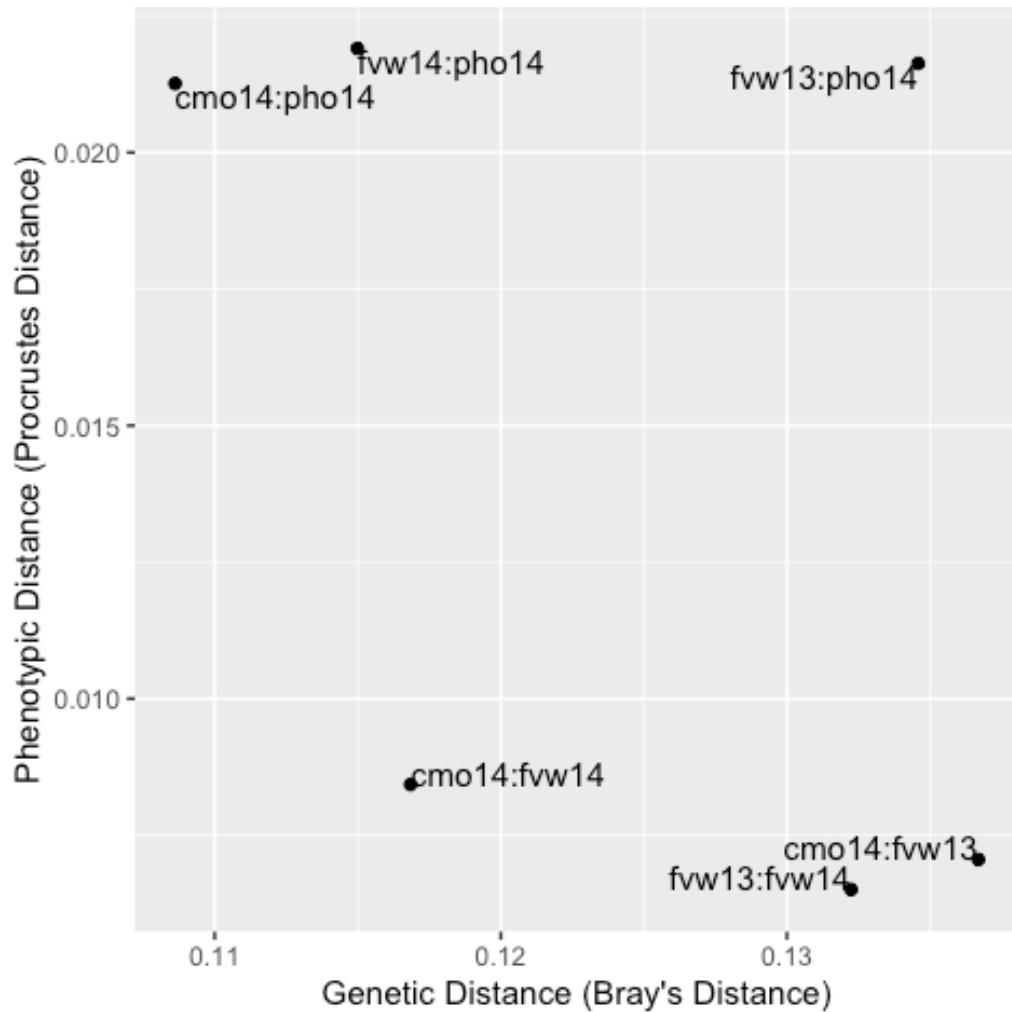
Supplemental Figure 15: Permutation test for over representation of hippo signaling terms in outlier regions. The red line represents the observed ratio of significant to expected genes in outlier windows (F_{ST} greater than 3 standard deviations from the mean). The permutation test selected random windows from the genome, equal in number to those identified as outliers. For each random draw of loci, the ratio of significant to expected genes in the term of interest was calculated for 1000 permutations.

Supplemental Figure 16. Response to selection based on projections onto *emc* shape change vector. Regression of cumulative selection differential onto cumulative response was used to estimate realized heritability in both "up" and "down" selection regimes independently.
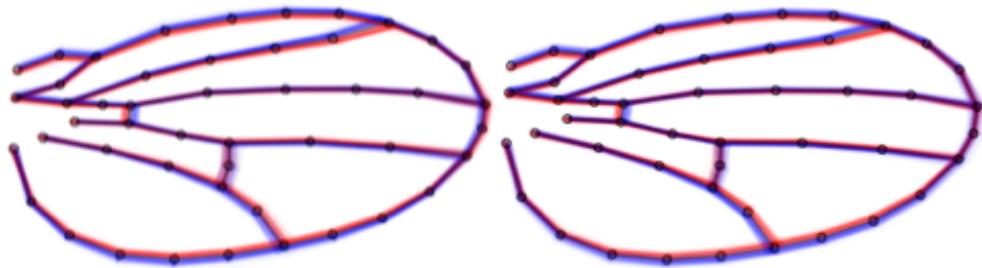
Supplemental Figure 17. Minimal genetic structure among wild populations used in this study. Principal coordinate plot calculated from Bray's distances estimated from allele frequency data (altF), between the four wild cohorts used in this analysis. Axis 1-3 explains 45%, 30% and 25% of the variance in genetic distances respectively.
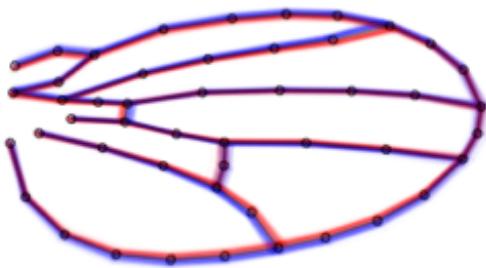
Supplemental Figure 18. Genetic distance is not correlated with phenotypic distance among wild cohorts used in this study. Phenotypic distances are the Euclidian distances from model estimated shape vectors for each population from a multivariate regression of shape onto for population and wing centroid size.
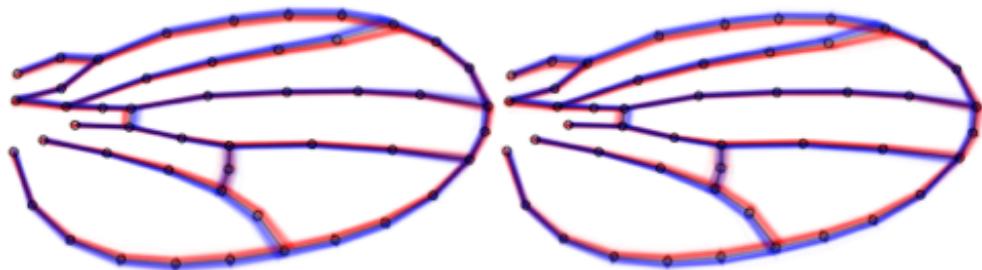
CMO   PHO
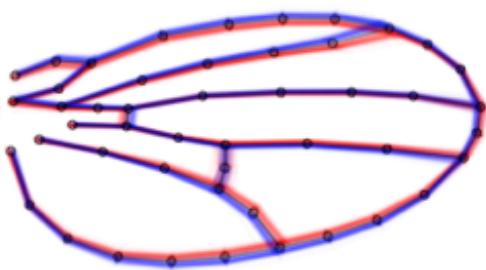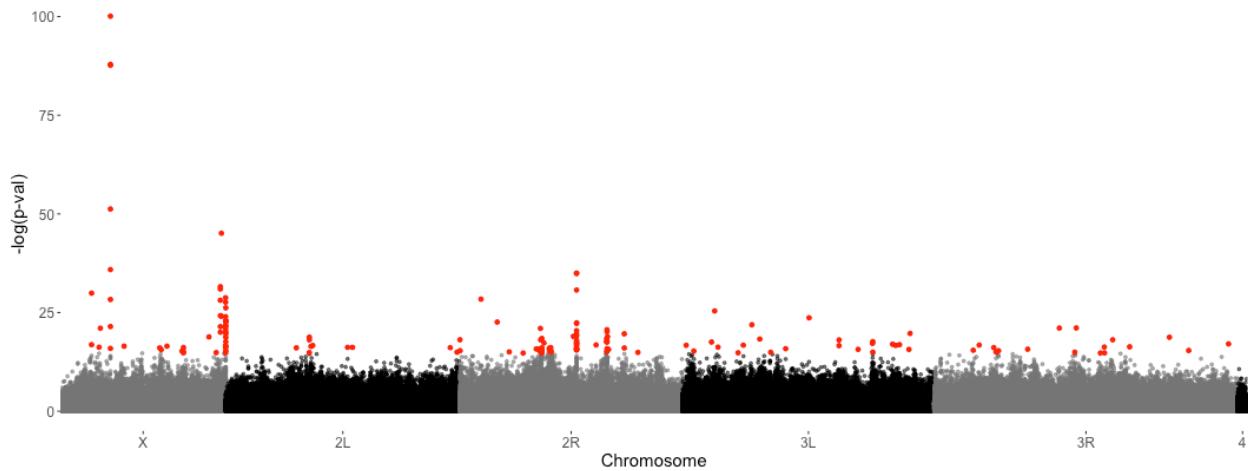
FVW13   FVW14
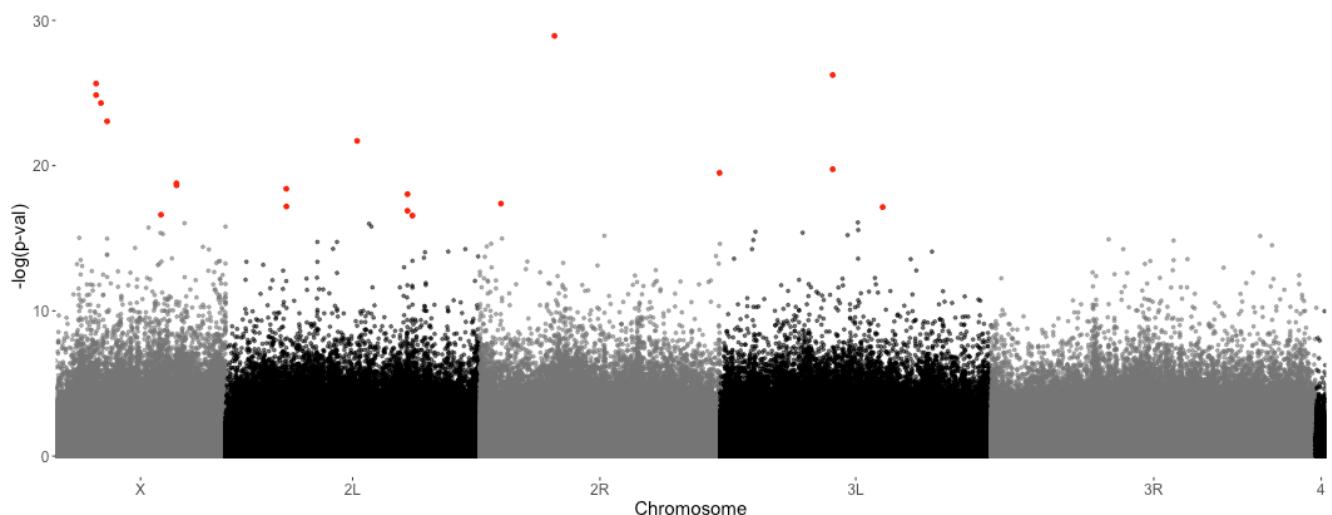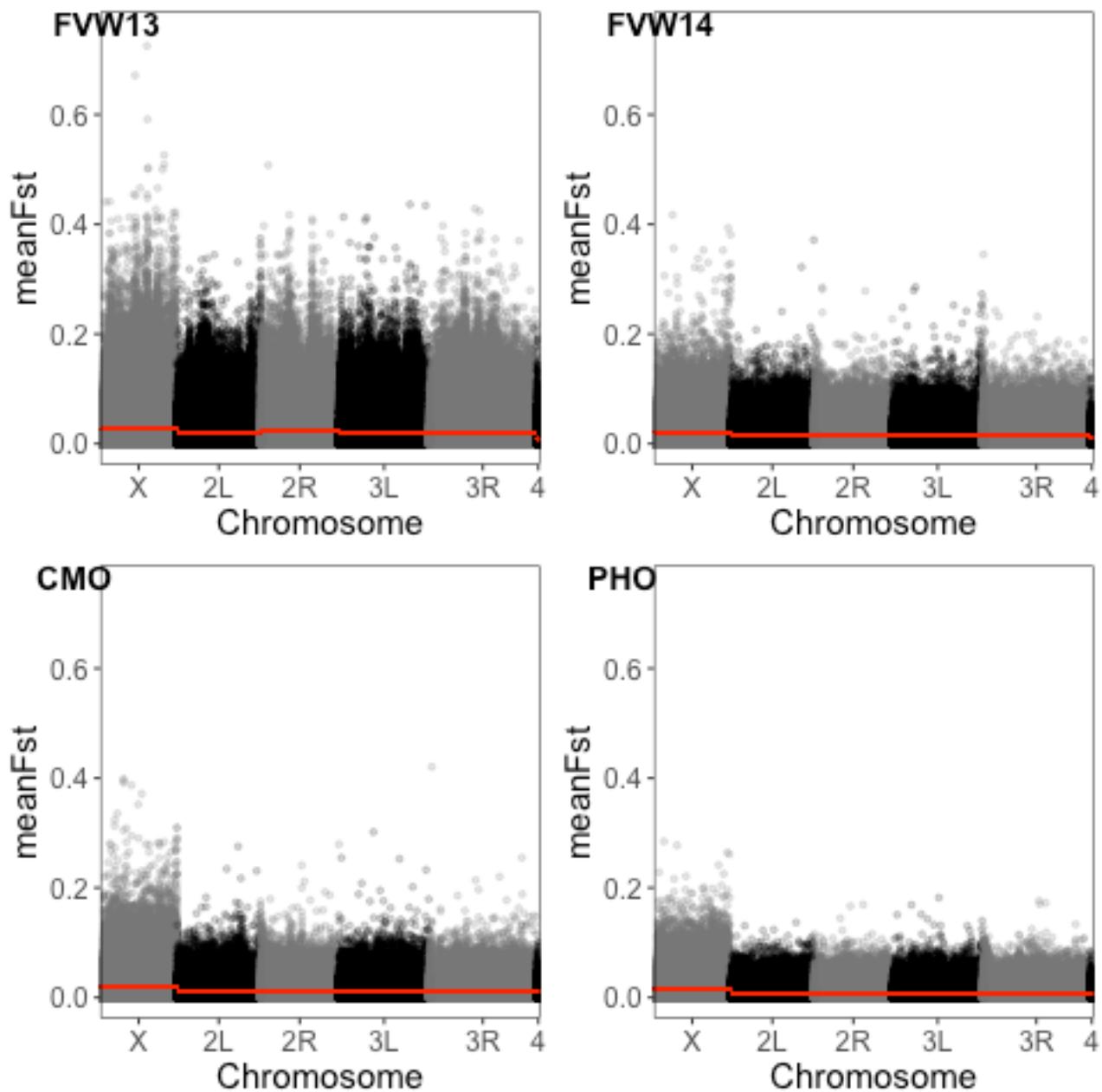
Supplemental Figure 19. Shape variation within *ds* selected pools for bulk segregant analysis by population. Wings within selected pools (one red and one blue, representing the "up" and "down" pools) are plotted to show the phenotypic extremes for the shape scores used for selecting individuals within and between pools. Black line indicates mean shape between pools.

Supplemental Figure 20. Removing PHO from the BSA analysis increases the number of differentiated sites between *ds* shape change pools. Genome-wide scan for differentiated loci between pools selected based on *ds* shape change vector using the CMH test using ACER. Variants in *ds* are still not implicated in this analysis. Points in red indicate sites with significant differentiation based with a FDR of 0.05.
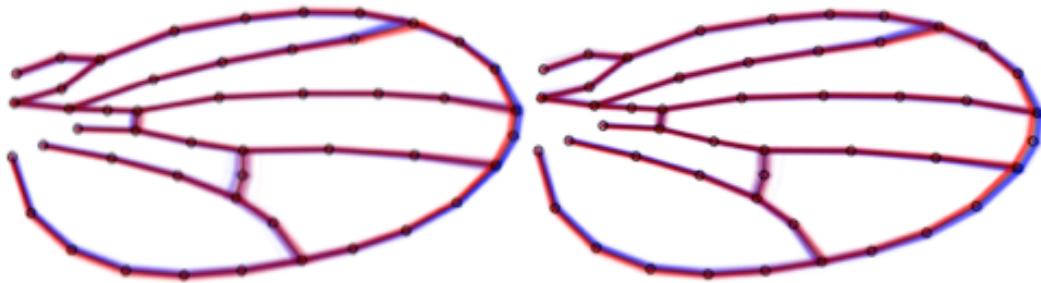

Supplemental Figure 21. Down sampling genome coverage does not impact results for CMH tests substantially. When all pools are sampled to a coverage depth of 75 reads, while preserving allele frequency, we do not find an increase in number of differentiated sites. Although the sites identified as significantly differentiated do change somewhat, in part due to sampling procedure as any sites without a minimum depth of 75 reads in each of the 4 populations was dropped from this analysis. Variants in *ds* are still not implicated in this analysis.

Supplemental Figure 22. F_{ST} between pools of individuals selected along the *ds* shape change axis within each population. Calculated in 100 bp windows using PoPoolation2 program. Elevated F_{ST} on the X chromosome is due to sampling of fewer X chromosomes, relative to autosomes as most pools, with the exception of FVW14, consist of only males. Red line indicates the mean F_{ST} for the chromosome, which as expected is very low.

CMO

PHO
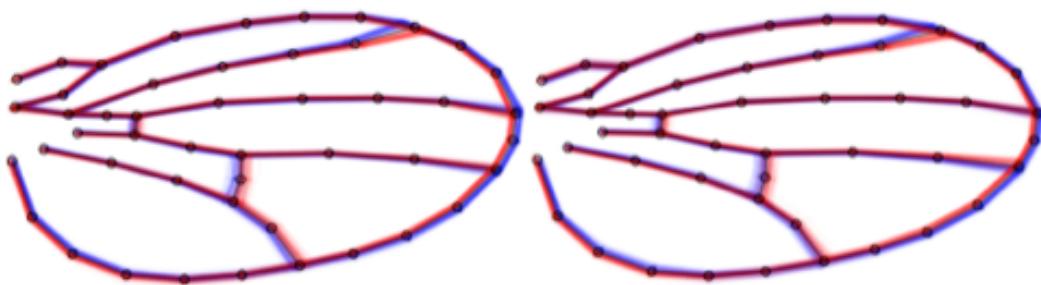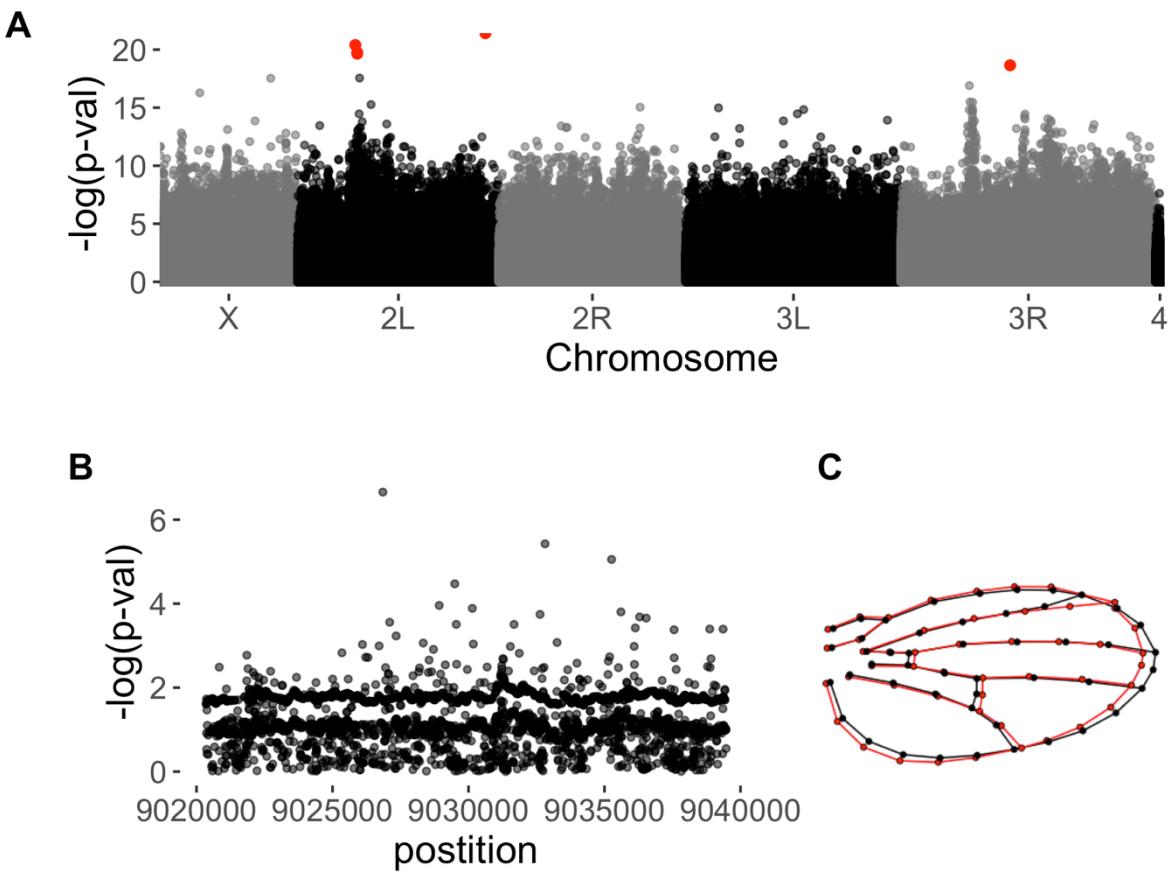
FVW13
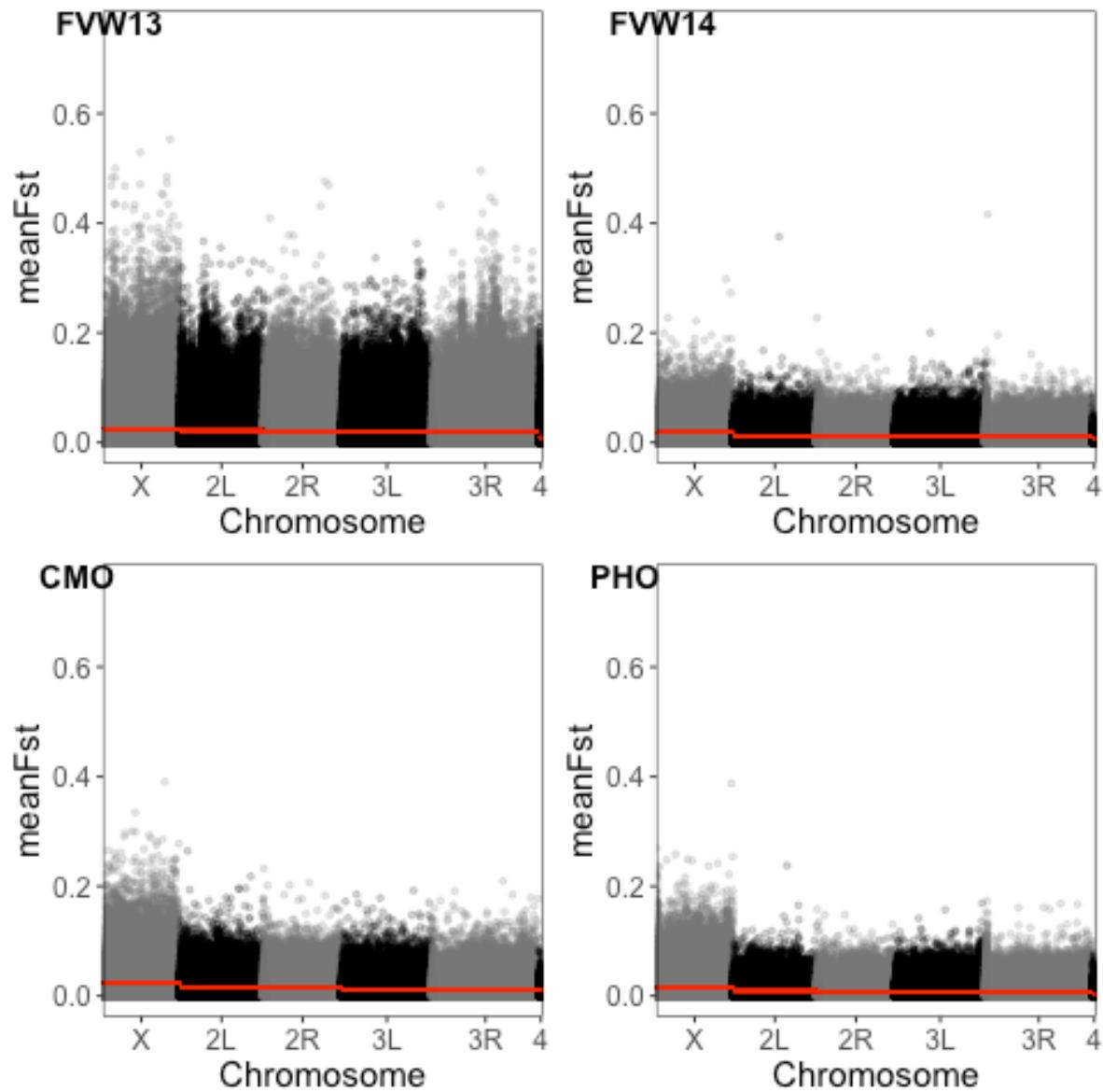
FVW14

Supplemental Figure 23. Shape variation within *neur* selected pools for bulk segregant analysis by population. Wings within selected pools (one red and one blue, representing "up" and "down" extreme pools respectively) are plotted.

A

B

C

Supplemental Figure 24. Genome-wide scan for differentiated loci between pools selected based on *neur* shape change vector using the CMH test implemented with ACER. (A) Whole genome scan for differentiation. Points in red indicate sites with significant differentiation based with a FDR of 0.05. (B) No significantly differentiated sites within *neur.* (C) shape change between selected pools based on *neur* shape change vector, effect size is multiplied by 2 for visualization.

Supplemental Figure 25. F_{ST} between pools of individuals selected along the *neur* shape change axis within each population. Calculated in 100 bp windows using PoPoolation2 program. Elevated F_{ST} on the X chromosome is due to sampling of fewer X chromosomes, relative to autosomes as most pools, with the exception of FVW14, consist of only males. Red line indicates the mean F_{ST} for the chromosome.

Supplemental Table 1. Number of individuals used for BSA from each wild-caught cohort

Collection	Females Phenotyped	Males Phenotyped	Females – "up" pool	Females – "down" pool	Males – "up" pool	Males – "down" pool
FVW13	0	2184	0	0	75	75
FVW14	403	797	59	0	16	75
PHO14	0	1232	0	0	75	75
CMO14	0	1001	0	0	75	75

Supplemental Table 2. Directions of major axes of variation differ somewhat between the among line (DGRP) covariance matrix and the wild caught cohorts, likely due to a combination of differences in allele frequencies and the major contribution of environmental variation (and potentially GxE) in the wild caught individuals. Correlations for the first three eigenvectors computed from among the DGRP, wild-cohorts (all populations) and CMO shape variation. |Estimates| < 10^{-15} are treated as 0.

	dgrp.PC1	dgrp.PC2	dgrp.PC3	wild.PC1	wild.PC2	wild.PC3	cmo.PC1	cmo.PC2	cmo.PC3
dgrp.PC1	1	0	0	-0.17	-0.46	0.26	-0.35	-0.54	0.066
dgrp.PC2		1	0	-0.18	-0.21	0.61	-0.10	-0.60	-0.27
dgrp.PC3			1	0.56	-0.12	0.40	-0.32	0.27	-0.71
wild.PC1				1	0	0	-0.39	0.68	-0.29
wild.PC2					1	0	0.90	0.31	-0.061
wild.PC3						1	0.056	-0.55	-0.74
cmo.PC1							1	0	0
cmo.PC2								1	0
cmo.PC3									1

Supplemental Table 3. Top 50 enriched GO terms of linked differentiated sites following artificial selection based on *ds* shape change
See: ds_artSelGOtop50.csv

Supplemental Table 4. Top 50 enriched GO terms of linked differentiated sites following artificial selection based on *emc* shape change
See: emc_artSelGOtop50_3sd.csv

Supplemental Table 5. Pairwise Procrustes (Euclidian approximation) distances between mean shapes across wild cohorts. Pairwise comparisons between population means estimated from a model with fixed effects for centroid size and population. Z-score and p-value are calculated using permutation of residuals as implemented in the RRPP package using a null model with only the effect of centroid size. Comparisons to the PHO samples, which show the largest pairwise differences are in bold.

Population Comparison	Distance	95% permutation threshold	Z-score	p
CMO-FVW13	0.0070	0.0011	9.70	0.001
CMO-FVW14	0.0084	0.0014	11.0	0.001
CMO-PHO	**0.021**	**0.0013**	**15.2**	**0.001**
FVW13-FVW14	0.0065	0.0012	7.25	0.001
FVW13-PHO	**0.022**	**0.0012**	**13.0**	**0.001**
FVW13-PHO	**0.022**	**0.0014**	**13.5**	**0.001**

Supplemental table 6. Significantly differentiated sites when PHO population is left out of CMH test.
See: ds_wild_significantSites_zeroGen_genes_fdr_noPho.csv

Supplemental table 7. GO analysis of significantly differentiated sites found in CMH test when PHO population is left out of analysis.
See: GOanalysisTop50_cmhSites_fdr_noPHO_ds_wild.csv

Supplemental Table 8. Significantly differentiated variants for *ds* shape change (CMH test) from the wild-caught cohorts (BSA) when all data is down-sampled to 75X coverage per site.

Location	CMH p-value (FDR corrected)	Gene	FlyBase ID	Distance from ORF (bp)
X:14833408	5.42×10^{-5}	*CG14411*	FBgn0030582	0
X:14833412	3.29×10^{-5}	*CG14411*	FBgn0030582	0
X:15381676	7.54×10^{-5}	*CG9164*	FBgn0030634	0
X:16037404	2.21×10^{-4}	*Muc14A*	FBgn0052580	0
X:22956575	4.58×10^{-2}	*lncRNA:CR45502*	FBgn0267058	1501
X:4118816	9.49×10^{-3}	*tyf*	FBgn0026083	0
X:4130993	9.84×10^{-3}	*GlcAT-I*	FBgn0066114	218
2L:14793860	1.15×10^{-2}	*CG18420*	FBgn0028866	656
2L:1480390	3.03×10^{-2}	*CG31928*	FBgn0051928	11905
2L:20626829	7.29×10^{-4}	*lncRNA:CR44909*	FBgn0266214	623
2L:4481772	1.54×10^{-2}	*dpy*	FBgn0053196	0
2L:4481788	3.70×10^{-2}	*dpy*	FBgn0053196	0
2L:4843588	4.62×10^{-2}	*mxt*	FBgn0031637	0
2R:11876761	2.73×10^{-2}	*CG13185*	FBgn0033661	0
2R:16405045	3.70×10^{-6}	*Sema2b*	FBgn0264273	0
2R:9943212	5.15×10^{-3}	*Mef2*	FBgn0011656	0
3L:19082031	4.49×10^{-3}	*Mkp3*	FBgn0036844	0
3L:19082032	2.73×10^{-3}	*Mkp3*	FBgn0036844	0
3L:23237904	3.05×10^{-2}	*nAChRalpha4*	FBgn0266347	0

Supplemental table 9. Significantly differentiated variants for *neur* shape change from the wild-caught cohorts (BSA).

Location	CMH p-value (FDR corrected)	Gene	FB ID	Distance from ORF (bp)
2L:15775767	0.026	*CG43760*	FBgn0264260	675
2L:15967220	0.017	*Beat-Ib*	FBgn0028645	4785
2L:15967222	0.017	*Beat-Ib*	FBgn0028645	4787
2L:8651597	0.015	*Sema1a*	FBgn0011259	0
3R: 21898160	0.024	*CG6678*	FBgn0038917	0