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]*, Anna Mammel[2,7], Emmalee Northrop-Albrecht[2,5], Eladio J. Márquez[3,6], Rosa A. Moscarella[3,4], David Houle[3], Ian Dworkin[1,2]
*Co-first authors
Corresponding authors: dworkin@mcmaster.ca, dhoule@bio.fsu.edu

Author affiliations: 1- Department of Biology, McMaster University

2 - Department of Integrative Biology, Michigan State University

3- Department of Biological Science, Florida State University

4-Current address: Department of Biology, University of Massachusetts

5- Current address Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester MN

6 – Current address: Dragonfly Therapeutics, Waltham MA

7-Current address: Neurocode USA. 3548 Meridian St, Bellingham, WA 98225

Abstract

Identifying the genetic architecture of complex traits is of interest to many geneticists, including those interested in human disease, plant and animal breeding and evolutionary genetics. Despite advances in sequencing technologies and GWAS statistical methods improving our ability to identify variants with smaller effect sizes, many of these identified polymorphisms fail to be replicated in subsequent studies. In addition to sampling variation, this 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 GWAS (Pitchers et al. 2019) in three genes: *dachsous (ds)*, *extra-macrochaete (emc)* and *neuralized (neur),* using artificial selection in the lab and bulk segregant mapping in natural populations. We demonstrate that shape changes associated with these genes is aligned with major axes of phenotypic and genetic variation in natural populations. Following 7 generations of artificial selection along *ds* and *emc* shape change vectors, we observe genetic differentiation of variants in *ds* and in genomic regions with other genes in the hippo signaling pathway, indicating

available genetic diversity of a population summarized in **G** influences 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, low minor allele frequencies coupled with small effect sizes of the contributing variants.

Introduction

Dissecting the genetic architecture underlying complex traits remains challenging because of the joint contributions of many alleles of small effect and genotype-by-environment interactions, among other factors. Advances in sequencing technology in conjunction with development of GWAS statistical methodologies, have enabled identification of loci contributing to numerous complex traits and diseases. However, such mapping approaches often identify only a subset of loci contributing to trait variation (Visscher et al., 2017). In part, this reflects the low power for detection of alleles of small effect or very rare alleles without the use of prohibitively large sample size (Tam et al., 2019). For alleles that are relatively common in a population, replication rates between GWAS studies is surprisingly 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 of the alleles identified in these studies are rare in a population and require very large cohorts to detect (Fritsche et al., 2016; Ioannidis et al., 2011). Despite this ability to detect associations with traits across studies, the magnitude of effect can vary between different cohorts or populations (CONVERGE consortium, 2015; Marigorta et al., 2018).

Replication of genetic effects can become more complicated when effects can vary in direction in addition to magnitude, as is the case for multivariate traits. Despite the commonly cited concern of the “curse of dimensionality” and associated needs for large sample sizes, studying genetic effects in a multivariate (or multidimensional) context is beneficial. 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 including dimensional reduction techniques (Fatumo et al., 2019; Pitchers et al., 2019; Porter and O’Reilly, 2017;

Shriner, 2012). Second, a multivariate approach considers pleiotropic effects in a statistically rigorous and justifiable manner (Melo et al., 2019). Additionally, the ability to study both the direction and magnitude of genetic effects provides an additional and important way of assessing repeatability. We may be less confident in assessing replication of genetic effect for univariate measures as half the time the "replicated" measure will be in the same direction as the original estimate, even if there is no true effect. By contrast, the probability of a "replicated" genetic effect sharing a similar direction by chance alone decreases as the number of dimensions of the trait of interest increases (Marquez and Houle, 2015; Stephens, 2013).

Understanding the replication of effects in different biological contexts is of fundamental interest to evolutionary biologists and breeders as these multivariate genetic effect vectors can be directly expressed in terms of the multivariate form of the breeder's equation, $\Delta \boldsymbol{z} = \boldsymbol{G\beta}$, which is important for 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 variance-covariance matrix (**G**), and in particular the axis of greatest genetic variation, $\mathbf{g}_{max}$. Theory and empirical studies demonstrate that direction of $\mathbf{g}_{max}$, can influence evolutionary trajectories (Blows and McGuigan, 2015; McGuigan, 2006; Schluter, 1996).The degree to which genetic effects associated with particular variants align to major axes of genetic 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 multivariate traits, it is important to consider not only the direction of effect for a single gene but also correlated effects that contribute to the phenotype. Interestingly, initial comparisons of directions of genetic effects among induced mutations in two *Drosophila melanogaster* wing development pathways showed only partial relationships within and between pathways (Dworkin and Gibson, 2006). These mutant alleles varied for magnitude of effects from quite weak to severe, while still allowing for wing shape to be measured in the heterozygous state. Despite this, recent work has demonstrated that the direction of genetic effects of SNP variants in a genome wide scan could be similar to those from validation experiments using RNAi knockdown of genes (Pitchers et al., 2019).

Pitchers et al. (2019) demonstrated that in several cases the RNAi effects of genes in particular pathways, such as hippo signalling, were aligned with SNP effects for other genes in the same pathway. Interestingly, the hippo pathway demonstrated an over-representation of SNPs associated with wing shape in the genome wide association. The degree to which these 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. There are some biological factors that might suggest that such an approach to validation may be unsuccessful. Given common dominance patterns, and the likely non-linear genotype-phenotype relationships of most genetic effects, small to moderate changes in gene function may result in modest phenotypic effects (Green et al., 2017; Melo et al., 2019; Wright, 1934). As such, large effect mutants and many RNAi knockdown studies generally reflect a moderate to large phenotypic effect that may not be reflective of the genetic effects (in direction as well as magnitude) of SNPs that contribute to phenotypic variance in natural populations.

The context-dependence of genetic effects for a multivariate trait has been demonstrated using *Drosophila* wing shape. Variants in *Egfr* influencing *Drosophila* wing shape are replicable in both lab reared and wild-caught cohorts (Dworkin et al., 2005; Palsson et al., 2005; Palsson and Gibson, 2004). 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 1/10 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, in one natural population the genetic effect on wing shape from this SNP largely disappeared (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), genetic background (Greene et al., 2009) among others. Because both environment and genetic background likely affect the genotype-phenotype map in a non-linear fashion (Wright, 1934), it is important to test observed associations in other experimental contexts.

An important but rarely used approach which has direct relevance to evolutionary biology is to test whether such candidate genetic variants respond to artificial selection in the direction of the inferred effect. In this study we used artificial selection for phenotypic directions defined by previously estimated genetic effects as well as a bulk segregant mapping approach using several natural populations. Using several candidate SNPs identified in a previous GWAS of Drosophila wing shape (Pitchers et al., 2019) we used vectors of shape change based on RNAi knockdowns of the relevant genes. We demonstrate that the direction of shape change for these genetic effects is aligned with major axes of natural phenotypic and genetic variation. Then, using artificial selection on a synthetic outcrossed population and sampling of natural populations, we were able to replicate effects of SNPs in some of these genes, in particular *dachsous* (*ds*), a component of the hippo signaling pathway. Interestingly we observed evidence that while selecting on a vector of shape change for *ds*, not only did variants in the *ds* locus respond to selection, but several other components of the hippo signaling pathway. 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 *Drosophila* genetic resource panel (DGRP) was collected for 184 lines as part of a GWAS study as described in Pitchers *et al* (2019). Genotype data for these lines was obtained from freeze 2 of the DGRP (Huang et al., 2014). For replication using artificial selection, 30 DGRP lines 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 lines were selected to increase genetic variation at the *ds* locus (Supplemental Figure 2, Table 1). Reciprocal pairwise crosses between the 30 selected DGRP lines were used to create heterozygotes and these 30 heterozygous genotypes were successively pooled for 4 subsequent generations, allowing for recombination. After pooling, the synthetic outbred

population was maintained for approximately 47 subsequent generations (generations allowing for recombination) before the start of artificial selection experiments.

For the replication in wild-caught populations, 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 FVW13 collection, only males were used in this study given difficulties differentiating *Drosophila melanogaster* and *D. simulans* females. 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). The LDA algorithm was trained using male wings from the collected *D. melanogaster* data set as well as males from a *D. simulans* data set acquired from the Hosken lab. There was 100% agreement between the classification of females within each species with our phenotypic classification, indicating that it is very unlikely that *D. simulans* females were included in our samples (Supplemental Figure 3).

Morphometric Data

Image collection for each experiment was performed as described below. Landmark and semi-landmark data were captured from black and white TIFF images using the pipeline described in Houle, 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 correct errors. All shape data was subjected to full Procrustes superimposition to scale images to the same size, translocate to the same location and rotate to minimize distance between points (Rohlf and Slice, 1990). Superimposition and extraction of 14 landmark and 34 semi landmark positions was done with the program CPR v1.11 (Márquez 2012–2014, Figure 1). Superimposition results in the loss of 4 possible dimensions of variation while semi-landmarks are constrained to vary over only one axis,

restraining these points to approximately a single dimension of variation. This results in a total of 58 available dimensions of shape variation, that can be summarized using the first 58 PCs. These principal components were used in all subsequent analysis, with the exemption of shape models fit using the Geomorph package in R, where Procrustes landmarks were used.

Generation of shape vectors for selection

A panel of shape change vectors was estimated using the progesterone-inducible Geneswitch Gal4 construct under the regulation of an ubiquitous *tubulin* driver to drive the expression of RNAi for genes of interest as previously described in Pitchers *et al* 2019. Knockdown was performed at varying concentrations of mifepristone, an analog of progesterone, [0, 0.3, 0.9, 2.7 μM] to titer the response. Using multivariate regression of phenotype on concentration of mifepristone the shape change associated with knockdown of the gene of interest was determined. Shape change vectors for *dachsous* (*ds*), *extramacrochaete* (*emc*) and *neuralized* (*neur*) were used in this experiment (Figure 1b, Supplemental Figure 1).

Shape data collected as part of a previous study (Pitchers et al., 2019) was used to assess the relationships between shape change vectors and $\mathbf{g}_{max}$. 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 onto the shape change vector defined by the knockdown experiment. Principal component analysis was performed on the model residuals and the correlation between the first 3 eigenvectors ("genetic PCs") and shape scores was calculated (Figure 1a). From this, *ds emc* and *neur* shape change vectors were selected for further experiments due to high correlation with directions of natural genetic variation.

Synthetic outbred population and artificial selection

The synthetic outbred population was used as the parent population for artificial selection. Each artificial selection experiment was carried out with three independent replicates of each "up" selection, "down" selection and 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). A shape score was calculated by projecting the data onto the *ds* shape change vector and the 40 individuals with highest shape

scores and 40 individuals with lowest scores, were selected for the next generation in the up and down generations, respectively. In control lineages, 40 individuals were randomly selected for the next generation. Following 7 generations of selection, 75 individuals from each lineage were selected for pooled sequencing. Genomic DNA was extracted using Qiagen DNeasy DNA extraction kit. The response to selection was evaluated both by computing Procrustes distance (PD) between average shape of wings between generations 1 and 7, and also using shape scores (projections) with a mixed effect model allowing for the fixed effect factors of treatment, 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 differential on cumulative selection response, averaging over sex and with a random effect of replicate lineage.

Wild populations

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 20X magnification. When possible, both left and right wings were dissected and imaged and averaged to calculate an individuals 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 Supplemental 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 FVW13 population and regressing shape onto centroid size and sex gave equivalent results (Supplemental Figure 4). 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, emc,* or *neur* knockdowns. The 75 most extreme individuals within each cohort on the shape score distribution were selected for pooled sequencing. Genomic DNA was extracted using a Qiagen DNeasy DNA extraction kit. The difference vector between the mean shapes of selected pools (within each population) was used to calculate the PD between pools and the correlation of this shape change vector with the selection vector used.

Sequencing and Genomic Analysis

Library preparation and Illumina sequencing was performed at the RTSF 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 (150bp) were sequenced using Illumina HiSeq 2500, with each sample 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 *Drosophila melanogaster* genome (v6.23) using BWA-MEM (v0.7.8) version (Li and Durbin, 2010). Sequencing replicates of the same biological samples were merged using SAMtools version. PCR duplicates were removed using Picard with the MarkDuplicates tool (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 (v 4.1.1) using default settings. INDELs and regions within 5bp of an indel were identified and masked using popoolation2 scripts. Population genetic statistics were calculated using PoPoolation (v 1.2.2) and PoPoolation2 (v1.201) (Kofler et al., 2011b, 2011a).

For wild cohorts, the Cochran-Mantel-Haenszel (CMH) test was used to measure significantly differentiated sites. 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 with an adjusted alpha of 0.05. For wild populations, each significant site from the CMH test, at a corrected p-value cut-off of 0.05 we identified the nearest gene using BEDtools (v2.19.1)(Quinlan and Hall, 2010).

For artificial selection experiments, F_{ST} was calculated in 5000bp sliding windows. We chose this larger 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 (E. G. King et al., 2012; Elizabeth G. King et al., 2012; Marriage et al., 2014). This statistic was used to compare the "left" selected pools to the "right" 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 using the bumphunter package (v. 1.24.5) in Bioconductor, clustering regions with a gap less than 10000bp between regions (Jaffe et al., 2012). High F_{ST} was defined as F_{ST} values greater than 3 standard deviations above the mean. GO terms associated with identified genes were annotated using TopGO package (v. 2.34.0) (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 with a FDR of 5%.

Results

dachsous (ds) shape change is aligned with major axes of genetic and phenotypic variation in natural populations

To assess the relationship between shape change vectors and axis of variation in the DGRP, mean shape vectors were calculated for each DGRP line then used in a PCA to summarize axes of variation. The mean shape vectors for each DGRP line were projected onto vectors for *ds* and *emc* defined from the RNAi knockdown and the correlations between them (Figure 1) and with PC1 generated from DGRP line means (*ds*: r = 0.56; *emc*: r = 0.45; Figure 1). As

expected, a similar response was observed with *ds* and *emc* shape change vectors due to the high correlation between shape change vectors (Figure 1,2, r = 0.65). *neur* shape change is aligned with PC1 (r = -0.69) and PC3 (r = -0.64), indicating that this is also an important axis of shape variation in this population (Figure 1) that differs from both *ds* (r = 0.034) and *emc* (r = 0.30) shape change vectors. 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, where phenotypic variance for shape is due to the joint contribution of genetic and environmental effects. As observed with the DGRP, there is a substantial correlation between *ds* shape change and PC1 in most of the sampled cohorts (PHO: r = 0.78; CMO: r = 0.87; FVW12: r = -0.22; FVW14: r = 0.95, Figure 2, Supplemental Figure 5). In the cohorts where the *ds* shape change vector was not correlated with PC1, specifically the FVW12 collection, this vector is correlated with PC2 (PHO: r = 0.12; CMO r = -0.44; FVW12: r = -0.63; FVW14: r = 0.19; Figure 2) from the wild caught individuals. We also observe a correlation between *neur* shape change and PC1 in most cohorts (PHO: r = 0.51; CMO: r = -0.051; FVW12: r = -0.95 FVW12; FVW14: r = -0.084; Figure 2). As with the *ds* shape change vector, in cases such as the CMO cohort and PC1 there is a correlation between the *neur* shape change vector and PC2 (PHO: r = 0.22; CMO: r = -0.57; FVW12: r = -0.0.059; FVW14: r = 0.85; Figure 2). Interestingly, in the CMO natural cohort of flies, the correlations between the projection of shape data onto the *ds* and *neur* shape change vectors is low (r = 0.12, Supplemental Figure 5).

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, replicating the findings of the earlier GWAS (Pitchers et al. 2019), we performed artificial selection experiment for wing shape along the *ds* shape change vector to examine genomic response to selection. We observed a shape change in both the “up” (females: Procrustes Distance (PD) = 0.039, males: 0.044) and “down” directions (females: PD = 0.022, males: PD = 0.022) compared to the first

generation of selection. In comparison, the shape change among the unselected control lineages was much smaller (females: PD = 0.005, males: PD = 0.005) (Figure 3). The direction of phenotypic shape change after 7 generations of selection was in a similar direction to the *ds* shape change vector (defined by the 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 (Supplemental Figure 7, up = 0.38, 95% CI: 0.25 – 0.50; down = 0.28, 95% CI: 0.24 – 0.50) were moderate.

Genome wide F_{ST} was estimated between up and down *ds* selection lineages. We observed strong genetic differentiation linked with the *ds* locus (Figure 3, Supplemental Figure 6). We also observed striking genetic differentiation across several regions in the genome. Gene ontology analysis for genes in regions of the genome with an F_{ST} greater than 0.263 (two standard deviations from mean F_{ST}), show enrichment for hippo signaling loci (Supplemental Table 2). In addition, one of the SNPs (2L:702560) identified in Pitchers *et al* though 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). The top 20 enriched terms are all related to cell signaling and development but of note is the specific inclusion of the terms for 'negative regulation of hippo signaling', 'hippo signaling' and 'regulation of hippo signaling' in this list (Supplemental Table 2, Supplemental Figure 6). Although hippo signaling is often also associated with changes in size (Pan, 2007), we do not observe a significant change of size in our selection lineages in either sex (Supplemental Figure 8). However, with more generations of selection it is possible we would have observed a clear change in size as there is a trend indicating such divergence (Supplemental Figure 8).

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) (figure 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 (Supplemental Figure 9, 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* (Figure 4, Supplemental Figure 6). Notably, as seen in supplementary figure 2, the site frequency spectrum (SFS) suggests only modest allelic variation at *emc* locus in the synthetic outbred population. Using a 2 standard deviation cut-off for F_{ST}, while we did observe enrichment for various developmental GO terms, we did not observe enrichment of specific hippo signaling terms (Supplemental Table 3, Supplemental Figure 6).

Bulk segregant analysis in wild caught cohorts do not recapitulate effects of the GWAS or artificial selection

Having demonstrated that variants in *ds* (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. As previously discussed, wild caught populations introduce considerably more environmental variation for shape along with a different site frequency spectrum for variants that can contribute 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 the alignment and variant calling in this region challenging. 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. 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*.

We observed modest, but significant wing shape differences among collection cohorts using a Procrustes ANOVA, along with permutations of the residuals for the relevant "null" model (Supplemental Table 4; R^2= 0.16, F = 351, Z_{RRPP} = 18.3, p = 0.001). This appears to be due to differences in wing shape between the PHO population and other populations based on pairwise Procrustes Distances. This should not influence downstream genomic analysis as

(Collyer and Adams, 2018), individuals used for generating pools were compared within each population. Because there is a single bout of selection distinguishing "up" and "down" pools in the BSA, shape and allele frequency changes are expected to be modest. We observe shape difference between selected pools within each population (PD = CMO: 0.033; PHO: 0.036; FVW14: 0.041; FVW12: 0.040). Correlations of the shape difference vectors of the pools (i.e. difference between "up" and "down" pools) and the direction of the *ds* shape change vector used for selection is high (CMO: 0.94, PHO: 0.79, FVW12: 0.92, FVW14: 0.90).

The genome scans from the bulk segregant analysis show little evidence of genetic differentiation linked with *ds* (Figure 5). Across the genome, 15 sites were detected as significantly differentiated between "up" and "down" selected pools based on a CMH test with a FDR cut-off of 5% (Figure 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 the PHO population had a distinct shape from the other populations and had a lower correlation of the difference vector between selected pools and the *ds* shape change vector, we repeated the CMH test with this population left out. We observe a significant differentiation at 174 sites between pools (Supplemental Table 5, Supplemental Figure 10). We identified the nearest gene to these sites and GO analysis indicated enrichment for wing development terms, in particular many terms related to Wnt signaling, but not specific hippo signaling terms (Supplemental Table 6). Importantly, we do not observe differentiation linked to *ds* or any other hippo loci.

In addition to *ds* selection, we also selected pools of individuals based on the *neur* shape change vector. This shape change vector is not aligned with *ds* but does align with $\mathbf{p}_{max}$ in wild populations (figure 1, 2). We observe shape changes between "up" and "down" selected pools (PD = CMO: 0.027; PHO: 0.028; FVW14: 0.041; FVW12: 0.038). As with the *ds* experiment, there is little evidence of genetic differentiation between *neur* selected pools (Supplemental figure 11). Only 4 sites were identified as being significantly differentiated between left and right pools and none of these sites are associated with wing development (Supplemental Table 7).

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 a phenotype and observing changes in allele frequencies. We observed that shape changes associated with *ds*, *emc* and *neur* were both similar to effects of other gene perturbations, and 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 (Figure 1, 2). After observing a strong response to artificial selection along two of these shape change vectors (*ds* and *emc*), we examined patterns of genomic differentiation and observed substantial changes in allele frequency for markers linked with *ds* itself, and also markers linked to numerous other genes associated with hippo signaling (Figure 3, 4). Despite this, using pools of wild caught individuals chosen to be phenotypically divergent on the same shape vectors, none of the same loci identified in the artificial selection experiments showed genetic differentiation (Figure 5, Supplemental Figure 11). 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 *ds* but also linked to a number of other hippo signalling loci (Figure 3, Supplemental Table 2). 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 created haplotype blocks containing many of the candidate SNPs that we could select upon to identify causative alleles though a frequency change. Although LD blocks in the outcrossing population used in this study remain large, *ds* variants exist on distinct haplotypes, allowing for an examination of allele frequency changes for particular variants. Of interest is one SNP, 2L:702560, which was previously identified though GWAS (Pitchers et al., 2019) as influencing wing shape variation being driven to near fixation in each of the artificial selection lineages (Table 1). Importantly, we recognize that the haplotype blocks are large, and contain many potential functional variants. However based on the results of both the current and previous study, these variants in *ds* represent good candidates for functional validation in future work.

We created a synthetic outbred population with the intention of maximizing genetic diversity at *ds.* When selecting on the *emc* shape change vector, which is similar but not identical to *ds* shape change, we observe only a modest allele frequency change at *emc* but a much more robust response at *ds*. This is not particularly surprising given the increased diversity at *ds* as compared to *emc* in the founding population. A study using *D. simulans*, demonstrated that alleles of intermediate frequency in parental populations contribute more to polygenic adaptation on short time scales (Kelly and Hughes, 2019). In this case, the orientation of **G** influenced what variants were captured by selection.

Previous studies of the direction of genetic effects on wing morphology using RNAi knockdown of gene activity in the *Egfr* and *Tgf-*β signaling pathways showed no relationship between the direction of effects within developmental pathways (Dworkin et al., 2005). However, models for the architecture of complex traits predict that many alleles of small effect will contribute to trait variation with many genes within developmental pathways contributing (Boyle et al., 2017; Wray et al., 2018). In contrast to studies titrating gene function using RNAi knockdown, which constrains genetic effects to a particular direction, this study used variants that occur in the DGRP. These variants may either increase or decrease gene function and increases the chance that two genetic effects in the pathway are in the same direction. This can help to explain the observation of the correlated response to selection of not just *ds* but also other hippo signaling loci (figure 3). This pathway response has also been demonstrated in human adaptation to pathogen resistance (Daub et al., 2013) and adaptation to high altitude (Gouy et al., 2017). These results demonstrate an exciting relationship between polymorphisms in the same developmental pathway having 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). With these variants at low frequencies, they are less likely contribute to short term selection, as even alleles with relatively large effects (a) will contribute relatively little to overall genotypic variance if frequencies (p, q) of one of the alternative alleles is low (i.e $V_A = 2pqa^2$). Furthermore such rare variants are even less likely to occur in the same individuals with other rare variants at other sites. Therefore, the outcrossed

population we created here may be an idealized situation with which to test the response to selection, rather than one reflective of natural populations where functional variants are rare, and environmental effects may play a substantial role in phenotypic variance.

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 bulk segregant analysis 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 also were found in genome wide associations (Pitchers et al., 2019). However, there are many explanations for why we may not have been able to detect these allele frequency changes in our experiment. First, the addition of environmental variation to the system introduces additional complications. In the aforementioned example of variants in *Egfr*, the genetic effect of the SNP in wild caught cohorts was only 10% of the magnitude compared with lab reared flies (Dworkin et al., 2005). 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 dominates and can result in low minor allele frequencies at these sites. Of note, the variants previously implicated in *ds* shape variation (Pitchers *et al* 2019), are at low frequency in the wild cohorts (Table 2). Because allelic contribution to wing shape are expected to be both rare in the 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 (ref) to account for sampling effects, we observe few differentiated sites, and none in *ds*, indicating that bulk segregant analysis (BSA) may not be a well-suited approach to identify modest allele frequency changes and as such may not be particularly effective for highly polygenic traits. In contrast to our study, in cases where BSA approaches have been able to identify variants in natural populations, these studies tend have a smaller number of contributing loci and identify a 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 approach (Bastide et al., 2013). Potentially pigmentation may represent a relatively ‘simpler’ genetic architecture

(fewer variants of individually larger genetic effect, smaller impact of environmental variation, smaller mutational target size). 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, likely making this approach less sensitive.

Our approach for the BSA was to use the phenotypic selection across 4 distinct "populations" but performing the same phenotypic selection within each. Even with 4 populations all sampled from Michigan, it is important to recognize that there can be heterogeneity, not only in allele frequencies, but in environmental variance and potentially GxE effects. The population from the Phillips Orchard (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, despite using the same projection vector (Supplemental Figure 10). 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 the genetic effects show a degree of GxE (with a specific environment in PHO), and the genetic backgrounds and environment for the PHO population leads to alleles contributing to shape variation in the *ds* direction in PHO. Because phenotypic variance has both genetic (**G**) and environmental (**E**) components, a unique **E** in one population aligned with *ds* shape change could impact the individuals sampled for the bulk segregant analysis. Such obfuscating effects have been observed before with the previously discussed *Egfr* example, where the same SNP effect that was initially identified and validated in multiple contexts (Dworkin et al., 2005; Palsson et al., 2005; Palsson and Gibson, 2004), could not be detected in one natural population, despite being at intermediate frequencies in each sample (Palsson et al., 2005). Importantly, although we do detect differentiation at sites associated with developmental processes in the wild cohorts, the differentiated sites are not linked to *ds* or other hippo signaling loci (Supplemental Table 5,6).

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 axis 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**, may at first 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 populations 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 'favoured' in evolution (Gompel and Prud'homme, 2009; Martin and Orgogozo, 2013; Stern and Orgogozo, 2008). 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. In particular, we think that ongoing coupling of more traditional mapping approaches like GWAS with short term artificial selection, may not only provide additional routes to validation and replication of genetic effect but can also yield a broader assessment of shared directions of genetic effects, in particular when using a multivariate perspective.

Acknowledgements:

We would like to thank the Hosken lab for *D. simulans* wing shape data.

References:

Adams, D.C., Otárola-Castillo, E., 2013. geomorph: an R package for the collection and analysis of geometric morphometric shape data. Methods Ecol. Evol. 4, 393–399. https://doi.org/10.1111/2041-210X.12035

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

Bastide, H., Betancourt, A., Nolte, V., Tobler, R., Stöbe, P., Futschik, A., Schlötterer, C., 2013. A Genome-Wide, Fine-Scale Map of Natural Pigmentation Variation in Drosophila melanogaster. PLoS Genet. 9, e1003534. https://doi.org/10.1371/journal.pgen.1003534

Blows, M.W., McGuigan, K., 2015. The distribution of genetic variance across phenotypic space and the response to selection. Mol. Ecol. 24, 2056–2072. https://doi.org/10.1111/mec.13023

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boyle, E.A., Li, Y.I., Pritchard, J.K., 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169, 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J., Myers, R.M., Petrov, D., Jónsson, B., Schluter, D., Bell, M.A., Kingsley, D.M., 2010. Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. Science 327, 302–305. https://doi.org/10.1126/science.1182213

Collyer, M.L., Adams, D.C., 2018. RRPP: An R package for fitting linear models to high-dimensional data using residual randomization. Methods Ecol. Evol. 9, 1772–1779. https://doi.org/10.1111/2041-210X.13029

CONVERGE consortium, 2015. Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature 523, 588–591. https://doi.org/10.1038/nature14659

Daub, J.T., Hofer, T., Cutivet, E., Dupanloup, I., Quintana-Murci, L., Robinson-Rechavi, M., Excoffier, L., 2013. Evidence for Polygenic Adaptation to Pathogens in the Human Genome. Mol. Biol. Evol. 30, 1544–1558. https://doi.org/10.1093/molbev/mst080

Dworkin, I., Gibson, G., 2006. Epidermal Growth Factor Receptor and Transforming Growth Factor-β Signaling Contributes to Variation for Wing Shape in *Drosophila melanogaster*. Genetics 173, 1417–1431. https://doi.org/10.1534/genetics.105.053868

Dworkin, I., Palsson, A., Gibson, G., 2005. Replication of an *Egfr* -Wing Shape Association in a Wild-Caught Cohort of *Drosophila melanogaster*. Genetics 169, 2115–2125. https://doi.org/10.1534/genetics.104.035766

Fatumo, S., Carstensen, T., Nashiru, O., Gurdasani, D., Sandhu, M., Kaleebu, P., 2019. Complimentary Methods for Multivariate Genome-Wide Association Study Identify New

Susceptibility Genes for Blood Cell Traits. Front. Genet. 10, 334. https://doi.org/10.3389/fgene.2019.00334

Fritsche, L.G., Igl, W., Bailey, J.N.C., Grassmann, F., Sengupta, S., Bragg-Gresham, J.L., Burdon, K.P., Hebbring, S.J., Wen, C., Gorski, M., Kim, I.K., Cho, D., Zack, D., Souied, E., Scholl, H.P.N., Bala, E., Lee, K.E., Hunter, D.J., Sardell, R.J., Mitchell, P., Merriam, J.E., Cipriani, V., Hoffman, J.D., Schick, T., Lechanteur, Y.T.E., Guymer, R.H., Johnson, M.P., Jiang, Y., Stanton, C.M., Buitendijk, G.H.S., Zhan, X., Kwong, A.M., Boleda, A., Brooks, M., Gieser, L., Ratnapriya, R., Branham, K.E., Foerster, J.R., Heckenlively, J.R., Othman, M.I., Vote, B.J., Liang, H.H., Souzeau, E., McAllister, I.L., Isaacs, T., Hall, J., Lake, S., Mackey, D.A., Constable, I.J., Craig, J.E., Kitchner, T.E., Yang, Z., Su, Z., Luo, H., Chen, D., Ouyang, H., Flagg, K., Lin, D., Mao, G., Ferreyra, H., Stark, K., von Strachwitz, C.N., Wolf, A., Brandl, C., Rudolph, G., Olden, M., Morrison, M.A., Morgan, D.J., Schu, M., Ahn, J., Silvestri, G., Tsironi, E.E., Park, K.H., Farrer, L.A., Orlin, A., Brucker, A., Li, M., Curcio, C.A., Mohand-Saïd, S., Sahel, J.-A., Audo, I., Benchaboune, M., Cree, A.J., Rennie, C.A., Goverdhan, S.V., Grunin, M., Hagbi-Levi, S., Campochiaro, P., Katsanis, N., Holz, F.G., Blond, F., Blanché, H., Deleuze, J.-F., Igo, R.P., Truitt, B., Peachey, N.S., Meuer, S.M., Myers, C.E., Moore, E.L., Klein, R., Hauser, M.A., Postel, E.A., Courtenay, M.D., Schwartz, S.G., Kovach, J.L., Scott, W.K., Liew, G., Tan, A.G., Gopinath, B., Merriam, J.C., Smith, R.T., Khan, J.C., Shahid, H., Moore, A.T., McGrath, J.A., Laux, R., Brantley, M.A., Agarwal, A., Ersoy, L., Caramoy, A., Langmann, T., Saksens, N.T.M., de Jong, E.K., Hoyng, C.B., Cain, M.S., Richardson, A.J., Martin, T.M., Blangero, J., Weeks, D.E., Dhillon, B., van Duijn, C.M., Doheny, K.F., Romm, J., Klaver, C.C.W., Hayward, C., Gorin, M.B., Klein, M.L., Baird, P.N., den Hollander, A.I., Fauser, S., Yates, J.R.W., Allikmets, R., Wang, J.J., Schaumberg, D.A., Klein, B.E.K., Hagstrom, S.A., Chowers, I., Lotery, A.J., Léveillard, T., Zhang, K., Brilliant, M.H., Hewitt, A.W., Swaroop, A., Chew, E.Y., Pericak-Vance, M.A., DeAngelis, M., Stambolian, D., Haines, J.L., Iyengar, S.K., Weber, B.H.F., Abecasis, G.R., Heid, I.M., 2016. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. Nat. Genet. 48, 134–143. https://doi.org/10.1038/ng.3448

Gilchrist, A.S., Partridge, L., 2001. The contrasting genetic architecture of wing size and shape in Drosophila melanogaster. Heredity 86, 144–152. https://doi.org/10.1046/j.1365-2540.2001.00779.x

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

Gouy, A., Daub, J.T., Excoffier, L., 2017. Detecting gene subnetworks under selection in biological pathways. Nucleic Acids Res. 45, e149–e149. https://doi.org/10.1093/nar/gkx626

Green, R.M., Fish, J.L., Young, N.M., Smith, F.J., Roberts, B., Dolan, K., Choi, I., Leach, C.L., Gordon, P., Cheverud, J.M., Roseman, C.C., Williams, T.J., Marcucio, R.S., Hallgrímsson, B., 2017. Developmental nonlinearity drives phenotypic robustness. Nat. Commun. 8, 1970. https://doi.org/10.1038/s41467-017-02037-7

Greene, C.S., Penrod, N.M., Williams, S.M., Moore, J.H., 2009. Failure to Replicate a Genetic Association May Provide Important Clues About Genetic Architecture. PLoS ONE 4, e5639. https://doi.org/10.1371/journal.pone.0005639

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

Huang, W., Massouras, A., Inoue, Y., Peiffer, J., Ramia, M., Tarone, A.M., Turlapati, L., Zichner, T., Zhu, D., Lyman, R.F., Magwire, M.M., Blankenburg, K., Carbone, M.A., Chang, K., Ellis, L.L., Fernandez, S., Han, Y., Highnam, G., Hjelmen, C.E., Jack, J.R., Javaid, M., Jayaseelan, J., Kalra, D., Lee, S., Lewis, L., Munidasa, M., Ongeri, F., Patel, S., Perales, L., Perez, A., Pu, L., Rollmann, S.M., Ruth, R., Saada, N., Warner, C., Williams, A., Wu, Y.-Q., Yamamoto, A., Zhang, Y., Zhu, Y., Anholt, R.R.H., Korbel, J.O., Mittelman, D., Muzny, D.M., Gibbs, R.A., Barbadilla, A., Johnston, J.S., Stone, E.A., Richards, S., Deplancke, B., Mackay, T.F.C., 2014. Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome Res. 24, 1193–1208. https://doi.org/10.1101/gr.171546.113

Ioannidis, J.P.A., Tarone, R., McLaughlin, J.K., 2011. The False-positive to False-negative Ratio in Epidemiologic Studies: Epidemiology 22, 450–456. https://doi.org/10.1097/EDE.0b013e31821b506e

Jaffe, A.E., Murakami, P., Lee, H., Leek, J.T., Fallin, M.D., Feinberg, A.P., Irizarry, R.A., 2012. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int. J. Epidemiol. 41, 200–209. https://doi.org/10.1093/ije/dyr238

Kelly, J.K., Hughes, K.A., 2019. Pervasive Linked Selection and Intermediate-Frequency Alleles Are Implicated in an Evolve-and-Resequencing Experiment of *Drosophila simulans*. Genetics 211, 943–961. https://doi.org/10.1534/genetics.118.301824

King, Elizabeth G., Macdonald, S.J., Long, A.D., 2012. Properties and Power of the Drosophila Synthetic Population Resource for the Routine Dissection of Complex Traits. Genetics 191, 935–949. https://doi.org/10.1534/genetics.112.138537

King, E. G., Merkes, C.M., McNeil, C.L., Hoofer, S.R., Sen, S., Broman, K.W., Long, A.D., Macdonald, S.J., 2012. Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. Genome Res. 22, 1558–1566. https://doi.org/10.1101/gr.134031.111

Kofler, Orozco-terWengel, P., De Maio, N., Pandey, R.V., Nolte, V., Futschik, A., Kosiol, C., Schlötterer, C., 2011a. PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. PLoS ONE 6, e15925. https://doi.org/10.1371/journal.pone.0015925

Kofler, Pandey, R.V., Schlotterer, C., 2011b. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27, 3435–3436. https://doi.org/10.1093/bioinformatics/btr589

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Marigorta, U.M., Rodríguez, J.A., Gibson, G., Navarro, A., 2018. Replicability and Prediction: Lessons and Challenges from GWAS. Trends Genet. 34, 504–517. https://doi.org/10.1016/j.tig.2018.03.005

Marquez, E.J., Houle, D., 2015. Dimensionality and the statistical power of multivariate genome-wide association studies (preprint). Genomics. https://doi.org/10.1101/016592

Marriage, T.N., King, E.G., Long, A.D., Macdonald, S.J., 2014. Fine-Mapping Nicotine Resistance Loci in *Drosophila* Using a Multiparent Advanced Generation Inter-Cross Population. Genetics 198, 45–57. https://doi.org/10.1534/genetics.114.162107

Martin, A., Orgogozo, V., 2013. The Loci of Repeated Evolution: A Catalog of Genetic Hotspots of Phenotypic Variation. Evolution 67, 1235–1250. https://doi.org/10.1111/evo.12081

McGuigan, K., 2006. Studying phenotypic evolution using multivariate quantitative genetics: EVOLUTIONARY QUANTITATIVE GENETICS. Mol. Ecol. 15, 883–896. https://doi.org/10.1111/j.1365-294X.2006.02809.x

Melo, D., Marroig, G., Wolf, J.B., 2019. Genomic Perspective on Multivariate Variation, Pleiotropy, and Evolution. J. Hered. 110, 479–493. https://doi.org/10.1093/jhered/esz011

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

Palsson, A., Gibson, G., 2004. Association Between Nucleotide Variation in Egfr and Wing Shape in Drosophila melanogaster. Genetics 167, 1187–1198. https://doi.org/10.1534/genetics.103.021766

Pan, D., 2007. Hippo signaling in organ size control. Genes Dev. 21, 886–897. https://doi.org/10.1101/gad.1536007

Pitchers, W., Nye, J., Márquez, E.J., Kowalski, A., Dworkin, I., Houle, D., 2019. A Multivariate Genome-Wide Association Study of Wing Shape in *Drosophila melanogaster*. Genetics 211, 1429–1447. https://doi.org/10.1534/genetics.118.301342

Porter, H.F., O'Reilly, P.F., 2017. Multivariate simulation framework reveals performance of multi-trait GWAS methods. Sci. Rep. 7, 38837. https://doi.org/10.1038/srep38837

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Reddiex, A.J., Chenoweth, S.F., 2021. Integrating genomics and multivariate evolutionary quantitative genetics: a case study of constraints on sexual selection in *Drosophila serrata*. Proc. R. Soc. B Biol. Sci. 288, 20211785. https://doi.org/10.1098/rspb.2021.1785

Rohlf, F.J., Slice, D., 1990. Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. Syst. Zool. 39, 40. https://doi.org/10.2307/2992207

Schluter, D., 1996. Adaptive Radiation Along Genetic Lines of Least Resistance. Evolution 50, 1766. https://doi.org/10.2307/2410734

Shriner, D., 2012. Moving toward System Genetics through Multiple Trait Analysis in Genome-Wide Association Studies. Front. Genet. 3. https://doi.org/10.3389/fgene.2012.00001

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

Steiner, C.C., Weber, J.N., Hoekstra, H.E., 2007. Adaptive Variation in Beach Mice Produced by Two Interacting Pigmentation Genes. PLoS Biol. 5, e219. https://doi.org/10.1371/journal.pbio.0050219

Stephens, M., 2013. A Unified Framework for Association Analysis with Multiple Related Phenotypes. PLoS ONE 8, e65245. https://doi.org/10.1371/journal.pone.0065245

Stern, D.L., Orgogozo, V., 2008. THE LOCI OF EVOLUTION: HOW PREDICTABLE IS GENETIC EVOLUTION? Evolution 62, 2155–2177. https://doi.org/10.1111/j.1558-5646.2008.00450.x

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., Meyre, D., 2019. Benefits and limitations of genome-wide association studies. Nat. Rev. Genet. 1. https://doi.org/10.1038/s41576-019-0127-1

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., Yang, J., 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am. J. Hum. Genet. 101, 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Wray, N.R., Wijmenga, C., Sullivan, P.F., Yang, J., Visscher, P.M., 2018. Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. Cell 173, 1573–1580. https://doi.org/10.1016/j.cell.2018.05.051

Wright, S., 1934. Physiological and Evolutionary Theories of Dominance. Am. Nat. 68, 24–53. https://doi.org/10.1086/280521

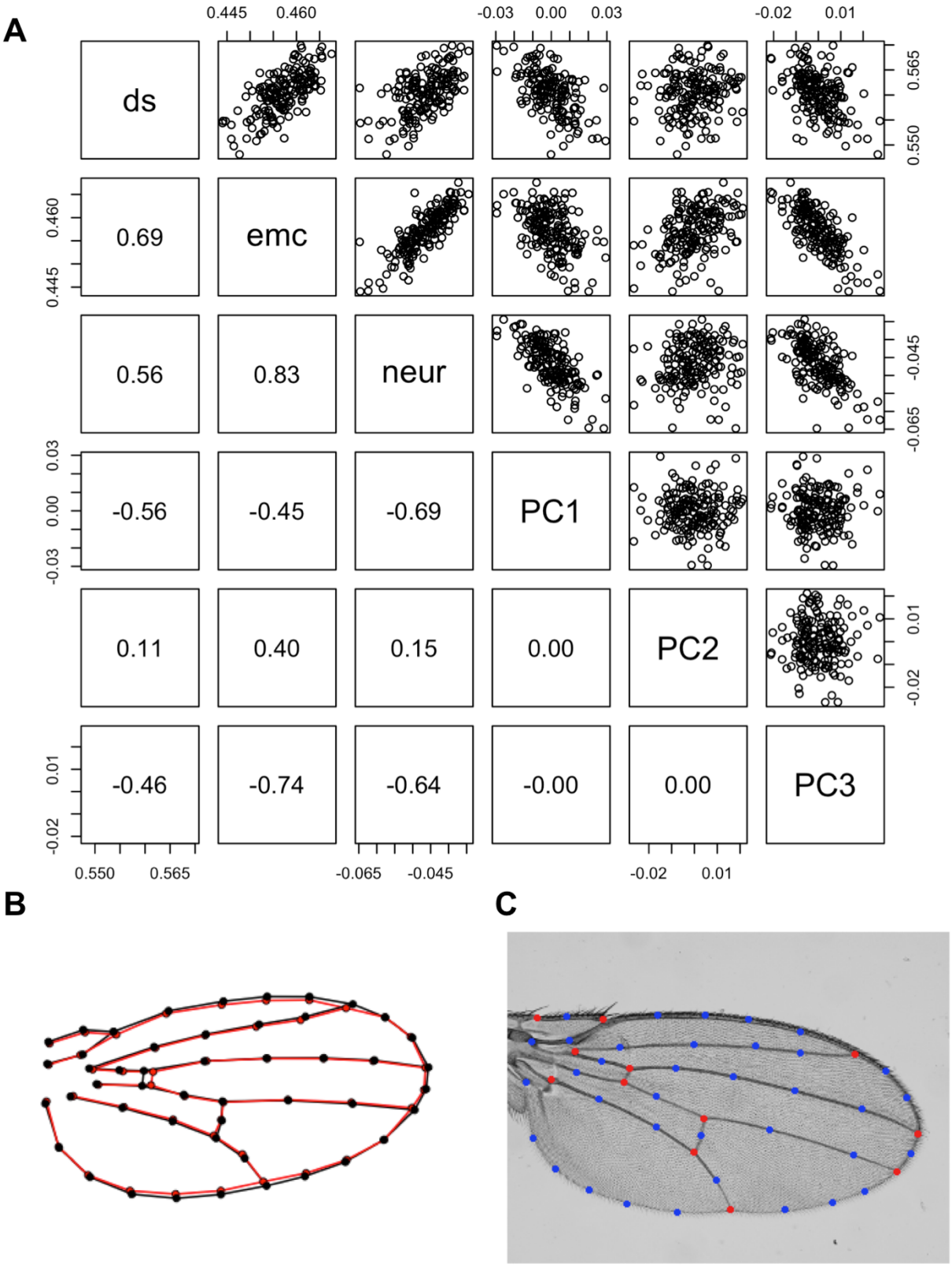


Figure 1. Shape change vectors are correlated with directions of shape variance in DGRP panel. (A) Shape change vectors for *ds, emc* and *neur* were used for projections for DGRP lines to calculate shape scores. Eigenvectors for PCA were estimated based on the same estimated line means for shape. (B) Effect of *ds* shape change estimated from RNAi knockdown, effect size is scaled by 0.5. (C) Landmarks (red) and semi-landmarks (blue) used in geomorphic morphometric analysis.

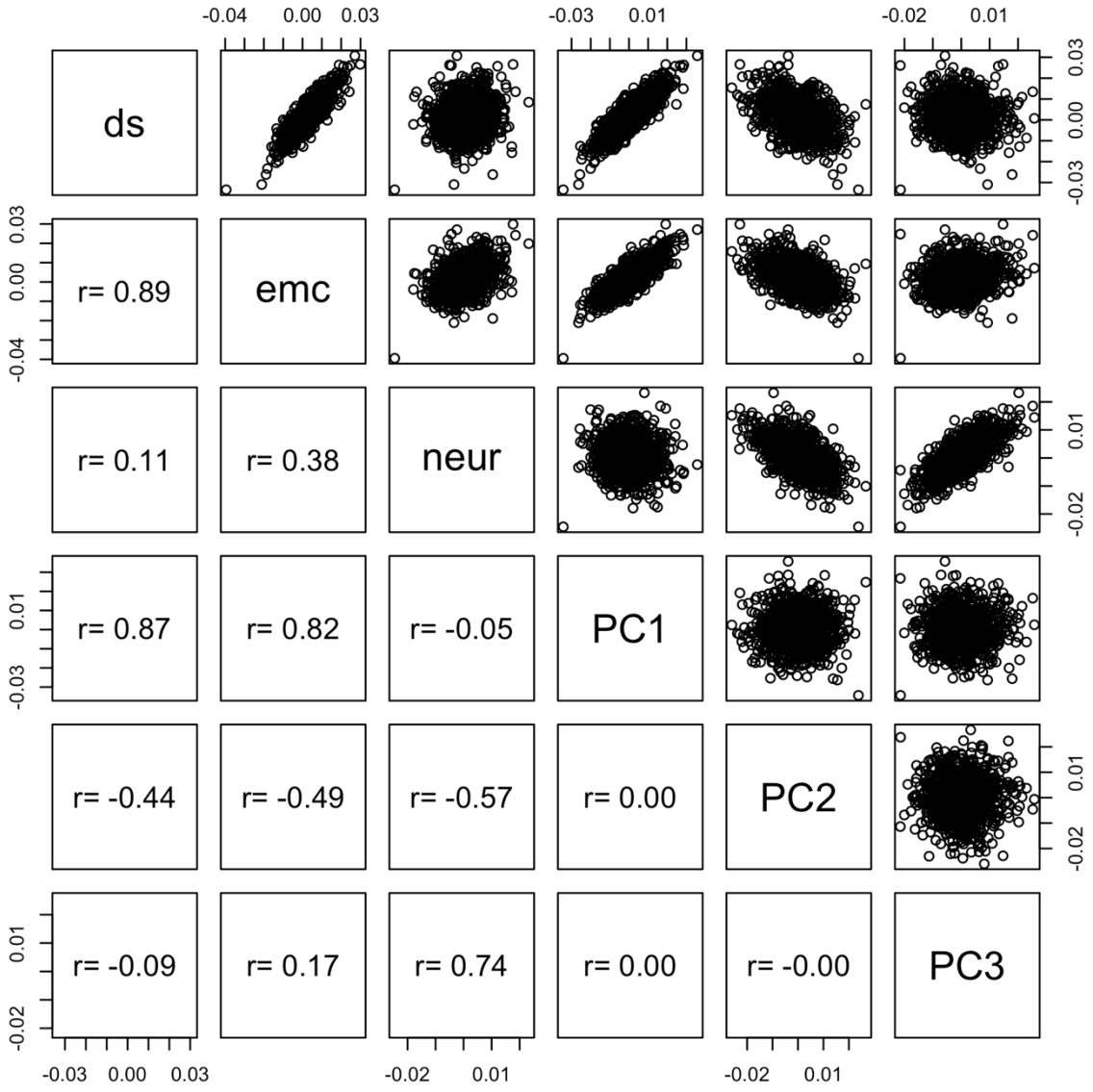


Figure 2. Relationship between shape change vectors and eigen vectors in wild cohort. Correlation between projection of shape data from CMO population onto *ds, emc* and *neur* shape change vectors and the first three eigenvectors are calculated from shape data from all wings in the CMO population for PCA.

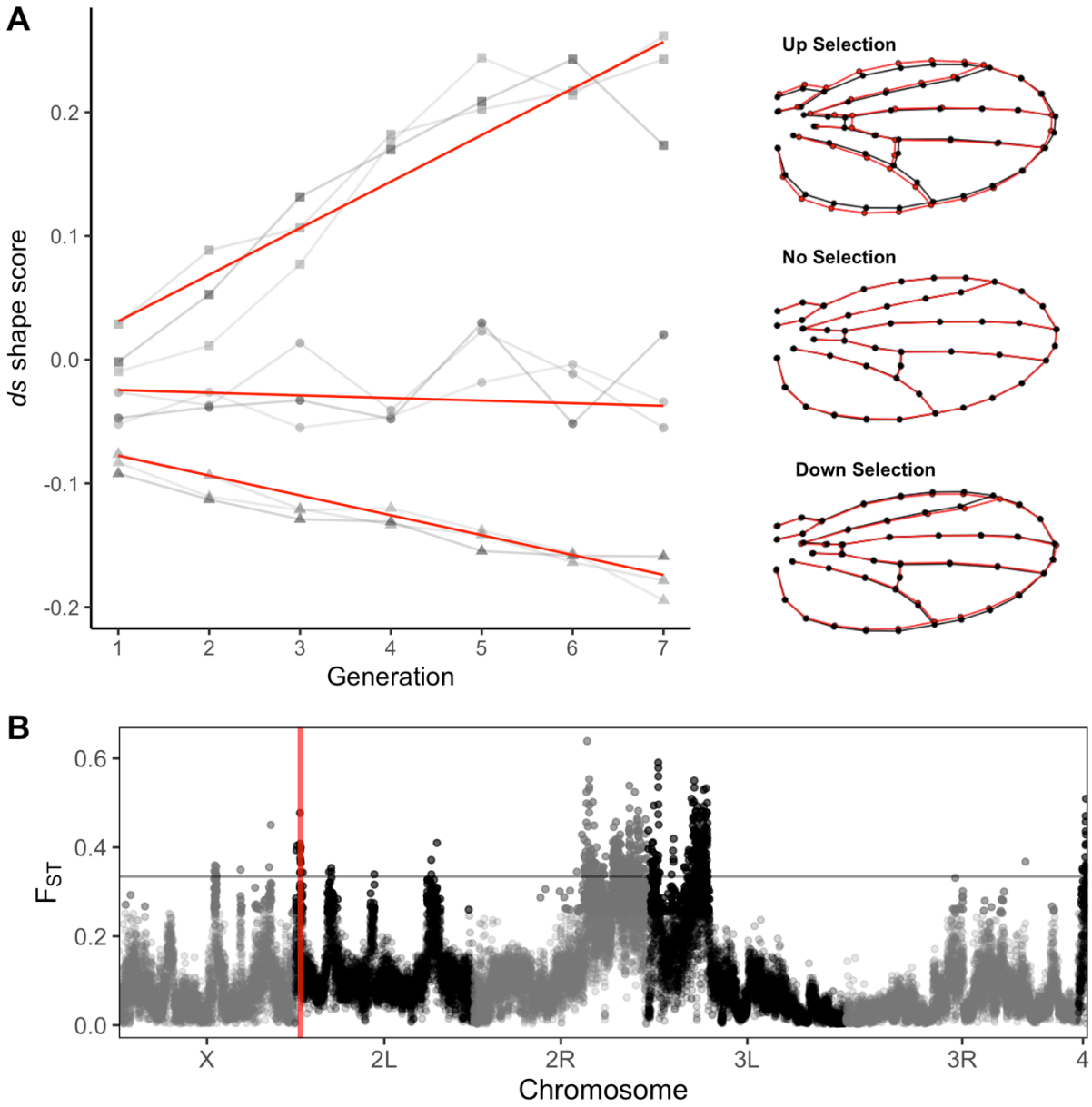


Figure 3. Artificial selection based on *ds* shape change vector. (A) Phenotypic response to selection based on the *ds* shape change vector. Each replicate of up (squares), control (dots) and down (triangles) selection lineages are plotted in greys. Red lines indicate the estimated response to selection. Wing plots represent the effect of selection (red). (B) Genomic differentiation between up and down selection lineages (F_{ST}) measured in 5000bp sliding windows. Red line represents the location of *ds.* Grey line represents 3sd from genome wide mean F_{ST}.

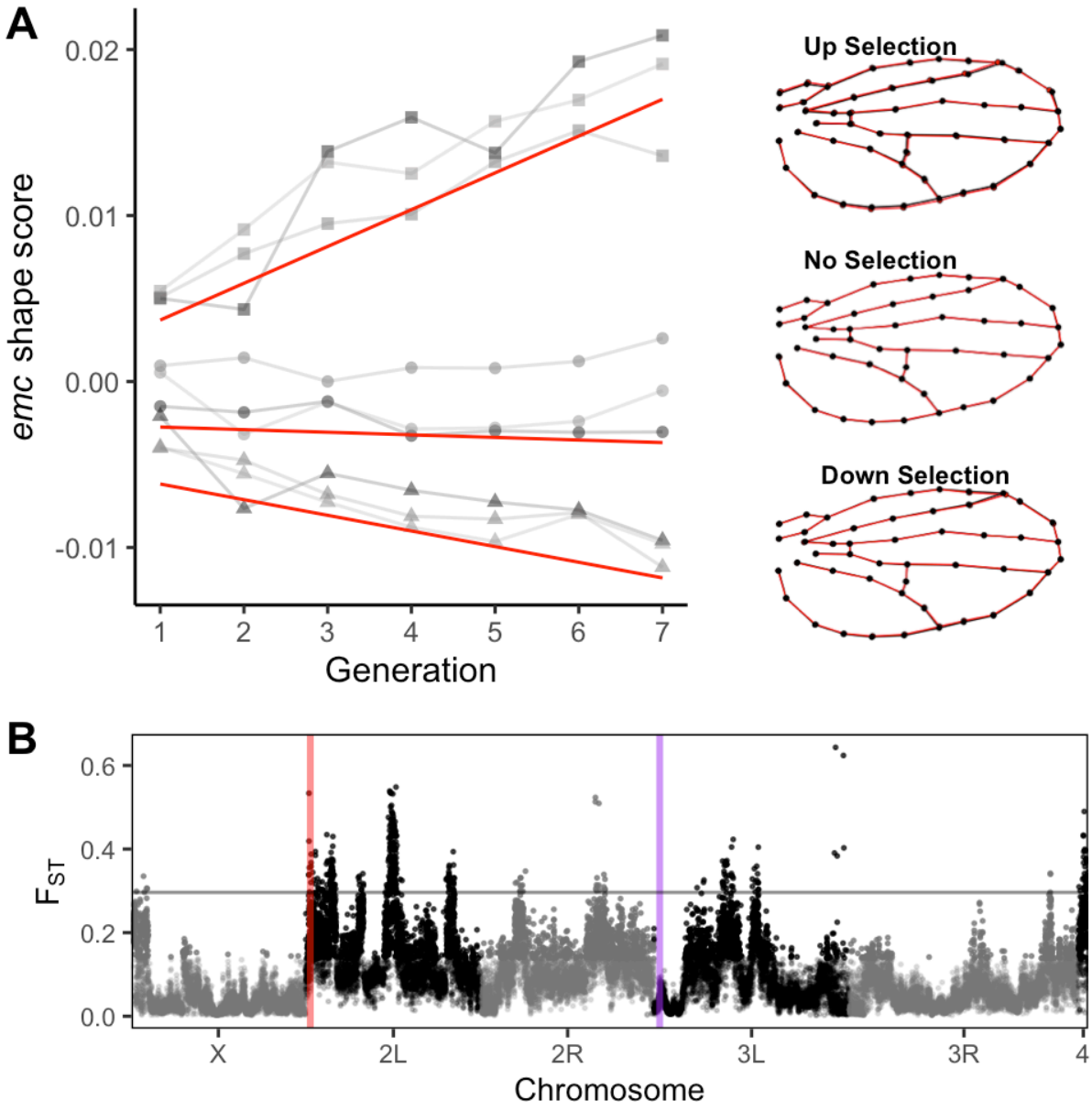


Figure 4. Artificial selection based on *emc* shape change vector. (A) Phenotypic response to selection based on the *ds* shape change vector. Each replicate of up (squares), control (dots) and down (triangles) selection lineages are plotted in greys. Red lines indicate the estimated response to selection. Shape change between generation 1 and 7 is indicated on the right. Effects have been magnified 5x for visualization. (B) Genomic differentiation between up and down selection lineages (F_{ST}) measured in 5000bp sliing windows. Red line represents the location of *ds,* purple line represents the location of *emc.* Grey line represents 3sd from genome wide mean F_{ST}.

.

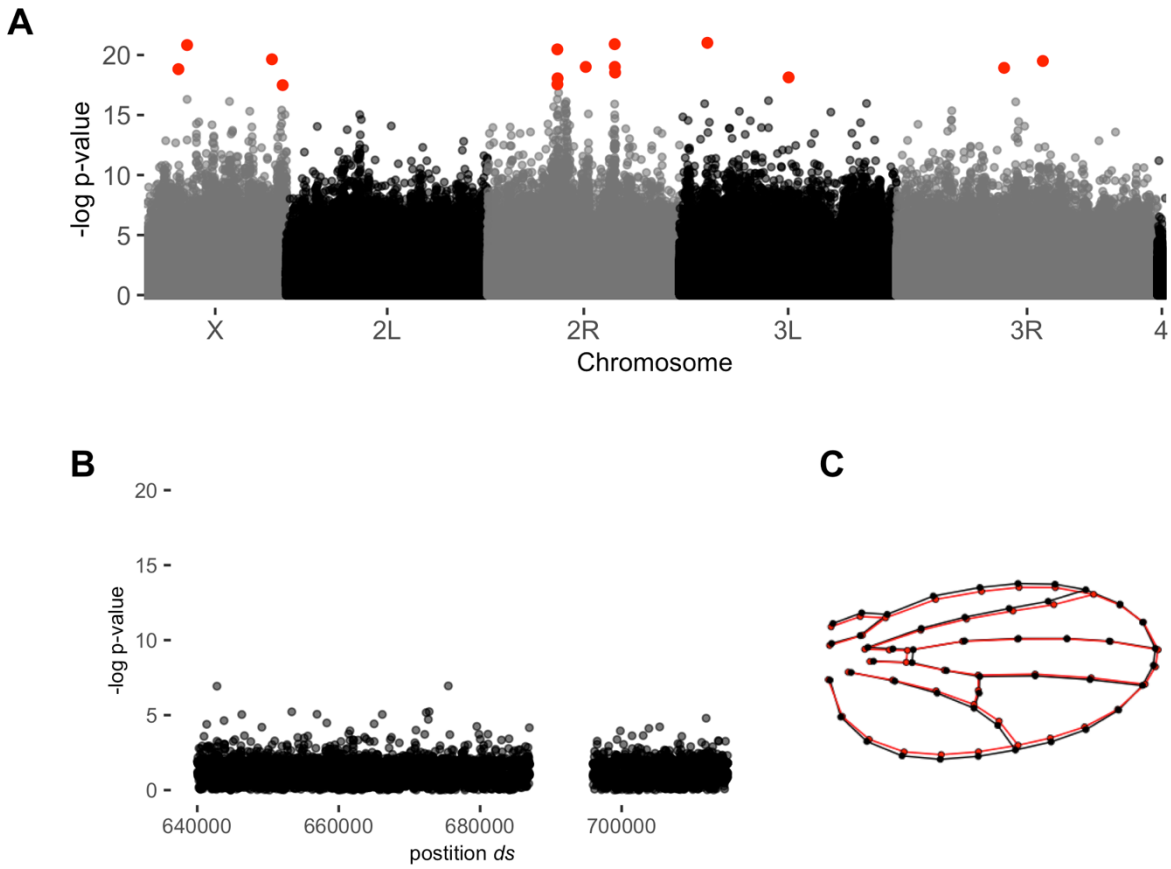


Figure 5. Differentiation between pools selected based on *ds* shape change. (A) Genome-wide scan for differentiated loci between pools selected based on *ds* shape change vector using the CMH test. Points in red indicate sites with significant differentiation based on p-value with FDR of 0.05. (B) Genomic differentiation at *ds* between pools selected based on *ds* shape change vector. No sites are significantly differentiated. (C) Shape difference between selected pools. CMO population used as example. Mean left pool shape in red and mean right pool shape in black.

Supplemental Table 1. Number of individuals in wild-caught cohorts for BSA analysis

Collection	Number of Females Phenotyped	Number of Males Phenotyped	Number of Females – "up" pool	Number of Females – "down" pool	Number of Males – "up" pool	Number of 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

Table 1. Variants from Pitchers (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.

Variant	Estimated Effect (Pitchers 2019)	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	**0.159**	**0.056**	**0.06**	**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)

Supplemental Table 2. Top 50 enriched GO terms of linked differentiated sites following artificial selection based on *ds* shape change

SEE EXCEL SHEET.

Supp. Table 3. Top 50 enriched GO terms of linked differentiated sites following artificial selection based on *emc* shape change

SEE EXCEL SHEET

Table 2. Variants from Pitchers (2019) in *ds* wild populations. Estimated effect sizes for SNPs are estimated from a GWAS in the DGRP with LASSO regularized coefficients. MAF in wild cohorts was estimated from pools of 75 random individuals that were sequenced.

Variant	Estimated Effect (Pitchers 2019)	DGRP MAF	Estimated MAF CMO	Estimated MAF FVW12	Estimated MAF FVW14	Estimated MAF PHO
2L:655894	0.072	0.445	0	0	0	0
2L:702560	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

Supp. Table 4. Pairwise Procrustes (Euclidian approximation) distances between mean shapes of 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.

Population Comparison	Distance	Upper confidence limit (95%)	Z-score	p
CMO-FVW12	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**
FVW12-FVW14	0.0065	0.0012	7.25	0.001
FVW12-PHO	**0.022**	**0.0012**	**13.0**	**0.001**
FVW12-PHO	**0.022**	**0.0014**	**13.5**	**0.001**

Table 3. Significantly differentiated sites between wild caught pools selected based on *ds* shape change.

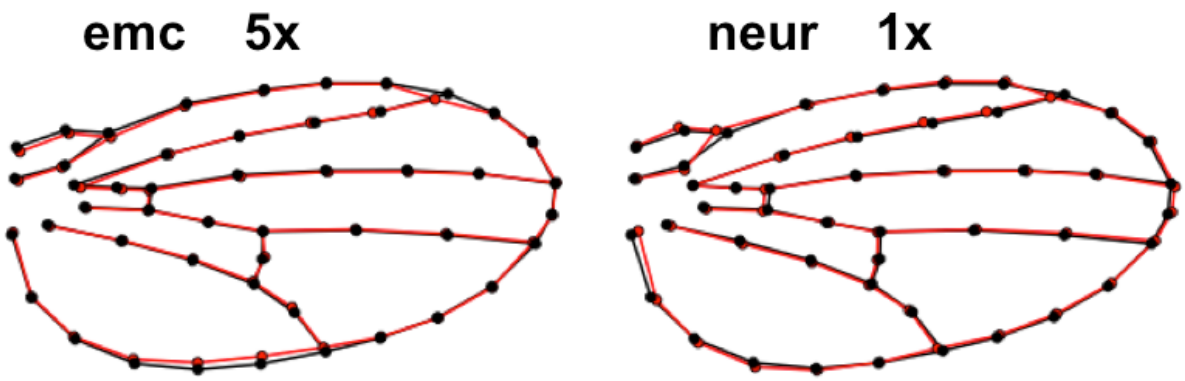
Location	CMH p-value (FDR corrected)	Gene	FB ID	Distance from ORF (bp)
2R:17491270	0.026	NT5E-2	FBgn0050104	0
2R17498059	0.034	CG30103	FBgn0050103	2061
2R:17515133	0.022	CG4853	FBgn0034230	0
2R: 20537878	0.013	CG13423	FBgn0034513	0
2R:23601278	0.005	CG10332	FBgn0260455	0
2R:23601278	0.005	IM18	FBgn0067903	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

Sup table 5. PHO left out differentiated sites.
See file

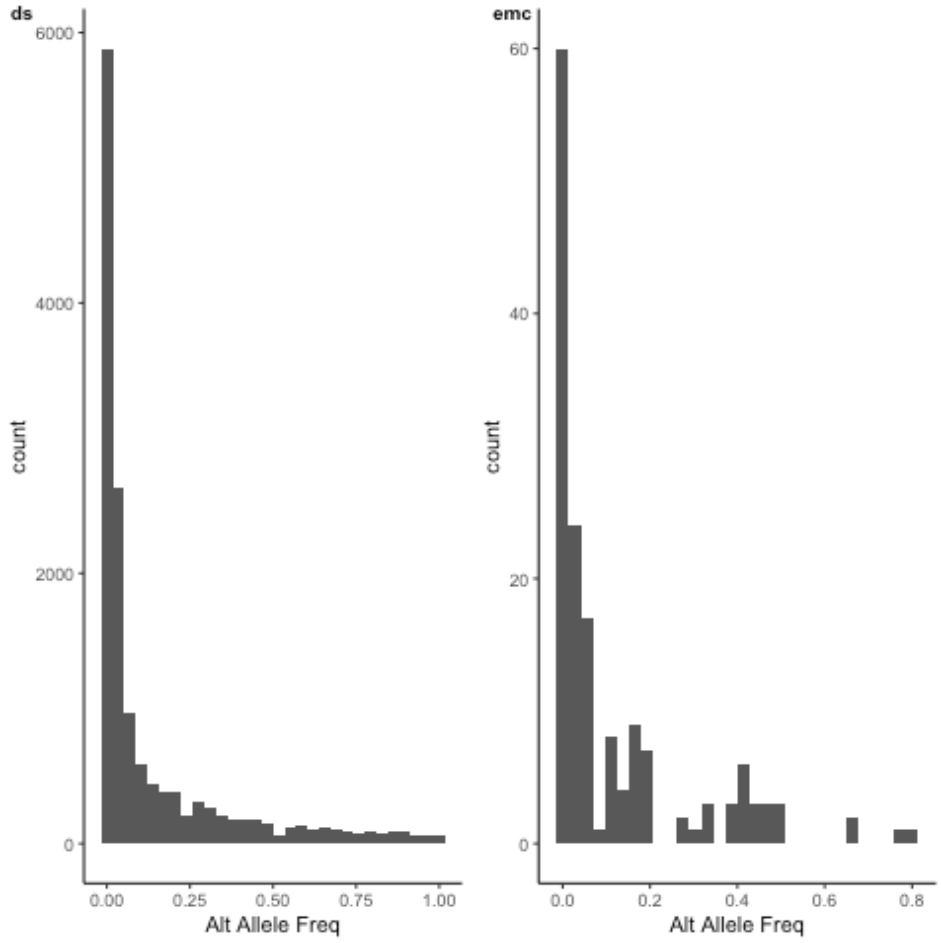
Sup table 6. GO analysis of sites for PHO left out
See file

Sup table 7. Significantly differentiated sites between wild caught pools selected based on *neur* shape change.

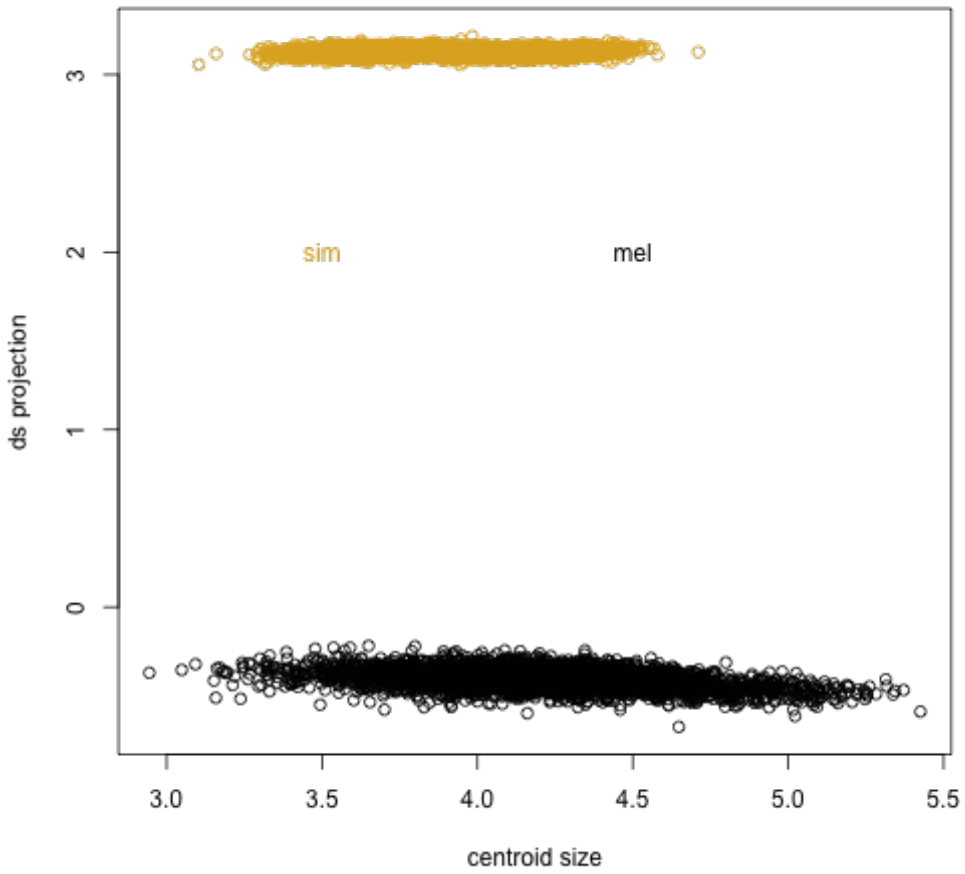
Location	CMH p-value (FDR corrected)	Gene	FB ID	Distance from ORF (bp)
2L:15775767	0.026	CG43760	FBgn0264260	675
2L:15967220	0.017	Beat-Ib	FBgn0028645	4785
2L:15967222	0.017	Beat-Ib	FBgn0028645	4787
2L:8651597	0.015	Sema1a	FBgn0011259	0
3R: 21898160	0.024	CG6678	FBgn0038917	0



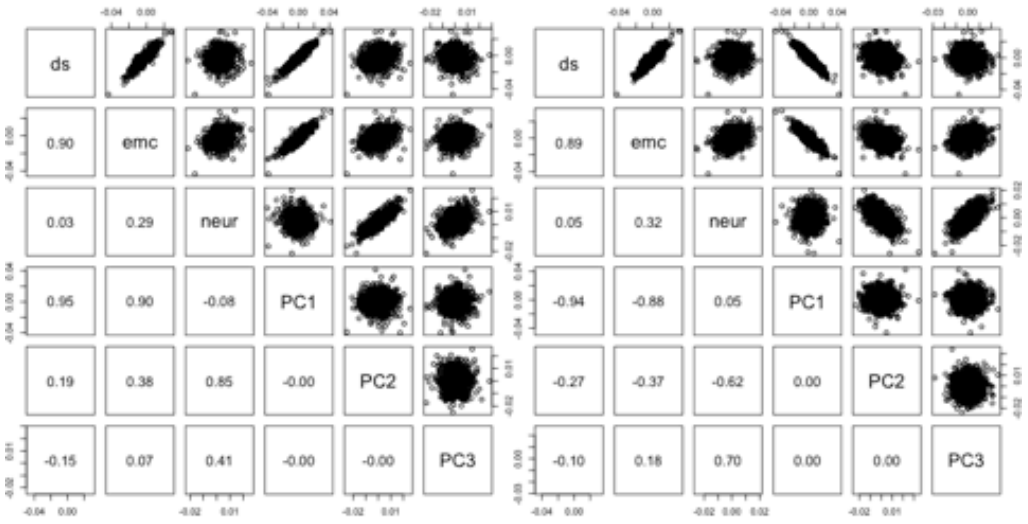
Supplemental Figure 1. RNAi knockdown shape change effects of *emc* and *neur. emc* effect is scaled by 5x and *neur* effect is scaled by 1x for better visualization.



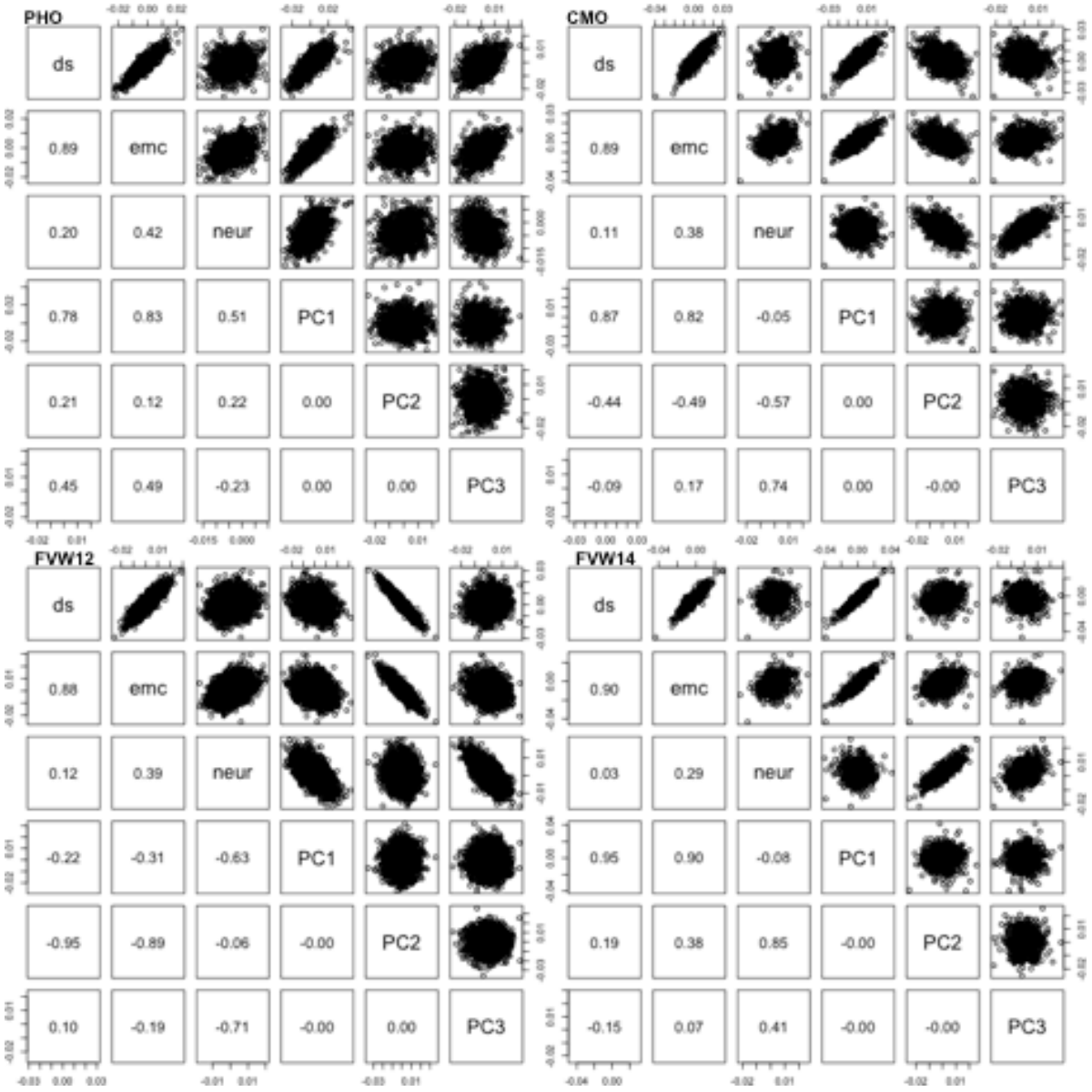
Supplemental Figure 2. Allele Frequency spectra. Estimated alternate allele frequencies at *ds* and *emc* in synthetic outbred population. Alternate allele frequencies are estimated using parental strain genotype data and assuming an equal contribution from each parent to the founding population. Note the different axis scale between genes.



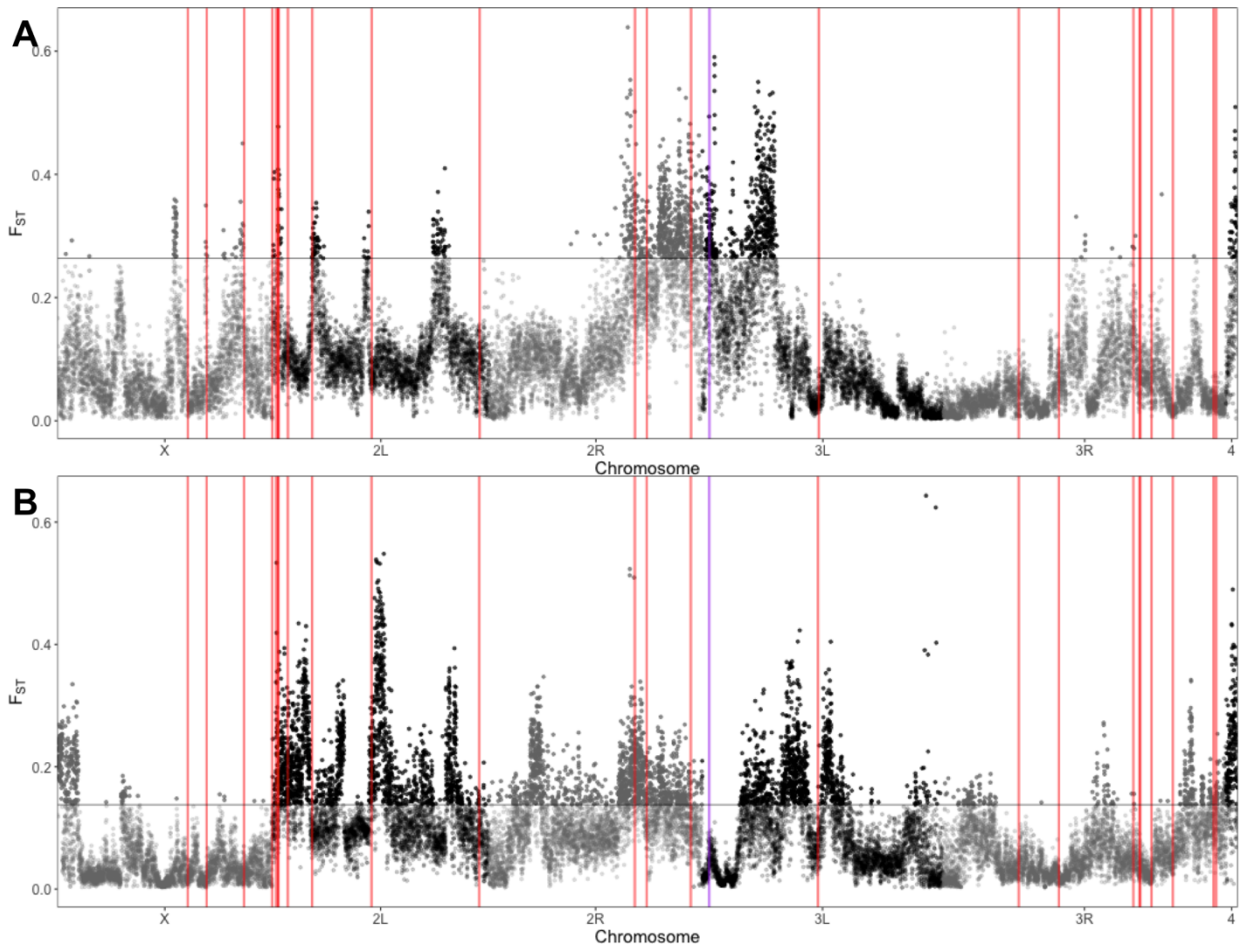
Supplemental Figure 3 Projection of FVW12 *D. melanogaster* wings onto *ds* shape change vector is distinct from *D. simulans.* Clear separation between the FVW14 samples (black) and *D. simulans* data (gold) indicates that *D. melanogaster* females were accurately identified.



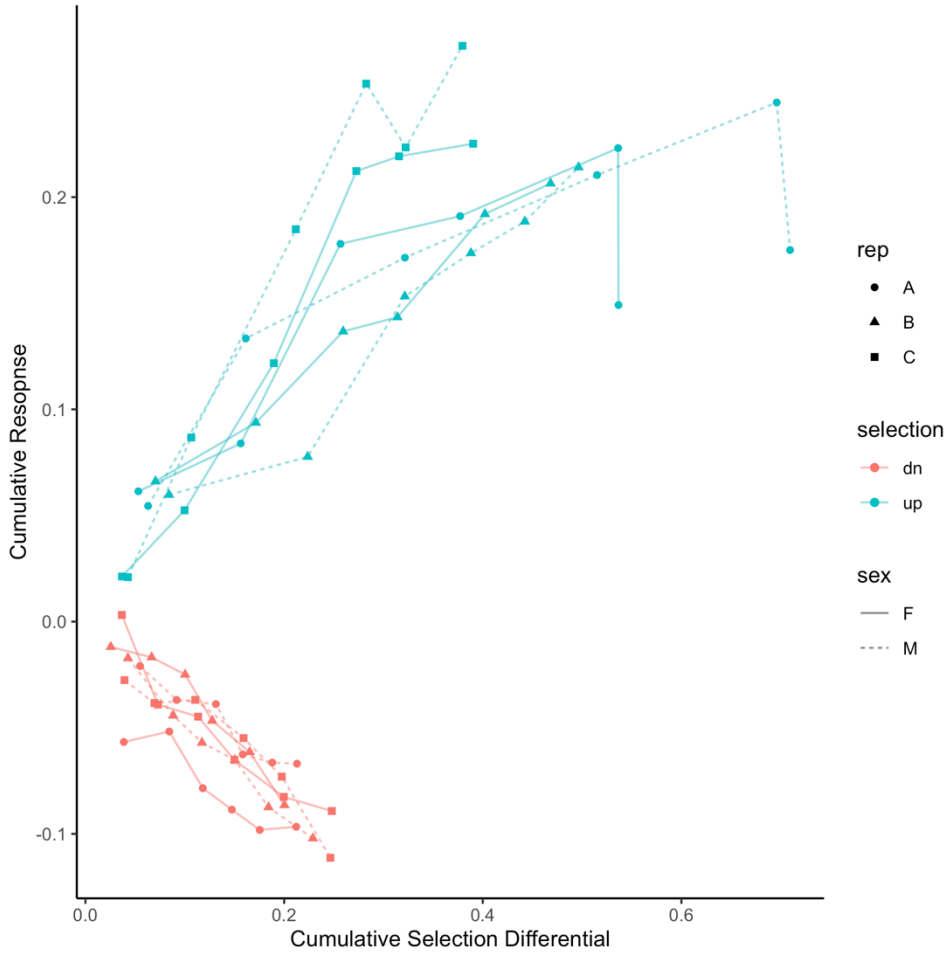
Supplemental Figure 4. Including females in analysis does not change interpretation for FVW14 data. Correlation between projection of shape data from FVW14 population onto *ds, emc* and *neur* shape change vectors and the first three eigenvectors are calculated from shape data from all wings in the FVW14 population for PCA. Females are included in panel B and do not change the conclusions drawn from the relationships between directions shape change vectors and PCs.



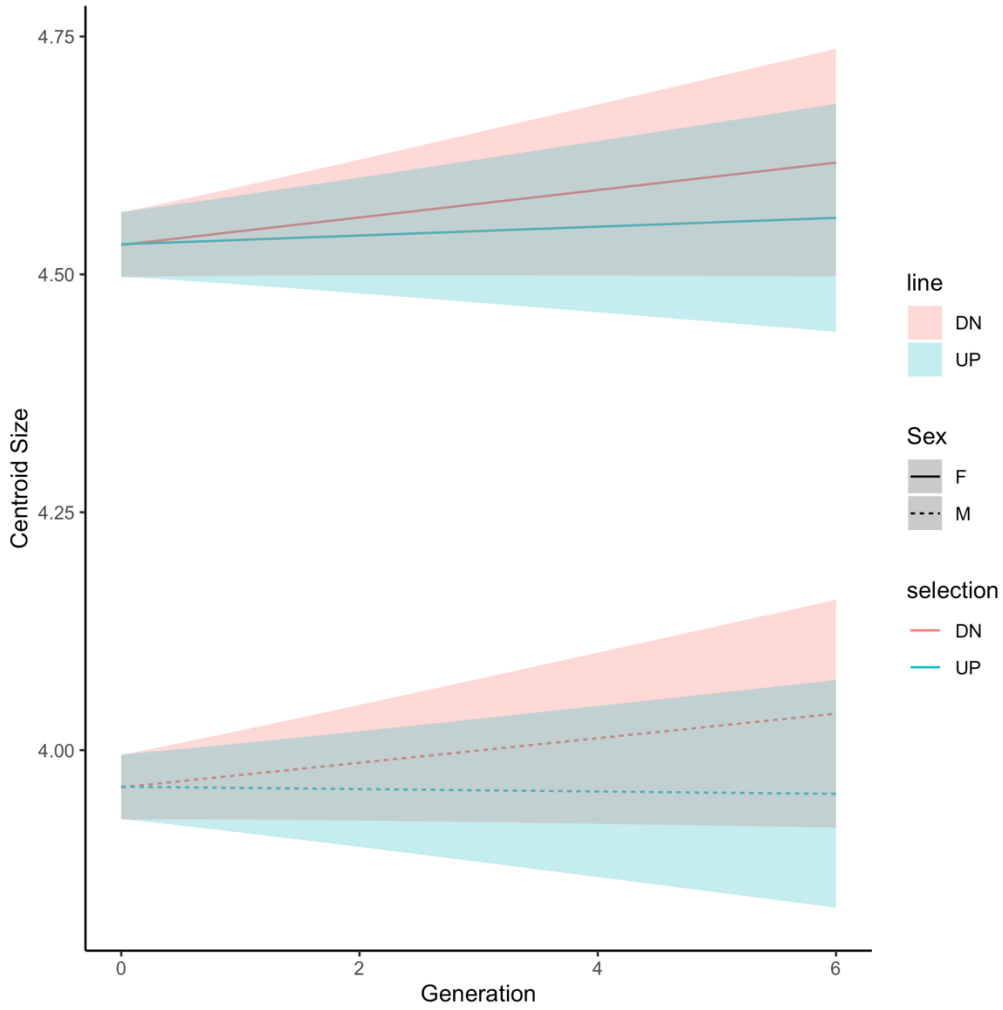
Supplemental Figure 5. Correlation between projection of shape data from wild cohorts onto *ds, emc* and *neur* shape change vectors and the first three eigenvectors are calculated from shape data from all wings in each cohort independently.



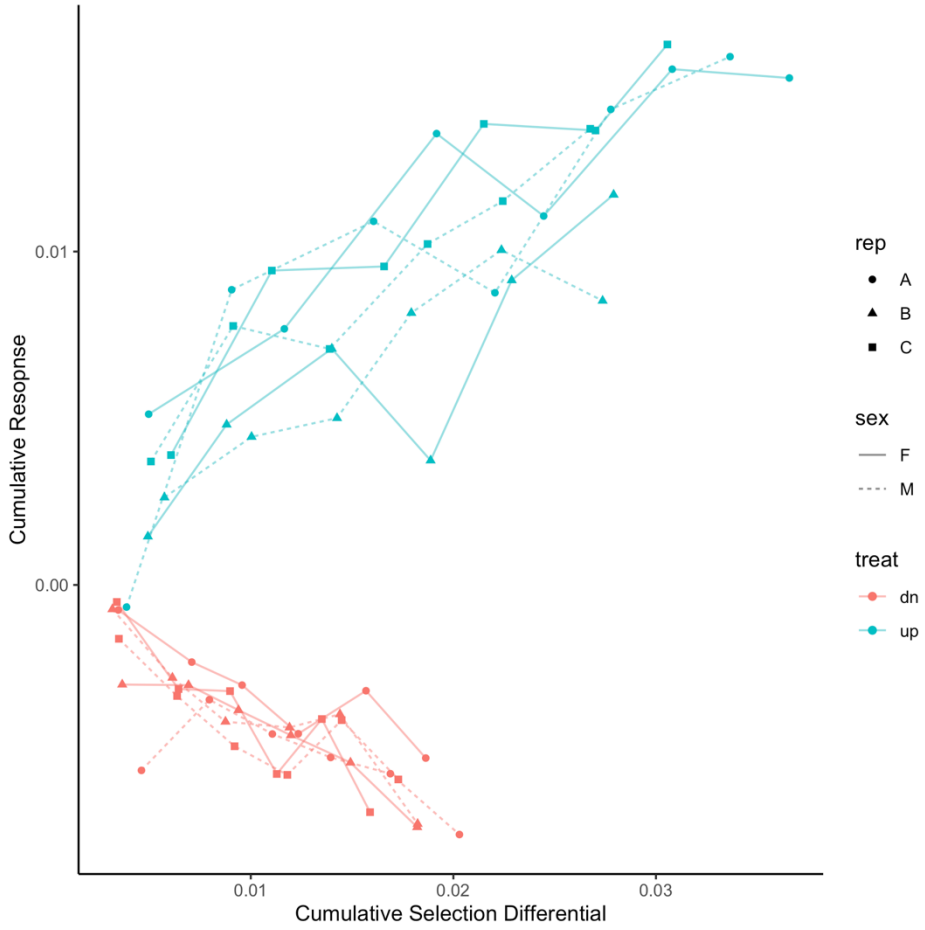
Supplemental Figure 6. Genetic differentiation between artificial selection pools with core hippo signaling loci (red) and *emc* (purple) marked. Genomic differentiation between up and down selection lineages (F_{ST}) measured in 5000bp sliding windows for *ds* selection (A) and *emc* selection (B). Horizonal line represents 3sd from the mean.



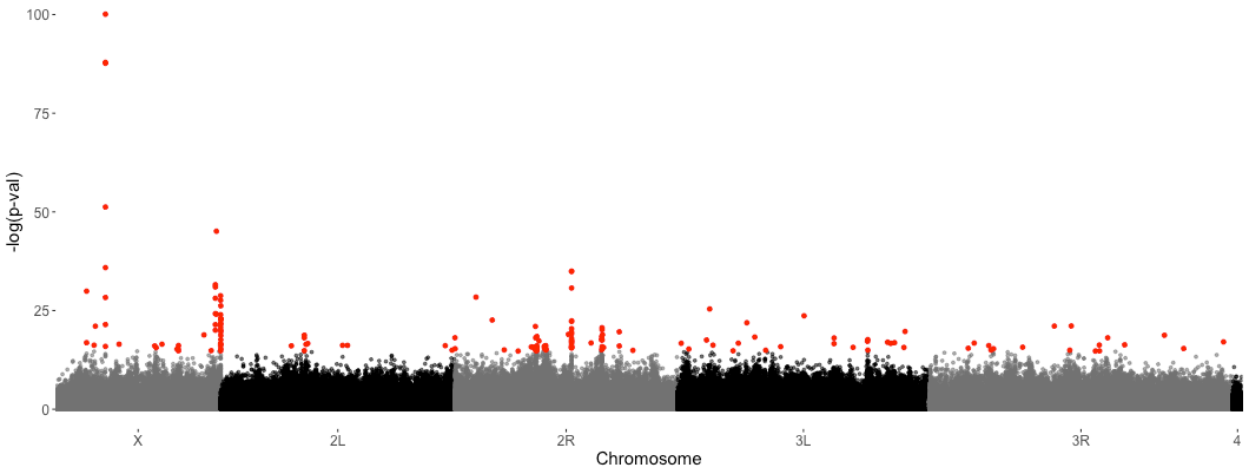
Supplemental Figure 7. Response to selection based on *ds* shape change. Regression of cumulative selection differential onto cumulative response was used to estimate realized heritability in both lineages independently.



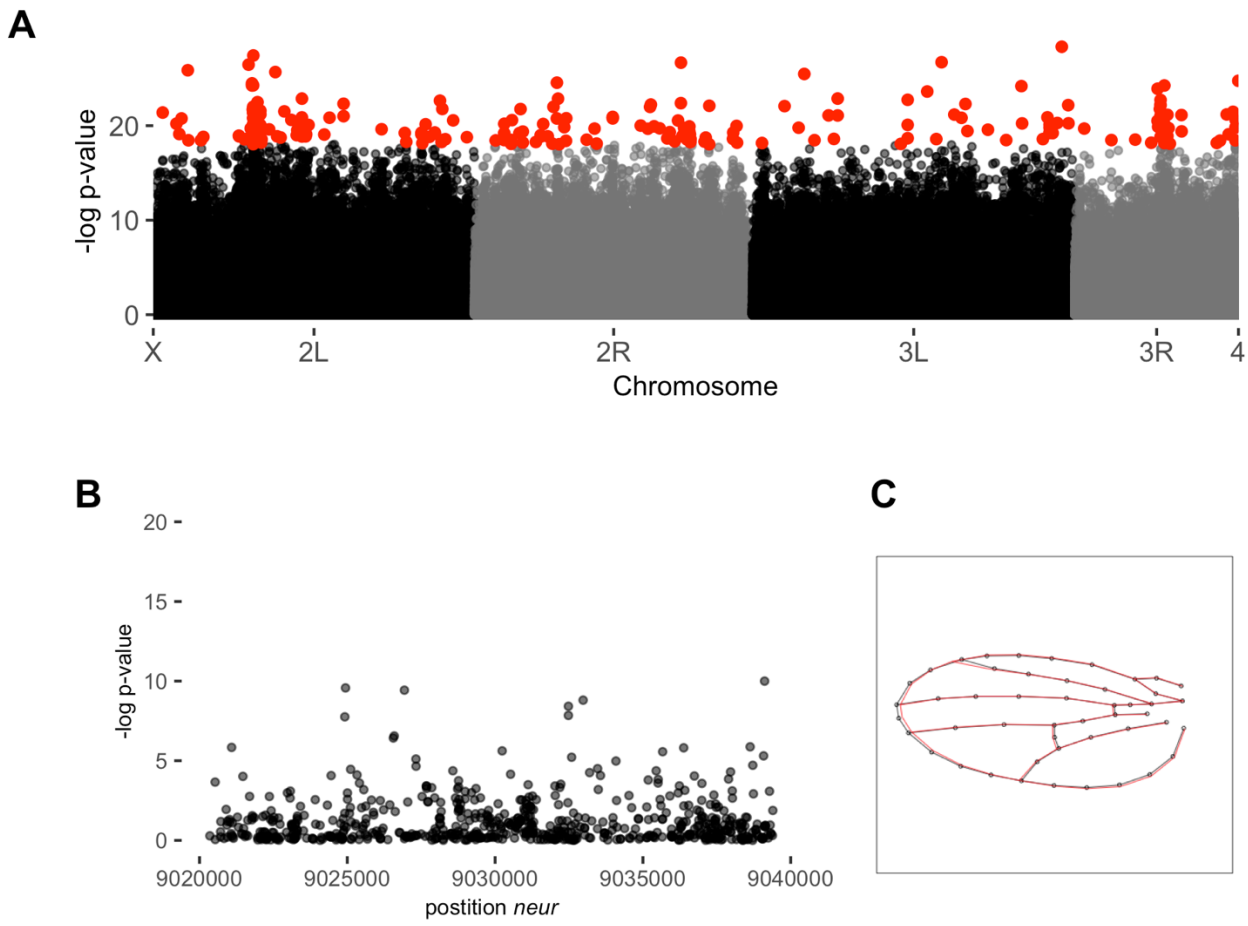
Supplemental Figure 8. No consistent and substantial change in wing size following artificial selection based on *ds* shape change. Mean size estimated for up and down *ds* selection lineages estimated from a linear mixed model with replicate lineage fitted as a random effect. Shaded regions represent 95% confidence bands for the correlated response to selection.



Supplemental Figure 9. Response to selection based on *emc* shape change. Regression of cumulative selection differential onto cumulative response was used to estimate realized heritability in both lineages independently.



Supplemental Figure 10. Removing PHO from genomic 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 with the ACER package in R.. PHO data is excluded from analysis. Points in red indicate sites with significant differentiation based on p-value with a FDR of 0.05.



Supplemental figure 11. Genome-wide scan for differentiated loci between pools selected based on *neur* shape change vector using the CMH test implemented with the ACER package in R. (A) Whole genome scan for differentiation. Points in red indicate sites with significant differentiation based on p-value with a FDR of 0.05. (B) No differentiated sites within *neur.* (C) shape change between selected pools based on *neur* shape change vector.