Genetic and environmental canalization are not associated among altitudinally varying populations of *Drosophila melanogaster*

Maria Pesevski[1] and Ian Dworkin[1,2]

[1]Department of Biology, McMaster University, Hamilton, Ontario, L8S 4K1, Canada

[2]Corresponding author, dworkin@mcmaster.ca

July 25, 2019

running title: Environmental canalization in diverged populations

Author contributions: Experimental design - MP & ID, Ran experiments - MP, Analysis - MP & ID, Writing and editing - MP & ID

Data Accessibility: Upon publication all data and scripts will be made available on both github and Dryad.

Abstract

Organisms are exposed to environmental and mutational effects influencing both mean and variance of phenotypes. These potentially deleterious effects can be ameliorated by the evolution of buffering (canalizing) mechanisms, resulting in reduced phenotypic variation. Under some population genetics models, the circumstances under which genetic canalization evolves is limited. As such, it has been argued that canalizing mechanisms for environmental and mutational stresses may co-evolve, and will be correlated. Yet, empirical evidence has not consistently supported this prediction. In a recent study, a population of *Drosophila melanogaster* adapted to high altitude showed evidence of genetic decanalization relative to low-altitude populations. Using these populations, we tested if these populations also varied for environmental canalization. We reared strains derived from these populations under different temperature rearing environments. Using the *Drosophila* wing, we quantified size, shape, cell density and frequencies of mutational defects. We observed the expected differences in size, shape, cell density and mutational defects between the high- and low-altitude populations. However, we observed little evidence for a relationship between a number of measures of environmental canalization with population of origin or visible defect frequency. Our results do not support the predicted association between genetic and environmental canalization.

Key Words

Canalization, *Drosophila*, Adaptation, high-altitude, Wing shape, Body size, Integration, Geometric morphometrics, Cell density

Introduction

In addition to differences in trait means, there can be considerable variation in how much variation is observed among individuals of a given genotype (Waddington, 1942; Rendel, 1963; Félix and Barkoulas, 2015; Flatt, 2005; Visser et al., 2003; Siegal and Leu, 2014; Klingenberg, 2019; Mayer and Hansen, 2017; Pélabon et al., 2010; Gibson and Wagner, 2000). Theoretical work has examined this propensity to vary with respect to the evolution of phenotypic robustness or canalization. Such evolved properties are important to examine, as environmental and mutational variation influence trait variance, ultimately influencing organismal performance and fitness (Arnold, 2003, 1983). The sensitivity of a given genotype in its response to mutational or environmental influences can vary *among* genotypes. In particular, empirical evidence has demonstrated that under mutational or environmental perturbation, there is often the expression of cryptic genetic variation which has previously been used as evidence for genetic canalization (Paaby et al., 2015; Paaby and Rockman, 2014; Scharloo, 1991; Gibson and van Helden, 1997; Dworkin, 2005a). Theory suggests that robustness to environmental variation — environmental canalization — can readily evolve as organisms are constantly exposed to the influence of environmental effects (Eshel and Matessi, 1998; Wagner et al., 1997). Yet, as deleterious mutations are often purged by natural selection, this can result in weak selection for genetic canalization, making it difficult to evolve (Wagner et al., 1997; Proulx and Phillips, 2005; Visser et al., 2003; Gibson and Wagner, 2000).

The congruence hypothesis was proposed in order to solve this apparent inconsistency between theoretical and empirical work regarding the evolution of genetic canalization (Wagner et al., 1997). This hypothesis predicts that genetic canalization evolves as a correlated response during selection for environmental canalization (Wagner et al., 1997; Gibson and

Wagner, 2000; Visser et al., 2003). Empirical evidence for associations between genetic and environmental canalization is mixed. Some studies provide supporting evidence both from simulations (Ancel and Fontana, 2000; Siegal and Bergman, 2002; Shu et al., 2007) and empirical work (Lehner, 2010; Stearns and Kawecki, 1994; Stearns et al., 1995; Szöllősi and Derényi, 2009). However, some explicit tests for the congruence model did not find support for it (Dworkin, 2005a,c; Borenstein and Ruppin, 2006). The most likely explanation is that the evolution of genetic and environmental canalization are not homogeneous, given the complex interplay of selection, mutation rates, genetic architecture, and evolutionary history.

There are a number of important methodological and conceptual issues that influence the debate on the congruence hypothesis, and the study of the evolution of canalization more generally. For instance, the conditions in which the release of cryptic genetic variation can be used to infer genetic canalization has been discussed and its utility may be more limited than once thought (Hermisson and Wagner, 2004; Geiler-Samerotte et al., 2019). It is clear that mutation accumulation, or mutagenesis experiments, are likely to be more fruitful for investigating genetic canalization. Environmental canalization is often measured using multiple approaches that differ with respect to what aspects of environmental robustness they seek to capture. Within-individual variation (fluctuating asymmetry), among-individual within-genotype variation, and reaction norm of trait means under common or different environmental treatments (Dworkin, 2005b) have all been employed. Using these methods to study genetic and environmental canalization, some studies have seen modest evidence of association between degree of sensitivity to genetic perturbation (changes in trait means) and within- or among-individual variance within a genotype (Dworkin, 2005a,c; Chari and Dworkin, 2013; Camara and Pigliucci, 1999; Chandler et al., 2017). However, a number of other studies do not show a consistent relationship between magnitude of perturbation

and among-individual within-line or within-individual variance (Haber and Dworkin, 2017; Debat et al., 2009; Levy and Siegal, 2008). This suggests there are multiple, partially distinct properties when considering robustness for a given genotype.

Surprisingly, one issue that has not been broadly considered in the literature regarding the evolution of canalization is the influences of both lab domestication and the use of lab induced mutations. Most studies of canalization and robustness use adapted to lab environments (lab domestication), and use lab induced mutations as a source for genetic perturbations (Paaby et al., 2015; Gibson and van Helden, 1997; Dworkin, 2005a,c; Haber and Dworkin, 2017; Hallgrímsson et al., 2006; Debat et al., 2011; Levy and Siegal, 2008). However, lab domestication and lab induced mutations may be unrepresentative of natural populations (Dittmar et al., 2016; Rockman, 2008; Orgogozo et al., 2015). Lab induced mutations likely do not reflect the spectrum of mutational effects experienced by natural populations, providing an inaccurate examination of the ability of genotypes to buffer against the majority of mutations organisms are exposed to during their evolutionary history. Furthermore, when considering the evolution of canalization, the evolutionary history of the experimental populations matters. In some experimental studies, collections of natural lineages or families are used that have heterogeneous geographical origins, and/or have been maintained in the lab for long periods of time (i.e. Dworkin (2005a,c)). Thus, explicit tests for the correlated evolution of genetic and environmental canalization (*sensu* Wagner et al. (1997)) is difficult without knowledge of the evolutionary history of such populations.

Another shortcoming of many empirical studies examining properties of phenotypic robustness and canalization is that they have examined phenotypic variation in a decidedly univariate context (Dworkin, 2005a,c), even when examining many traits (Levy and Siegal, 2008; Takahashi et al., 2011, 2010; Takahashi, 2017). Considerable evidence and theory have demonstrated that a multivariate perspective on evolutionary change ($\Delta\bar{\mathbf{z}} = \mathbf{G}\beta$) improves

predictions and understanding of evolutionary responses to selection (Lande and Arnold, 1983; Lande, 1979; Schluter, 1996; Walsh and Blows, 2009; Houle et al., 2017; Blows and Mcguigan, 2014; Mcguigan and Blows, 2007; Pitchers et al., 2014; Agrawal and Stinchcombe, 2009; Hansen and Houle, 2008). Yet, this perspective has only been considered in a modest number of studies examining variational properties of phenotypes (Debat et al., 2009, 2011; Hallgrímsson et al., 2009, 2006; Green et al., 2017; Debat et al., 2006; Breuker et al., 2006; Cheverud et al., 1983; Pavlicev et al., 2009). In a recent study examining variation in $\mathbf{E}$ (from $\mathbf{P} = \mathbf{G} + \mathbf{E}$) across both naturally derived strains and lab induced mutations, it was demonstrated that some properties, such as changes to mean shape and matrix orientation, were much more variable than matrix eccentricity (Haber and Dworkin, 2017). A multivariate perspective needs to be consistently applied to studies examining trait variation (Klingenberg, 2019).

Thus what has been lacking for studies testing evolutionary models of canalization is a model system with the relevant natural history that can be studied in a multivariate perspective. Lack et al. (2016a), among other recent studies, demonstrated that populations of *Drosophila melanogaster* from sub-Saharan Africa recently adapted to a high-altitude environment via increased body size and wing morphology, among other traits (Pitchers et al., 2013; Lack et al., 2016a,b; Bastide et al., 2016; Fabian et al., 2015; Klepsatel et al., 2014). Intriguingly, the high-altitude population show a substantial increase in the frequency of qualitative mutational defects of wing morphology (Lack et al., 2016a). Inbred strains derived from the high-altitude population have defect frequencies as high as 40-50% (Lack et al., 2016a). This increase in frequency was not simply a result of hitchhiking of deleterious alleles, or a strong bottleneck, (Pool et al., 2012) but due to reduced mutational robustness, as assessed using mutagenesis experiments (Lack et al., 2016a). Importantly, the population-specific mutational sensitivities are pleiotropically linked to variants that influ-

ence the increase in wing size. This appears to be the case whether considering the variants among the high- and low-altitude populations (Lack et al., 2016a), or even in putative ancestral lowland populations (Groth et al., 2018). Currently, it is inferred that the increase in mutational sensitivity in the high-altitude population may have been a result of strong selection leading to rapid adaptation, but with potentially pleiotropic consequences. This may represent a situation similar to that envisioned by Waddington (1942), where the population has not yet re-evolved its canalization mechanism after a long bout of strong directional selection for larger body size, wing size and shape.

The evolutionary history of the high- and low-altitude populations provides an ideal opportunity to test the relationship between genetic and environmental canalization. While there is considerable strain-to-strain variation that occurs within populations, strains collected from a high-altitude population in Ethiopia from an elevation of ∼3000 m are much larger in overall size and wing size, have distinct wing shape and have a greater frequency of qualitative 'mutant' phenotypes than low-altitude populations from Zambia from an elevation of ∼500m (Lack et al., 2016a). Wing size and shape in *D. melanogaster* is a model system for studies of plasticity, sensitivity to mutational perturbation and within- and among-individual variability using both natural and lab-induced variation (Haber and Dworkin, 2017; Debat et al., 2011; Breuker et al., 2006; Debat et al., 2006; Soto et al., 2008; Klingenberg and Zaklan, 2000; Pélabon et al., 2006). In this study, we used a number of measures for both macro- and micro-environmental robustness between the high and low-altitude populations. We used different measures of environmental canalization: within-line, among-individual variation (micro-environmental variation), within-individual variation (fluctuating asymmetry) and phenotypic plasticity across a temperature gradient (macro-environmental variation). Further, we correlated the different measures of environmental canalization and mutational perturbation in order to test the congruence hypothesis and de-

termine whether strains with greater proportion of defects are also more variable (aka more decanalized). Despite demonstrating substantial population and environmental differences in wing size, shape, cell density and penetrance of mutational perturbation consistent with previous studies, we observed no consistent differences in measures of micro-environmental and macro-environmental canalization among populations. These results are discussed within the context of our ongoing understanding of the evolutionary mechanisms that influence trait variability.

Materials and Methods

Fly strains and Growth conditions

Drosophila melanogaster strains used in the current study represent a subset of those from Lack et al. (2016a,b). The high-altitude inbred strains were derived from flies collected in Fiche, Ethiopia at an altitude of 3070 m in December 2011. The low-altitude strains were collected in Siavonga, Zambia at an altitude of 530 m, and a 3125 km linear distance away from the high-altitude Ethiopian population in July 2010.

The flies for the micro-environmental variation experiments were raised as per Lack et al. (2016a). Flies were raised at 25°C, in 70% humidity, with 12:12 hour light/dark chamber, on standard cornmeal molasses food at a low larval density.

The flies for the temperature plasticity (macro-environmental variation) and fluctuating asymmetry experiments were raised on a 1:1.5 protein to sugar ratio diet; recipe outlined in Table S3 at 24°C for two generations prior to the experiment. Newly emerged adults, 10-20 males and females, were collected and placed in egg collection chambers with apple agar

plates with yeast. Eggs were collected, 50 at a time, and placed into vials with food. The flies were raised at 18°C, 24°C, and 28°C in 12:12 hour light/dark chambers until emergence. Adults were collected within 2 days of emergence and preserved in 70% ethanol.

Phenotyping

Wing size and shape - Micro-environmental canalization

The right wing of each fly was dissected and imaged using an Olympus DP30B camera mounted on an Olympus BX51 microscope (Olympus software V.3,1,1208) using a 2x objective (20X total magnification). Landmark and semi-landmark data was captured using a modified version of the "WINGMACHINE" pipeline (Houle et al., 2003; Pitchers et al., 2019). Coordinates of two starting landmarks were recorded using tpsDig2 software (V2.16). These coordinates are the humeral break on the leading edge of the wing and the alula notch on the trailing edge of the wing. B-splines were fit to veins and wing margin for each image using Wings (V3.7), reviewed and manually adjusted, if necessary. Landmark and semi-landmark positions were extracted and the shape information after adjusting for size, position, and rotation information using CPReader software (V1.12r). This produced data composed of 12 landmarks and 36 semi-landmarks (Figure 1B) as well as centroid size of for each specimen. The strains used in this experiment are outlined in Table S1.

Measuring trichome (cell) density

A subset of strains used for the the initial size and shape analysis were re-imaged with a higher resolution camera. We chose 15 strains from each population as follows: five lines were chosen with the highest within-line coefficient of variation (CV) for wing size; five lines were

chosen with the lowest within-line CV for wing size (See Analysis for CV calculation); five lines were chosen at random from each population. This allowed us to maximize the variation we examined within each population. We phenotyped 15-20 males and females from each line. Wings were imaged using an Olympus DP80 camera mounted on an Olympus BX43 microscope, using a 4X objective (total 40X magnification). Images were captured with cellSens Standard (V1.14) software. Cell density was quantified by counting trichomes on the surface of the wing using the ImageJ FijiWings macro (V2.2) (Dobens and Dobens, 2013). Each trichome represents a single cell (Dobzhansky, 1929). We used a $0.0065mm^2$ (75x75 px) measurement area in each of 16 different locations in the wing (Figure 4A).

Wing size and shape - Macro-environmental canalization

The right wing of each fly was imaged using the same microscope settings as the cell density experiments. Wing size and shape were quantified using the same pipeline as the micro-environmental canalization experiment. We used 3 replicate vials per line per rearing temperature. Strains used for this experiment are outlined in Table S2.

Fluctuating Asymmetry of wing size and shape

Left and right wings were phenotyped for two lines from each population (E39 and E73 from high-altitude, Z254 and Z311 from low-altitude, total number of indivudals: 509) to asses fluctuating asymmetry. Duplicate measurements were taken of the left and right wings from 77 individuals chosen randomly from different populations, sexes and rearing temperatures to estimate measurement error. The same phenotyping methods were used as the micro-environmental canalization experiment.

Quantification of wing defects

Each wing image was manually scored for venation defects. For the the micro-environmental canalization experiment, proportion of defects was calculated as the ratio of the number of wings with defects to total wings for each line. For the macro-environmental canalization experiment, each individual wing was scored based on whether they have a defect or not, using a binary scale (1 for defect observed, 0 for defect not observed). The proportion of defects for each line was calculated by averaging the scores for all individuals within line and experimental treatment.

Analysis

Data was analyzed using R (v3.5.1) (R Core Team, 2018) in RStudio (v1.1.456) on a MacBook Pro, running macOS Mojave (V10.14.2). Mixed models were run using *lmer* and *glmer* from the package *lme4* (V1.1.19) (Bates et al., 2015), glmmTMB (Brooks et al., 2017) and *procD.lm* from the package *geomorph* (V3.0.7) (Adams et al., 2018). Generalized linear models were run using *glm* from the *stats* package (V3.5.1) (R Core Team, 2018).

Modeling wing size, shape, cell density and wing defects

Linear mixed models were fit with the wing size, and cell density data, and generalized mixed model (binomial distribution, with a logit link) was fit with the wing defects data using population, sex, rearing temperature (for macro-environmental canalization experiment) and their interactions as fixed effects. Where possible, the intercept and sex effects were allowed to vary as random effects of line. Additionally, for the cell density data, wing region was included as a fixed effect to test whether there is variation in cell density across the wing,

and an individual level random effect was included to account for the multiple measures per wing.

For wing shape, we fit a multivariate linear model using *procD.lm* estimating the effects size, population and sex and their interactions as fixed effects, and line nested within population as a random effect. Statistical inference was performed using a randomized residual permutation procedure in geomorph using 1000-2000 permutations for each effect.

Estimating among-individual, within-line variation of wing size, shape and cell density

For wing size, among-individual, within-line variation was estimated in two ways for each strain. Using the coefficient of variation ($CV = \frac{\sigma}{\mu}$) and the median form of Levene's deviates (used for all formal statistical inference). (Van Valen, 2005; Dworkin, 2005b). For the macro-environmental canalization experiment, sex effects were first modeled out and then CV and Levene's deviates were calculated for each line at each temperature.

For wing shape, we use properties of the within-strain phenotypic variance co-variance (VCV) matrix such as matrix size (total variance), and matrix shape (eccentricity and standard deviation of eigenvalues), (Jones et al., 2003; Kirkpatrick, 2009; Haber, 2011; Pavlicev et al., 2009; Van Valen, 1974). The registration process (Procrustes superimposition) influences covariation within and among landmarks. As such, the use of Procrustes residuals for such analysis is of potential concern. However, in our previous study, we demonstrated using a variety of approaches, that at least for *Drosophila* wing shape, results from Procrustes superimposition were extremely similar to those generated via spatial interpolation of the data to generate multivariate variables (Haber and Dworkin, 2017). As such, for this study we used the Procrustes Residuals for simplicity.

Total variance is the trace of the VCV matrix and can be defined as the sum of all the eigenvalues of that matrix (Van Valen, 2005). Total variance estimates were multiplied by a factor of 1000.

The shape of the VCV matrix can be quantified using matrix eccentricity, defined as the ratio between the largest eigenvalue and the total variance. (Jones et al., 2003; Kirkpatrick, 2009)

Additionally, we calculated the relative standard deviation of eigenvalues (rSDE) and the relative standard deviation of the eigenvalues scaled by the Total variance (rSDE2) (Pavlicev et al., 2009; Van Valen, 1974; Haber, 2011). rSDE estimates were multiplied by a factor of 10000 and rSDE2 estimates were multiplied by a factor of 10.

Variation due to sex and size was modeled out prior to estimating total variance, eccentricity and rSDE for each strain.

CV and Levene's Deviates, were also calculated for trichome (cell) density for each strain. CV was calculated in two ways. First by averaging the cell density across the wing for each individual and calculating the within-individual CV, and then averaging the CV for line. Alternatively, cell density CV was calculated was by averaging cell density for each line first and then calculating CV. These two approaches of measuring CV produced similar results and only the first one is used in this paper.

Associations between the measures calculated above across strains was performed using a Pearson correlation coefficient.

We then examined whether any of the variation measures (CV and Levene's deviates for size, total variance, eccentricity, rSDE and rSDE2, as well as cell density CV and Levene's deviates, as response variables) varied due to the effects of population and sex. Generalized

mixed models were fit, sex (excluded for wing shape measures of variation and wing size measures of variation for macro-environmental experiment), population, temperature (for macro-environmental canalization experiment) and their interactions as fixed effects. The intercept and where possible sex were allowed to vary as random effects by line (nested within population). Given that all of these responses can only take on continuous positive values we assumed a Gamma distribution with an inverse link function.

Fluctuating Asymmetry of wing size and shape

To quantify measurement error, duplicate measures of left and right wing shape for 77 individuals was conducted. Wing size measurement error was estimated using an analysis of variance (ANOVA) and wing shape ME was estimated using a Procrustes ANOVA. For both wing size and shape, individual, side and their interaction were used as effects. The side effect represents directional asymmetry, the individual effect represents the variation among individuals and the side:individual interaction term represents fluctuating asymmetry The residual variance in this model estimates measurement error (Klingenberg and McIntyre, 1998; Debat et al., 2009; Palmer and Strobeck, 1986, 2003; Palmer, 1994). We compared the variation of the side:individual interaction term with the residual variation in order to determine whether the measurement error was negligible with respect to fluctuating asymmetry (Supplementary Tables S24 and S25).

Fluctuating asymmetry of wing size was calculated using the standard FA indices: FA1 ($FA1 = |R - L|$) and FA8a ($FA8a = |ln(\frac{R}{L})|$) (Palmer and Strobeck, 1986, 2003; Palmer, 1994).

In order to estimate and assess differences in developmental stability (based on FA), we fit a generalized linear mixed model using the FA indices for wing size $FA1$ and $FA8a$ as

response variables, temperature, sex, population and their interactions as fixed effects and line nested within population as a random effect assuming a Gamma distribution and an inverse link function for the response. The FA component for shape was extracted for each specimen using the *bilat.symmetry* function from *geomorph* to remove directional asymmetry. The extracted FA component was modeled by fitting a generalized linear model assuming a Gamma distribution using mean wing size (between left and right wing), population, sex and their interaction as fixed effects. As a confirmation of this analysis, wing shape FA was calculated as the Procrustes distance between the left and right wing for each individual PD_{RL}. We fit a generalized linear model using PD_{RL} as response variables, temperature, sex, population and their interactions as fixed effects and line as random effect assuming a Gamma distribution and an inverse link function. As a further confirmation of the FA analysis, morphological disparity analysis was performed to compare the difference in FA among groups. Each analysis provided largely similar results and only the first two are shown.

Results

Wing size, shape and wing defects vary between high- and low-altitude populations

First, we sought to confirm differences in trait means across populations. Consistent with previous findings (Lack et al., 2016a; Pitchers et al., 2013; Fabian et al., 2015), the high-altitude population has substantially larger wing size compared to the low-altitude population (Figure 1A, Table S4). Wing shape also varies in a manner consistent with previous studies (mean Procrustes distance of 0.013 between populations) (Pitchers et al., 2013) shown in

Figure 1C and Table S5.

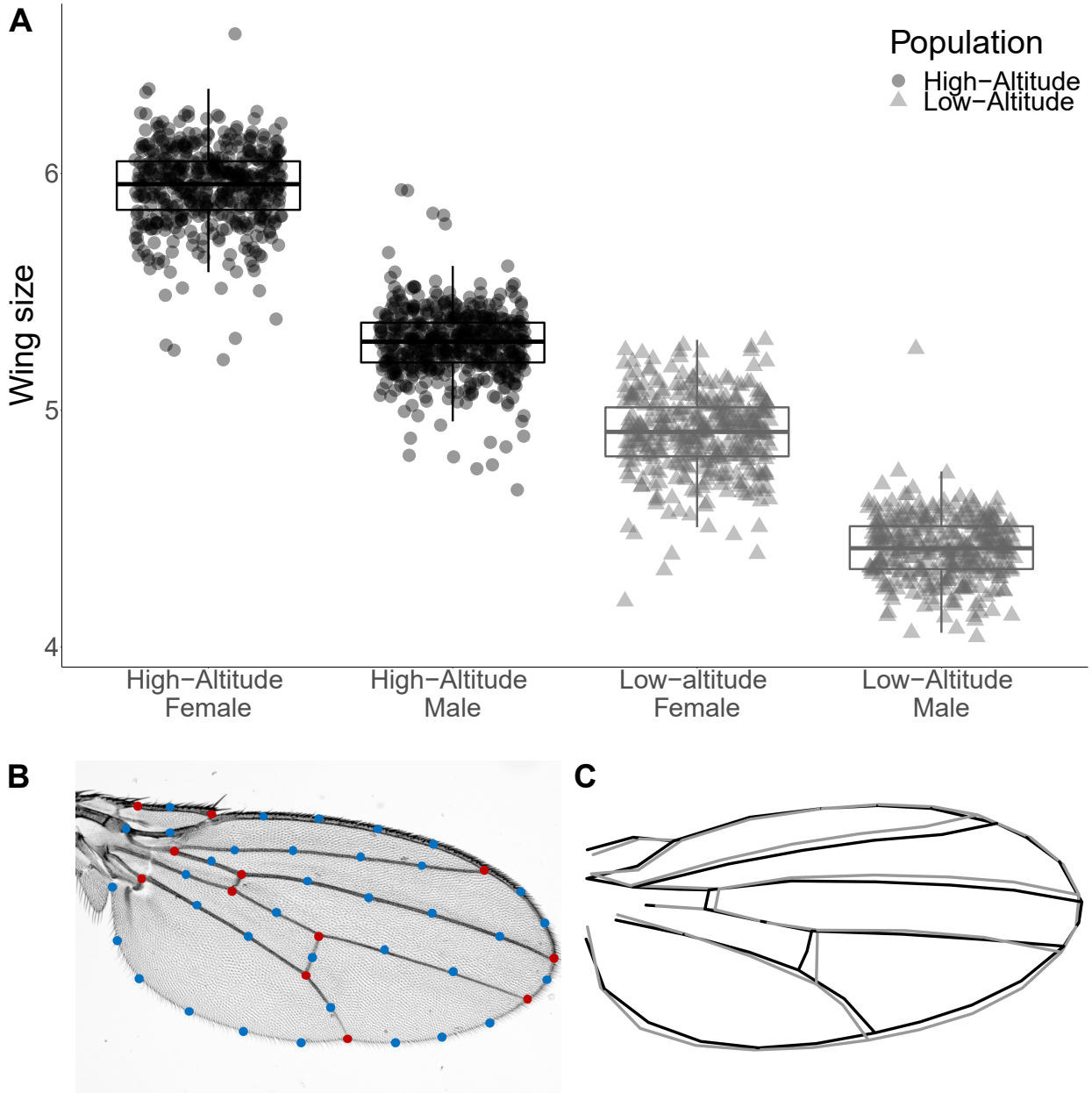

Figure 1: Wing size (centroid size) and shape variation within and among high- and low-altitude populations. (A) High-altitude population has larger wings than the low-altitude population (B) Landmarks (red) and semi-landmarks (blue) used in the analysis of wing shape. (C) Mean difference in wing shape between the high- and low-altitude population, scaled 2x. Procrustes distance between mean shapes of the two populations is 0.013

Consistent with Lack et al. (2016a), flies from the high-altitude population have a greater

proportion of wing defects compared to the low-altitude population (Figure S1, Table S8). On average, 26.7% (CI 22.8% - 29.9%) of high-altitude females and 22.0% (CI 18.6% - 25.8%) of males show such defects. In contrast, the average for the low-altitude females is 10.6% (CI 7.53% - 14.6%) and 12.2% (CIs 8.22% - 17.6%) for males.

Micro-environmental variation for wing size is similar between the high- and low-altitude populations

While it was previously demonstrated that the high altitude population is genetically decanalized (Lack et al., 2016a), it is unclear whether this is also associated with any form of environmental decanalization. To enable comparisons with previous studies we used both the coefficient of variation (CV) and Levene's deviates to measures among individual, within-line variation. CV was plotted for ease of interpretation, but all statistical analyses were performed using Levene's deviates (Van Valen, 2005; Dworkin, 2005b). However, Levene's deviates and CV are highly correlated (high-altitude r = 0.89 CIs 0.81 - 0.94; low-altitude r = 0.98 CIs 0.94 - 0.99) (Figure S3B). As shown in Figure 2A, Figure S3A and Tables S6 and S7 , measures of among individual, within-strain variability are similar between the high-altitude and low-altitude populations.

We examined the relationship between CV and proportion of defects, which showed a weak negative correlation, in the high-altitude population $r = -0.28$ (CIs -0.55 - 0.045), and a correlation close to zero in the low-altitude population $r = -7.45 \times 10^{-3}$ (CIs -0.46 - 0.45) (Figure 2B, Figure S4A) although confidence intervals included zero for both populations. Similar results were observed when comparing within-line Levene's Deviates with proportion of defects (Figure S4A).

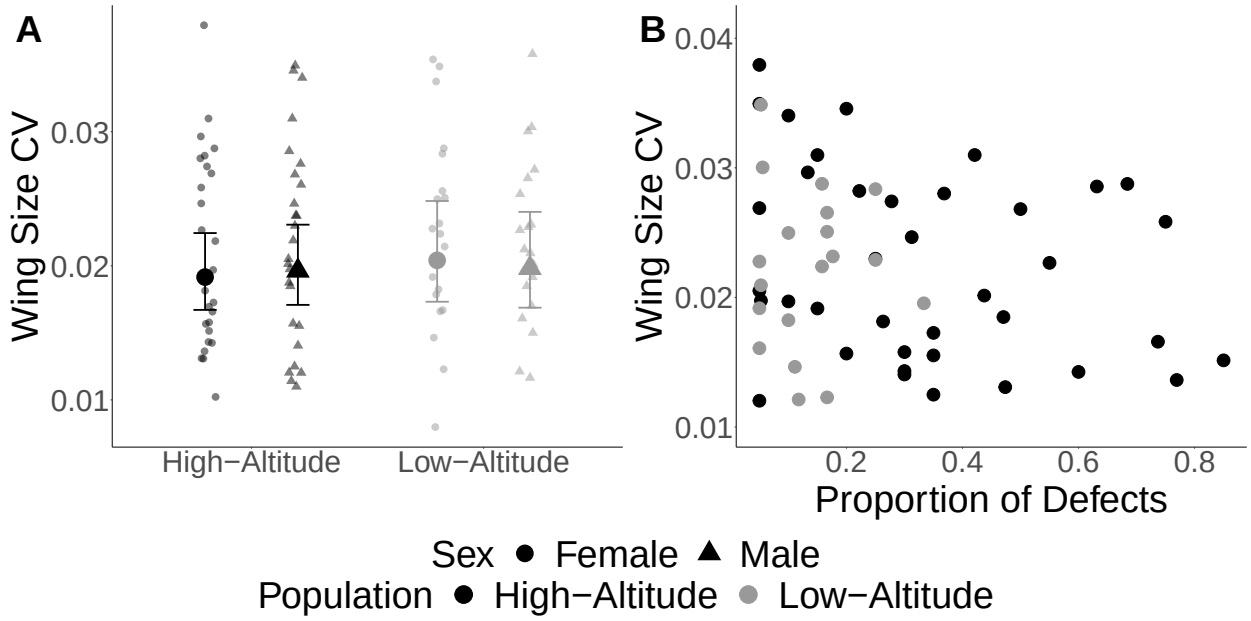

Figure 2: Within-line, among individual variation for wing size is similar across both high- and low-altitude populations. (A) Within-line variation measured as CV for wing size is similar in the high- and low-altitude populations. Large symbols represent fitted values, small symbols represent within-line CV for each line. Error bars are 95% CI (B)Within-line CV for wing size is not correlated with proportion of defects in both the high-altitude population $r = -0.28$ (CIs -0.55 - 0.045) or the low-altitude population $r = -7.45 \times 10^{-3}$ (CIs -0.46 - 0.45)

Micro-environmental variation for wing shape is similar between high- and low-altitude populations

We observed that strains derived from high- and low-altitude populations have similar measures of within-line total variance and eccentricity (Figure 3 and Tables S9 and S10). We confirmed this using the relative standard deviation of eigenvalues (Figure S5). Similar to the patterns for wing shape among populations, there is little evidence that total variance and eccentricity are correlated with frequency of wing defects (Figure S2) in either high-altitude (Total variance: r = 0.16 CI -0.26 - 0.53; eccentricity: r = -0.24 CI -0.59 - 0.18)

or low-altitude populations (total variance: r = 0.34 CI −0.24 - 0.74; eccentricity: r = −0.33 CI −0.73 - 0.24).

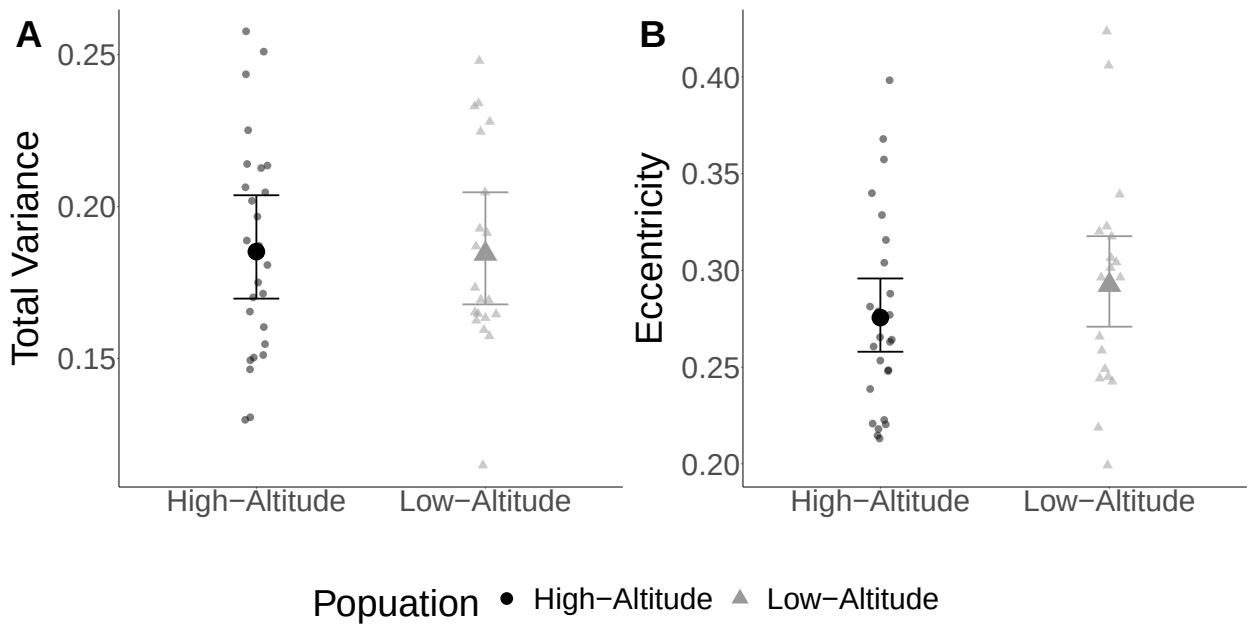

Figure 3: Similar patterns of matrix size and shape properties for wing shape between high- and low-altitude populations as measured using the (A) total variance of the VCV matrix (values multiplied by 1000), and (B) matrix eccentricity. Error bars are 95% CI

Cell Density varies across the wing, between population and sexes

Consistent with previous work, we observed that average cell density is lower (cell size is greater) in the high-altitude population relative to the low-altitude population (Figure 4B) (Fabian et al., 2015; Lack et al., 2016b) as well as for females versus males (Alpatov, 1930; Dobzhansky, 1929). Most studies count trichomes (cells) in a single small region of the wing. Yet, cell sizes are known to vary in different regions of the wing from less than 7.9μm to greater than 11.3μm (González-Gaitán et al., 1994). To account for local effects, we measured cell density in 16 regions across the wing (Figure 4A, Figure S6). While cell density varies considerably across the wing, with some intriguing interactions between sex, region of the

wing and population, the overall pattern as observed in previous work (Lack et al., 2016b; Fabian et al., 2015) remains (Figure S6; Table S11).

Variation within and among individuals for cell density is not associated with among individual variability in size or shape

After we confirmed and expanded upon the previously demonstrated association between cell density and wing size with respect to population and sex, we asked whether variation in cell density, within the wing was directly associated with variation among individuals in wing size and shape. That is, do lines that show the greatest degree of among individual, within-line variance for wing size and shape also show the greatest variance for cell density within and between individuals? We calculated the within-line CV for cell density across the wing in order to determine if there are any differences in within-line variation for cell density between the high- and low-altitude populations. As shown in Figure 4, we did not observe substantial differences in cell density CV between populations but observed an effect of sex that was consistent across both populations (Figure 4C; Table S12). Similarly, we did not observe differences in within-line cell density Levene's deviates between the high- and low-altitude populations (Figure S7A; Table S13).

Additionally, we did not observe a strong association between within-line cell density CV and within-line wing size CV for either the high-altitude population ($r = 0.076$ CIs -0.31 - 0.44) or the low-altitude population ($r = 0.25$ CIs -0.12 - 0.56)(Figure 4D). We observed a weak negative correlation between within-line cell density CV and within-line total variance,in the low-altitude population ($r = -0.41$ CIs -0.67 - -0.060;), but we did not observe any correlation between within-line cell density CV and within-line total variance for the high-altitude population (high-altitude $r = 0.062$ CIs -0.32 - 0.42). Similarly, we did

not observe any association between within-line cell density CV and within-line eccentricity in either the high- or the low-altitude population (high-altitude $r = 0.28$ CIs -0.10 - 0.59, low-altitude $r = -0.30$ CIs -0.60 - 0.063) (Figure S7 C, D). Further, we did not observe any evidence for associations between within-line cell density CV and proportion of wing defects for both the high- and low-altitude populations(high-altitude $r = -0.043$ CIs -0.44 - 0.37, low-altitude $r = 0.36$ CIs -0.076 - 0.68) (Figure S7B)

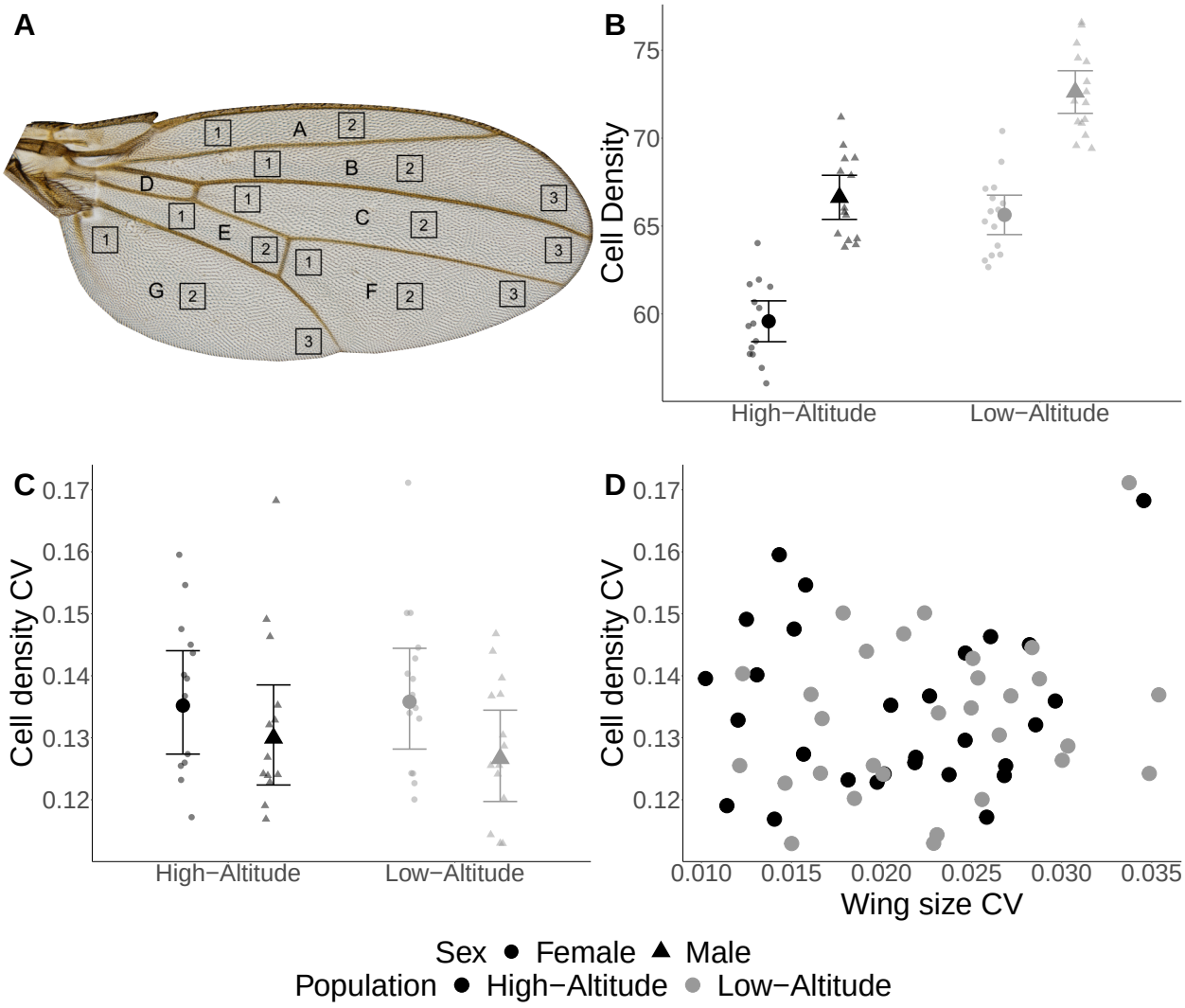

Figure 4: Mean cell density shows population and sex differences, but variability in cell density is not associated with within-line, among individual variation for size and shape. (A) Wing regions used for cell density measurement. Squares represent a $0.0065mm^2$ measurement area. (B) Cell density varies between the sexes and between the high- and low-altitude populations. Error bars are 95% CI (C) Within-line, among individual variation measured as CV for cell density is similar between the high- and low-altitude populations. (D) Cell density CV and wing size CV are not correlated in both the high-altitude population($r = 0.076$ CI -0.31 - 0.44), and the low-altitude population ($r = 0.25$ CI -0.12 - 0.56)

Temperature induced plasticity

To assess whether patterns of phenotypic plasticity varied among populations (macro-environmental canalization) we reared flies from both populations at three temperatures. Consistent with previous studies, wing size is larger for flies raised at lower temperature and the reaction norms showed modest evidence of a non-linearity (David et al., 1994; Partridge et al., 1994; James et al., 1997). Our data suggest that mean wing size of the high-altitude population may be more plastic compared to the low-altitude population (Figure 5A, Table S14, Table S15). We observed an increase in the Procrustes distance between mean shapes of the high- and low-altitude populations as temperature increases (Figure S10).

We quantified the proportion of wing defects for the high- and low-altitude populations at the three different rearing temperatures. Consistent with our previous results, the high-altitude population has a greater proportion of defects than the low-altitude population, however, we did not observe substantial differences in proportion of defects due to temperature (Figure S8A, Table S20).

In general, we did not observe any differences in the high- and low-altitude populations for among individual, within-line measures of variability at each temperature treatment. For the within-line CV for wing size, we observed an increase at both 18°and 28°C for high-altitude females and an increase in CV at 18°but not at 28°C for high-altitude males. Within-line CV for low-altitude males and females is consistent across temperatures (Figure 5B; Table S16). We observed a similar pattern when using Levene's deviates as we did for CV (Figure S8B; Table S17), with a modest effect of rearing temperature but little evidence for population level differences. For within-line wing shape total variance, we observe a consistent increase with temperature for both high- and low-altitude populations (Figure 5C; Table S18). Within-line wing shape eccentricity is similar across populations and temperatures (Figure 5D; Table

S19). Additionally, for within-line rSDE and rSDE we observe a similar pattern as we did for eccentricity (Figure S8C, D).

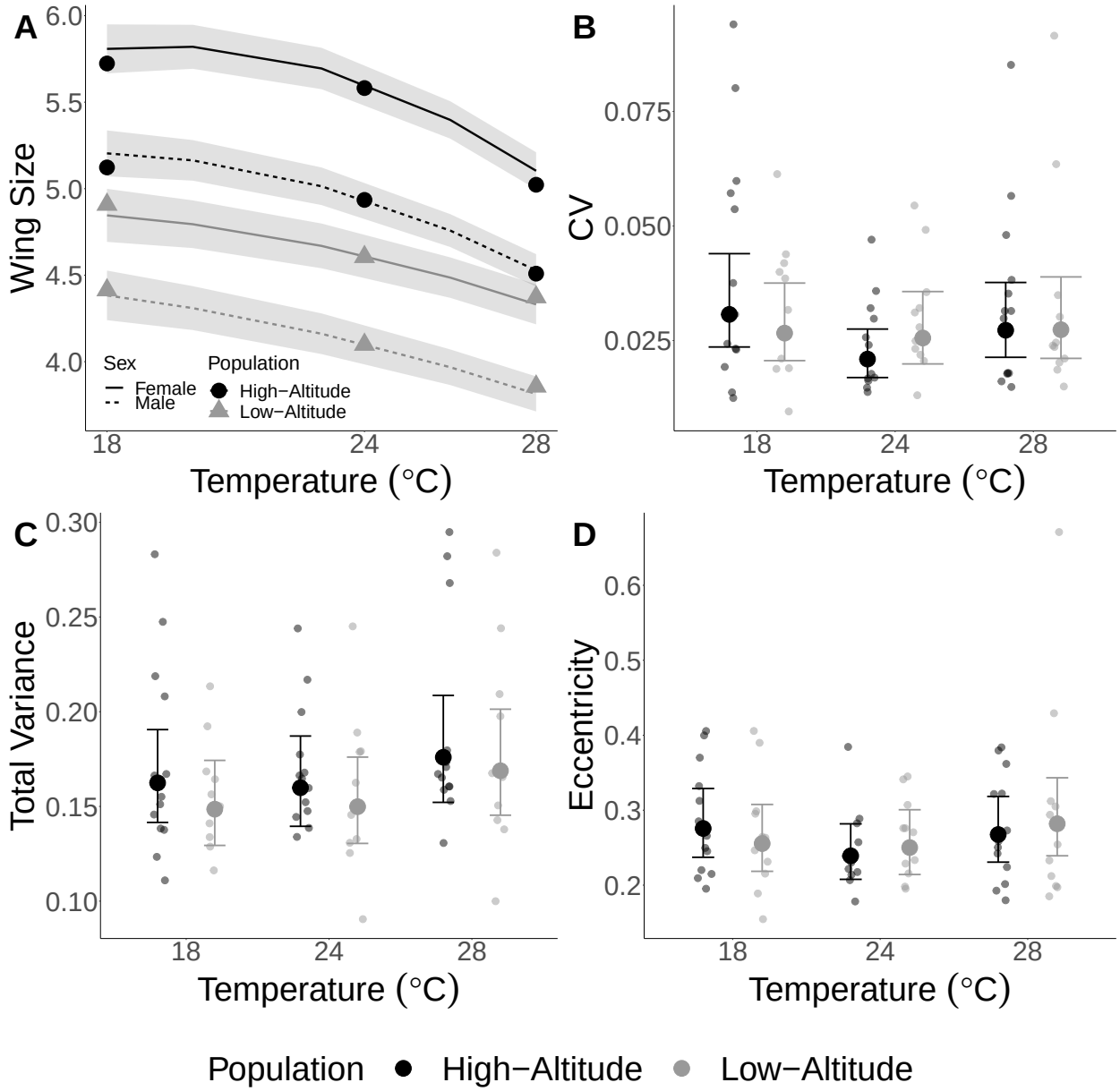

Figure 5: Considerable plasticity for mean wing size under different temperature rearing environments, but with minimal changes in trait variation (A) Mean changes in wing size (centroid size) (B) CV for wing size is similar across sexes and populations within each temperature treatment (C) total variance for wing shape (multiplied by a factor of 1000) and (D) eccentricity are similar between the high- and low-altitude populations across the different rearing temperatures. Gray shading and error bars are 95% CIs

Fluctuating asymmetry for wing size and shape is similar between high- and low-altitude populations

While among-individual, within-genotype variation and within-individual (among-sides) variation might be expected to capture similar aspects of developmental stability, empirical work shows that these do not always agree. Thus we measured asymmetry for both wing size and shape in a subset high- and low-altitude lines (2 strains for each population). We first estimated measurement error for wing size and shape and determined that measurement error was negligible with respect to FA for size (Tables S25 although had a larger impact on shape S26). Using FA8 as an index for wing size FA, we assessed and compared developmental instability for the high- and low-altitude populations at three different temperatures (18°C, 24°C and 28°C). We did not see any clear evidence of differences in FA8 between the high- and low-altitude populations or across the different temperatures, except for consistently lower FA8 in the high-altitude males at all temperatures (Figure 6A; Table S22). For comparison we also report FA1 as a measure of developmental instability ((Figure S11; Table S21)), but importantly FA1 does not account for mean trait size and should be interpreted with caution.

We measured FA for wing shape in two different ways. First, we calculated FA by removing directional asymmetry (DA) and then calculating the Procrustes distance (PD) for the FA component for each individual. The high-altitude females had a slight, but consistent increase in FA across all three temperatures while the other groups had similar FA to each other and across the different temperatures (Figure 6B Table S24). Alternatively, we calculated the PD between the left and right wing (PD_{LR}). We observed a similar pattern as we did using to first method to calculate wing shape FA, where there is consistently greater PD_{LR} in the high-altitude females across temperatures (Figure S11B; Table S23).

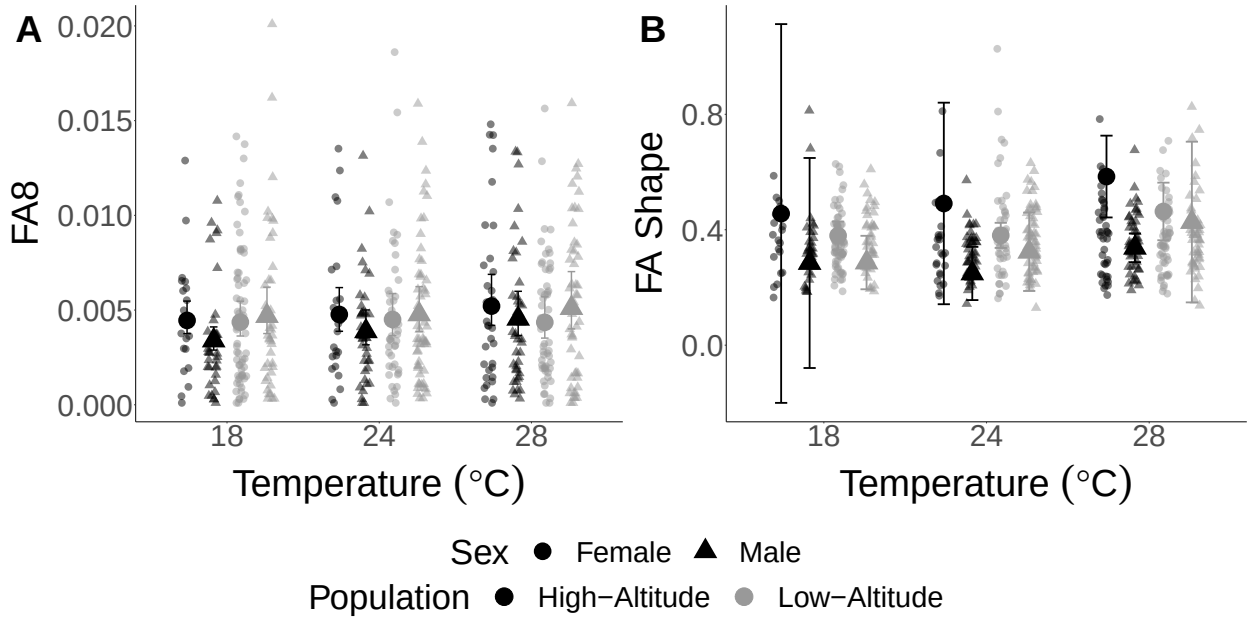

Figure 6: inconsistent differences in fluctuating asymmetry among the high- and low- altitude populations for (A) wing size ($FA8$) and (B) wing shape FA between the high- and low-altitude populations across temperatures. High-altitude males have consistently lower FA8 and high-altitude females have consistently greater wing shape FA across temperatures. Large symbols represent population means, small symbols represent individuals, error bars represent 95% CIs

Discussion

Since the proposal of the congruence hypothesis, researchers have tested for associations between genetic and environmental canalization. To date, empirical evidence has been equivocal. The majority of studies that support the congruence hypothesis were conducted in RNA viruses and microRNAs *in vivo* and *in silico*, which may not necessarily representative of multi-cellular organisms (Szöllősi and Derényi, 2009). Studies in other systems have been conducted and have not provided evidence to support the congruence hypothesis (Dworkin, 2005a,c). However, most empirical studies were conducted using lab domesticated lineages, laboratory-induced mutations, and with arbitrary measures of genetic canalization which

may not be representative of naturally occurring phenomena that would lead to congruent evolution of genetic and environmental canalization.

In this study, we address these issues using a naturally occurring system across adaptively diverged populations. The high-altitude population was previously demonstrated to have reduced mutational robustness (via mutagenesis), and these effects were pleiotropically linked to variants influencing changes in wing and body size that appear to be targets of selection (Lack et al., 2016a). Using strains derived from both high- and low-altitude populations, we examined multiple measures of environmental canalization. Despite recapitulating previously observed divergence in wing size & shape (Pitchers et al., 2013; Lack et al., 2016a,b) (Figures 1, 5A, S10), cell size (Figure 4, Figure S6) and frequencies of mutational defects (Figure 2B, Figure S1), we did not observe any evidence for associations between genetic and micro-environmental canalization (Figures 2, 3, 4C-D, 5B-D). Additionally, measures of among individual, within-line variance for wing size and shape were not correlated with the proportion of mutational defects both the high- and low-altitude populations (Figures 2B, 4D, Figures S2, S4, S7B, S9). We did observe greater temperature induced plasticity of mean wing size in the high-altitude population (Figure 5A). We observed a subtle increase in within-line variation for wing size at 18°C and 28°C compared to 24°C in both the high- and low-altitude population for wing size, and this increase was greater in the high-altitude population (Figure 5B, Figure S8B), although there is at best, marginal evidence for a significant treatment effect of population or its interaction with rearing temperature (Table S16). Intriguingly, we observed a decrease in developmental stability for wing size (measured using fluctuating asymmetry) for high-altitude females across all temperatures (Figure 6). Yet, we did not observe this same pattern in the males derived from the same population, nor did we see any increase in qualitative wing defects at varying temperatures and the measures of among-individual, within-line of variation were not correlated with the

proportion of defects at any of the developmental temperatures (S9). Therefore, our results are largely inconsistent with congruent evolution of genetic and environmental canalization and that with respect to adaptation to life at high altitudes driving changes in both trait means and variances, they are likely to evolve via separate underlying mechanisms.

Our study is one of the few to test of the congruence hypothesis using strains derived from natural populations with known evolutionary histories. However, we are aware of several important caveats. Our study compares a single high-altitude population to a single low-altitude population. Replication of our experiments with additional populations from independent altitudinal clines would provide stronger support of our findings. The strains used in this study were collected 7-8 years ago and may have undergone some degree of lab adaptation (although N_e is extremely small within each strain, so selection would be extremely weak). Although the two populations have modest genetic differentiation ($F_{ST} = 0.15$), lines derived from African populations tend to have have high residual heterozygosity even after 8 generations of inbreeding (Lack et al., 2016a). Additionally, for tests of plastic responses to environmental variables, rearing temperature is only one of several ecologically relevant variables.

While the variants contributing to divergence in size both between the high- and low-altitude population (Lack et al., 2016a) and under artificial selection derived from the low-altitude population (Groth et al., 2018) appear to be pleiotropically linked to the mechanism influencing sensitivity to mutational perturbation, these are not in fact the same traits. The penetrance of wing abnormalities among lines (and their increase under mutagenesis) may reflect one aspect of genetic canalization (that is linked to variants influencing mean size and shape), but they do not necessarily influence variance for these traits. Indeed, under high temperature stress (31°C) one of three replicates of lineages artificially selected for increased size (from a low-altitude ancestral population) showed a substantial increase in penetrance

of wing abnormalities (Groth et al., 2018). Interestingly we observed no increase in such abnormalities at our high temperature rearing (28°C) for the high-altitude strains. Whether this reflects insufficient stress or a difference in response is unclear. However, it is clear that the degree of genetic correlation between trait means and variances for wing size, shape and penetrance of abnormalities demonstrates considerable complexity.

While the work of Lack et al. (2016a,b) and Groth et al. (2018) clearly demonstrate evolutionary changes in genetic canalization associated with adaptive trait evolution (body size and wing morphology), the most likely evolutionary scenario is one of the loss of genetic canalization associated with the pleiotropic effects of variants that have contributed to the evolution of mean trait expression at high altitudes. This is reminiscent of the evolution of insecticide resistance in blowflies which showed a pleiotropic increase in fluctuating asymmetry due to the resistance locus or in sticklebacks with the effects of the *Eda* locus on both the expression of armour plates, and on fluctuating asymmetry of plates (McKenzie and Clarke, 1988; Morris et al., 2019). Consistent with Waddington's model for the evolution of canalization, it could be that modifiers that increase the mutational robustness of wing morphology in *Drosophila* have not yet risen to appreciable frequency in the high altitudinal populations, as it occured with the modifier variants influencing asymmetry in the blowflies (Davies et al., 1996). Future work will focus on identifying the variants influencing the adaptive divergence in wing size and morphology and how they shape mutational robustness are necessary.

Where does this leave the congruence scenario? While the results from the current study, and some previous studies (Dworkin, 2005a,c; Borenstein and Ruppin, 2006) are not consistent with the congruence hypothesis, it is perhaps best to consider under what conditions the direct evolution of genetic canalization or its evolution as a correlated response are probable. The model observation that finds that deleterious alleles are purged efficiently enough that

adaptive genetic canalization will not have the opportunity to evolve does not consider that deleterious alleles are not always purged efficiently for numerous reasons (pleiotropy, linkage with beneficial mutations, drift, GxE, fluctuating selection etc). Indeed, with the system we investigated in this study, the reduction in mutational robustness appears to be a direct pleiotropic consequence of the allelic effects on organismal size (Lack et al., 2016a; Groth et al., 2018). In such instances, genetic canalization may evolve to suppress deleterious mutational effects (ie. alter patterns of pleiotropy). As has been demonstrated previously, the likelihood of evolving either genetic or environmental canalization is in part a function of the fitness load imposed by the frequency and magnitude of environmental and genetic perturbations (Proulx and Phillips, 2005; Hermisson et al., 2003). However, our knowledge of the distribution of this fitness load and the frequency of relevant environmental perturbations is quite unclear. While studies of mutational load give us some idea of the distribution of fitness effects of new mutations, this is less clear in natural environments. Indeed, it has been argued that new allelic combinations produced due to the normal processes of mating and recombination may act as a "genetic perturbation" to the input of new mutations (Stearns et al., 1995), as genetic backgrounds are constantly shuffled. Alternatively, it may be that genetic canalization may be beneficial when it occurs, but rarely the result of persistent and direct selection (Wagner et al., 1997; Gibson and Wagner, 2000; Visser et al., 2003; Siegal and Bergman, 2002; Proulx and Phillips, 2005). Experimental evolution and artificial selection may continue to be the strongest framework to test the theory and understand under what conditions both genetic and environmental canalization are direct targets of selection. This approach should be coupled with studies of adaptively diverged natural populations that are likely to share the appropriate evolutionary history to address these questions (i.e. (Morris et al., 2019; McKenzie and Clarke, 1988)). Finally, this work suggests that trying to clearly delineate between selection for ‘environmental’ and ‘genetic’ canalization may be difficult given the interplay between genotypic and environmental effects in terms of trait expression

and variation.

Acknowledgements

We wish to thank Dr. John Pool and Dr. Justin Lack for sending us strains. We also thank Megan Cermak, Christian Marier, Sachin Davis, Melissa Rezik, Ravina Dhami, and Vikram Bhagavat for help with dissections and imaging. Dr. Greg Gibson and Dr. John Pool provided feedback on an earlier version of this manuscript. This work was funded by a NSERC (Canada) Discovery and Discovery accelerator grant to ID.

References

Adams, D., Collyer, M., and Kaliontzopoulou, A. 2018. Geomorph: Software for geometric morphometric analyses. R package version 3.0.6.

Agrawal, A. F. and Stinchcombe, J. R. 2009. How much do genetic covariances alter the rate of adaptation? *Proceedings of The Royal Society B: Biological sciences* 276:1183–1191.

Alpatov, W. W. 1930. Phenotypical variation in body and cell size of *Drosophila Melanogaster*. *The Biological Bulletin* 58:85–103.

Ancel, L. W. and Fontana, W. 2000. Plasticity, evolvability, and modularity in RNA. *Journal of Experimental Zoology* 288:242–283.

Arnold, S. 1983. Morphology, performance and fitness. *American Zoologist* 23:347–361.

Arnold, S. 2003. Performance surfaces and adaptive landscapes. *Integrative And Comparative Biology* 43:367–375.

Bastide, H., Lange, J. D., Lack, J. B., Yassin, A., and Pool, J. E. 2016. A variable genetic architecture of melanic evolution in *Drosophila melanogaster*. *Genetics* 204:1307–1319.

Bates, D., Mächler, M., Bolker, B., and Walker, S. 2015. Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software* 67:1–48.

Blows, M. W. and Mcguigan, K. 2014. The distribution of genetic variance across phenotypic space and the response to selection. *Molecular Ecology* 24:2056–2072.

Borenstein, E. and Ruppin, E. 2006. Direct evolution of genetic robustness in microRNA. *Proceedings of the National Academy of Sciences* 103:6593–6598.

Breuker, C. J., Patterson, J. S., and Klingenberg, C. P. 2006. A single basis for developmental buffering of *Drosophila* wing shape. *PLoS ONE* 1:e7.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., and Bolker, B. M. 2017. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal* 9:378–400.

Camara, M. D. and Pigliucci, M. 1999. Mutational contributions to genetic variance-covariance matrices: an experimental approach using induced mutations in *Arabidopsis thaliana*. *Evolution* 53:1692–1703.

Chandler, C. H., Chari, S., Kowalski, A., Choi, L., Tack, D., DeNieu, M., Pitchers, W., Sonnenschein, A., Marvin, L., Hummel, K., Marier, C., Victory, A., Porter, C., Mammel, A., Holms, J., Sivaratnam, G., and Dworkin, I. 2017. How well do you know your mutation? Complex effects of genetic background on expressivity, complementation, and ordering of allelic effects. *PLoS Genetics* 13:e1007075.

Chari, S. and Dworkin, I. 2013. The conditional nature of genetic interactions: The consequences of wild-type backgrounds on mutational interactions in a genome-wide modifier screen. *PLoS Genetics* 9:e1003661.

Cheverud, J. M., Rutledge, J. J., and Atchley, W. R. 1983. Quantitative genetics of development: Genetic correlations among age-specific trait values and the evolution of ontogeny. *Evolution* 37:895–905.

David, J., Moreteau, B., Gauthier, J., Pétavy, G., Stockel, A., and Imasheva, A. 1994. Reaction norms of size characters in relation to growth temperature in *Drosophila melanogaster*: an isofemale lines analysis. *Genetics Selection Evolution* 26:229–251.

Davies, A. G., Game, A. Y., Chen, Z., Williams, T. J., Goodall, S., Yen, J. L., McKenzie, J. A., and Batterham, P. 1996. Scalloped wings is the *Lucilia cuprina* notch homologue and a candidate for the modifier of fitness and asymmetry of diazinon resistance. *Genetics* 143:1321–1337.

Debat, V., Bloyer, S., Faradji, F., Gidaszewski, N., Navarro, N., Orozco-terWengel, P., Ribeiro, V., Schlötterer, C., Deutsch, J. S., and Peronnet, F. 2011. Developmental stability: A major role for cyclin g in *Drosophila melanogaster*. *PLoS Genetics* 7:e1002314.

Debat, V., Debelle, A., and Dworkin, I. 2009. Plasticity, canalization, and developmental stability of the *Drosophila* wing: Joint effects of mutations and developmental temperature. *Evolution* 63:2864–2876.

Debat, V., Milton, C. C., Rutherford, S., Klingenberg, C. P., and Hoffmann, A. A. 2006. Hsp90 and the quantitative variation of wing shape in *Drosophila melanogaster*. *Evolution* 60:2529–2538.

Dittmar, E. L., Oakley, C. G., Conner, J. K., Gould, B. A., and Schemske, D. W. 2016. Factors influencing the effect size distribution of adaptive substitutions. *Proceedings Of The Royal Society B: Biological Sciences* 283:20153065.

Dobens, A. C. and Dobens, L. L. 2013. Fijiwings: An open source toolkit for semiautomated morphometric analysis of insect wings. *G3* 3:1443–1449.

Dobzhansky, T. 1929. The influence of the quantity and quality of chromosomal material on the size of the cells in *Drosophila melanogaster*. *Wilhelm Roux' Archiv fur Entwicklungsmechanik der Organismen* 115:363–379.

Dworkin, I. 2005a. A study of canalization and developmental stability in the sternopleural bristle system of *Drosophila melanogaster*. *Evolution* 59:1500–1509.

Dworkin, I. 2005b. Canalization, cryptic variation, and developmental buffering: A critical examination and analytical perspective, pp. 131–158. *In* Variation. Academic Press.

Dworkin, I. 2005c. Evidence for canalization of Distal-less function in the leg of *Drosophila melanogaster*. *Evolution and Development* 7:89–100.

Eshel, I. and Matessi, C. 1998. Canalization, Genetic Assimilation and Preadaptation: A Quantitative Genetic Model. *Genetics* 149:2119–2133.

Fabian, D. K., Lack, J. B., Mathur, V., Schlötterer, C., Schmidt, P. S., Pool, J. E., and Flatt, T. 2015. Spatially varying selection shapes life history clines among populations of *Drosophila melanogaster* from sub-Saharan Africa. *Journal of Evolutionary Biology* 28:826–840.

Félix, M.-A. and Barkoulas, M. 2015. Pervasive robustness in biological systems. *Nature Reviews Genetics* 16:483–496.

Flatt, T. 2005. The evolutionary genetics of canalization. *The Quarterly review of biology* 80:287–316.

Geiler-Samerotte, K., Sartori, F. M., and Siegal, M. L. 2019. Decanalizing thinking on genetic canalization. *Seminars in cell & developmental biology* 88:54–66.

Gibson, G. and van Helden, S. 1997. Is function of the *Drosophila* homeotic gene Ultrabithorax canalized? *Genetics* 147:1155–1168.

Gibson, G. and Wagner, G. 2000. Canalization in evolutionary genetics: a stabilizing theory? *BioEssays* 22:372–380.

González-Gaitán, M., Capdevila, M. P., and Garcia-Bellido, A. 1994. Cell proliferation patterns in the wing imaginal disc of *Drosophila*. *Mechanisms of Development* 46:183–200.

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., and Hallgrímsson, B. 2017. Developmental nonlinearity drives phenotypic robustness. *Nat Commun* 8:1970.

Groth, B. R., Huang, Y., Monette, M. J., and Pool, J. E. 2018. Directional selection reduces developmental canalization against genetic and environmental perturbations in *Drosophila* wings. *Evolution* 72:1708–1715.

Haber, A. 2011. A comparative analysis of integration indices. *Evolutionary Biology* 38:476–488.

Haber, A. and Dworkin, I. 2017. Disintegrating the fly: A mutational perspective on phenotypic integration and covariation. *Evolution* 71:66–80.

Hallgrímsson, B., Brown, J. J. Y., Ford-Hutchinson, A. F., Sheets, H. D., Zelditch, M. L., and Jirik, F. R. 2006. The brachymorph mouse and the developmental-genetic basis for canalization and morphological integration. *Evolution & Development* 8:61–73.

Hallgrímsson, B., Jamniczky, H., Young, N. M., Rolian, C., Parsons, T. E., Boughner, J. C., and Marcucio, R. S. 2009. Deciphering the palimpsest: Studying the relationship between morphological integration and phenotypic covariation. *Evolutionary Biology* 36:355–376.

Hansen, T. F. and Houle, D. 2008. Measuring and comparing evolvability and constraint in multivariate characters. *Journal Of Evolutionary Biology* 21:1201–1219.

Hermisson, J., Hansen, T. F., and Wagner, G. P. 2003. Epistasis in polygenic traits and the evolution of genetic architecture under stabilizing selection. *The American Naturalist* 161:708–734.

Hermisson, J. and Wagner, G. P. 2004. The population genetic theory of hidden variation and genetic robustness. *Genetics* 168:2271–2284.

Houle, D., Bolstad, G. H., van der Linde, K., and Hansen, T. F. 2017. Mutation predicts 40 million years of fly wing evolution. *Nature* 548:447–450.

Houle, D., Mezey, J., Galpern, P., and Carter, A. 2003. Automated measurement of *Drosophila* wings. *BMC Evolutionary Biology* 3:25.

James, A. C., Azevedo, R. B., and Partridge, L. 1997. Genetic and environmental responses to temperature of *Drosophila melanogaster* from a latitudinal cline. *Genetics* 146:881–890.

Jones, A. G., Arnold, S. J., and Bürger, R. 2003. Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution* 57:1747–1760.

Kirkpatrick, M. 2009. Patterns of quantitative genetic variation in multiple dimensions. *Genetica* 136:271–284.

Klepsatel, P., Gáliková, M., Huber, C. D., and Flatt, T. 2014. Similarities and differences in altitudinal versus latitudinal variation for morphological traits in *Drosophila melangoaster*. *Evolution* 68:1385–1398.

Klingenberg, C. P. 2019. Phenotypic plasticity, developmental instability, and robustness: The concepts and how they are connected. *Frontiers in Ecology and Evolution* 7:56.

Klingenberg, C. P. and McIntyre, G. S. 1998. Geometric morphometrics of developmental instability: Analyzing patterns of fluctuating asymmetry with procrustes methods. *Evolution* 52:1363.

Klingenberg, C. P. and Zaklan, S. D. 2000. Morphological integration between developmental compartments in the *Drosophila* wing. *Evolution* 54:1273–1285.

Lack, J. B., Monette, M. J., Johanning, E. J., Sprengelmeyer, Q. D., and Pool, J. E. 2016a. Decanalization of wing development accompanied the evolution of large wings in high-altitude *Drosophila*. *Proceedings of the National Academy of Sciences* 113:1014–1019.

Lack, J. B., Yassin, A., Sprengelmeyer, Q. D., Johanning, E. J., David, J. R., and Pool, J. E. 2016b. Life history evolution and cellular mechanisms associated with increased size in high-altitude *Drosophila*. *Ecology and Evolution* 6:5893–5906.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: Body size allometry. *Evolution* 33:402–416.

Lande, R. and Arnold, S. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–1226.

Lehner, B. 2010. Genes confer similar robustness to environmental, stochastic, and genetic perturbations in yeast. *PloS one* 5:e9035.

Levy, S. F. and Siegal, M. L. 2008. Network hubs buffer environmental variation in *Saccharomyces cerevisiae*. *PLoS Biology* 6:e264.

Mayer, C. and Hansen, T. F. 2017. Evolvability and robustness: A paradox restored. *Journal of Theoretical Biology* 430:78–85.

Mcguigan, K. and Blows, M. W. 2007. The phenotypic and genetic covariance structure of *Drosphilid* wings. *Evolution* 61:902–911.

McKenzie, J. A. and Clarke, G. M. 1988. Diazinon resistance, fluctuating asymmetry and fitness in the australian sheep blowfly, *Lucilia cuprina*. *Genetics* 120:213–220.

Morris, M. R. J., Kaufman, R., and Rogers, S. M. 2019. Heterozygosity and asymmetry: Ectodysplasin as a form of genetic stress in marine threespine stickleback. *Evolution* 73:378–389.

Orgogozo, V., Morizot, B., and Martin, A. 2015. The differential view of genotype-phenotype relationships. *Frontiers in Genetics* 6:179.

Paaby, A. B. and Rockman, M. V. 2014. Cryptic genetic variation: evolution's hidden substrate. *Nature Reviews Genetics* 15:247–258.

Paaby, A. B., White, A. G., Riccardi, D. D., Gunsalus, K. C., Piano, F., and Rockman, M. V. 2015. Wild worm embryogenesis harbors ubiquitous polygenic modifier variation. *eLife* 4:e09178.

Palmer, A. R. 1994. Fluctuating asymmetry analyses: a primer. *In* Contemporary Issues in Genetics and Evolution, pp. 335–364.

Palmer, A. R. and Strobeck, C. 1986. Fluctuating asymmetry: Measurement, analysis, patterns. *Annual Review of Ecology and Systematics* 17:391–421.

Palmer, A. R. and Strobeck, C. 2003. Fluctuating asymmetry analyses revisited. *In* Developmental Instability: Causes and Consequences, pp. 279–319.

Partridge, L., Barrie, B., Fowler, K., and French, V. 1994. Evolution and development of body size and cell size in *Drosophila melanogaster* in response to temperature. *Evolution* 48:1269–1276.

Pavlicev, M., Cheverud, J. M., and Wagner, G. P. 2009. Measuring Morphological Integration Using Eigenvalue Variance. *Evolutionary Biology* 36:157–170.

Pélabon, C., Hansen, T., Carter, A., and Houle, D. 2006. Response of fluctuating and directional asymmetry to selection on wing shape in *Drosophila melanogaster*. *Journal of Evolutionary Biology* 19:764–776.

Pélabon, C., Hansen, T. F., Carter, A. J. R., and Houle, D. 2010. Evolution of variation and variability under fluctuating, stabilizing, and disruptive selection. *Evolution* 64:1912–1925.

Pitchers, W., Nye, J., Marquez, E. J., Kowalski, A., Dworkin, I., and Houle, D. 2019. A multivariate genome-wide association study of wing shape in *Drosophila melanogaster*. *Genetics* 211:1429–1447.

Pitchers, W., Pool, J. E., and Dworkin, I. 2013. Altitudinal clinal variation in wing size and shape in African *Drosophila melanogaster*: one cline or many? *Evolution* 67:438–452.

Pitchers, W., Wolf, J. B., Tregenza, T., Hunt, J., and Dworkin, I. 2014. Evolutionary rates for multivariate traits: the role of selection and genetic variation. *Phil. Trans. R. Soc. B* 369:20130252.

Pool, J. E., Corbett-Detig, R. B., Sugino, R. P., Stevens, K. A., Cardeno, C. M., Crepeau, M. W., Duchen, P., Emerson, J. J., Saelao, P., Begun, D. J., and Langley, C. H. 2012. Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. *PLoS Genetics* 8:e1003080.

Proulx, S. and Phillips, P. 2005. The opportunity for canalization and the evolution of genetic networks. *The American Naturalist* 165:147–162.

R Core Team 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rendel, J. 1963. The nature of concealed genotypic variation. *In* Proceedings-National Poultry Breeders, pp. 4–16.

Rockman, M. V. 2008. Reverse engineering the genotype-phenotype map with natural genetic variation. *Nature* 456:738–744.

Scharloo, W. 1991. Canalization: Genetic and developmental aspects. *Annual Review Of Ecology And Systematics* 22:65–93.

Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. *Evolution* 50:1766–1774.

Shu, W., Bo, X., Ni, M., Zheng, Z., and Wang, S. 2007. *In silico* genetic robustness analysis of microrna secondary structures: potential evidence of congruent evolution in microRNA. *BMC Evolutionary Biology* 7:223.

Siegal, M. L. and Bergman, A. 2002. Waddington's canalization revisited: developmental stability and evolution. *Proceedings of the National Academy of Sciences* 99:10528–10532.

Siegal, M. L. and Leu, J.-Y. 2014. On the nature and evolutionary impact of phenotypic robustness mechanisms. *Annual Review of Ecology, Evolution, and Systematics* 45:496–517.

Soto, I., Carreira, V., Soto, E., and Hasson, E. 2008. Wing morphology and fluctuating asymmetry depend on the host plant in cactophilic *Drosophila*. *Journal of Evolutionary Biology* 21:598–609.

Stearns, S., Kaiser, M., and Kawecki, T. 1995. The differential canalization of fitness components against environmental perturbations in *Drosophila melanogaster*. *J. Evol. Biol* 8:539–557.

Stearns, S. C. and Kawecki, T. J. 1994. Fitness sensitivity and the canalization of life-history traits. *Evolution* 48:1438–1450.

Szöllősi, G. J. and Derényi, I. 2009. Congruent evolution of genetic and environmental robustness in micro-RNA. *Molecular biology and evolution* 26:867–874.

Takahashi, K. H. 2017. Little effect of HSP90 inhibition on the quantitative wing traits variation in *Drosophila melanogaster*. *Genetica* 145:9–18.

Takahashi, K. H., Okada, Y., and Teramura, K. 2011. Genome-wide deficiency mapping of the regions responsible for temporal canalization of the developmental processes of *Drosophila melanogaster*. *The Journal of Heredity* 102:448–457.

Takahashi, K. H., Rako, L., Takano-Shimizu, T., Hoffmann, A. A., and Lee, S. F. 2010. Effects of small Hsp genes on developmental stability and microenvironmental canalization. *BMC Evolutionary Biology* 10:284.

Van Valen, L. 1974. Multivariate structural statistics in natural history. *Journal of Theoretical Biology* 45:235–247.

Van Valen, L. 2005. The statistics of variation, pp. 29–47. *In* B. Hallgrimsson and B. K. Hall (eds.), Variation, chapter 3. Elsevier.

Visser, J., Hermisson, J., Wagner, G., Meyers, L., Bagheri-Chaichian, H., Blanchard, J., Chao, L., Cheverud, J., Elena, S., Fontana, W., Gibson, G., Hansen, T., Krakauer, D., Lewontin, R., Ofria, C., Rice, S., Dassow, G., Wagner, A., and Whitlock, M. 2003. Perspective: Evolution and Detection of Genetic Robustness. *Evolution* 57:1959–1972.

Waddington, C. H. 1942. Canalization of development and the inheritance of acquired characters. *Nature* 150:563–565.

Wagner, G. P., Booth, G., and Bagheri-Chaichian, H. 1997. A Population Genetic Theory of Canalization. *Evolution* 51:329–347.

Walsh, B. and Blows, M. W. 2009. Abundant Genetic Variation plus Strong Selection - Multivariate Genetic Constraints: A Geometric View of Adaptation. *Annual Review of Ecology, Evolution, and Systematics* 40:41–59.

Supplemental Figures

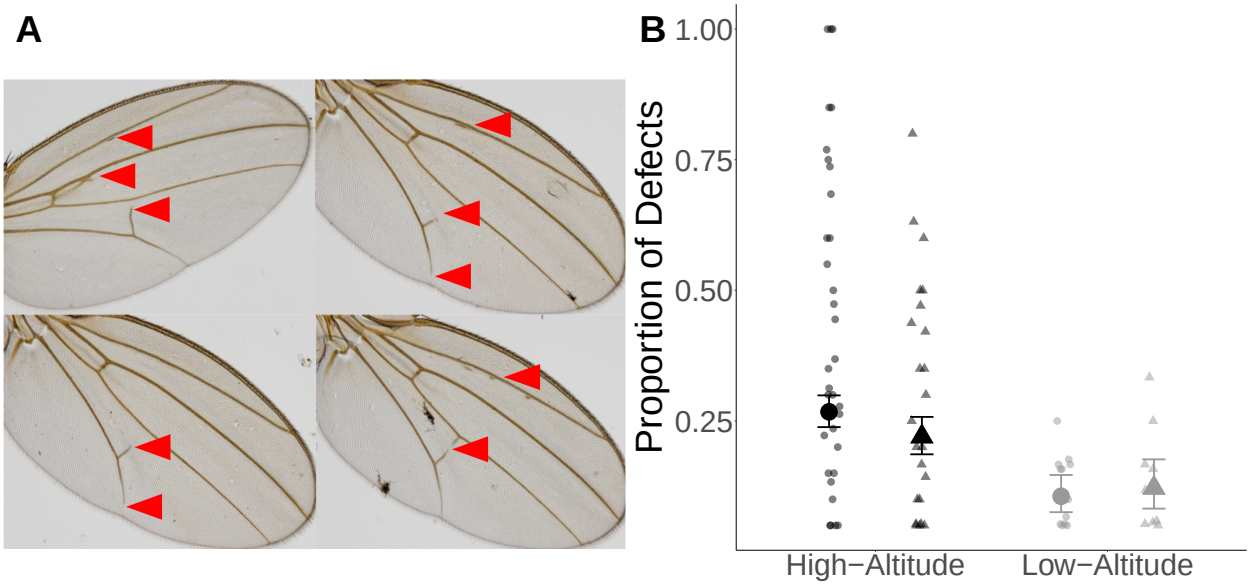

Figure S1: High-altitude population has a higher frequency of wing defects compared to the low-altitude population. (A) Examples of wing defects seen in the high-altitude population (red arrowheads pointing to venation defects) (B) Proportion of wing venation defects within lines is much greater in the high-altitude population than the low-altitude population

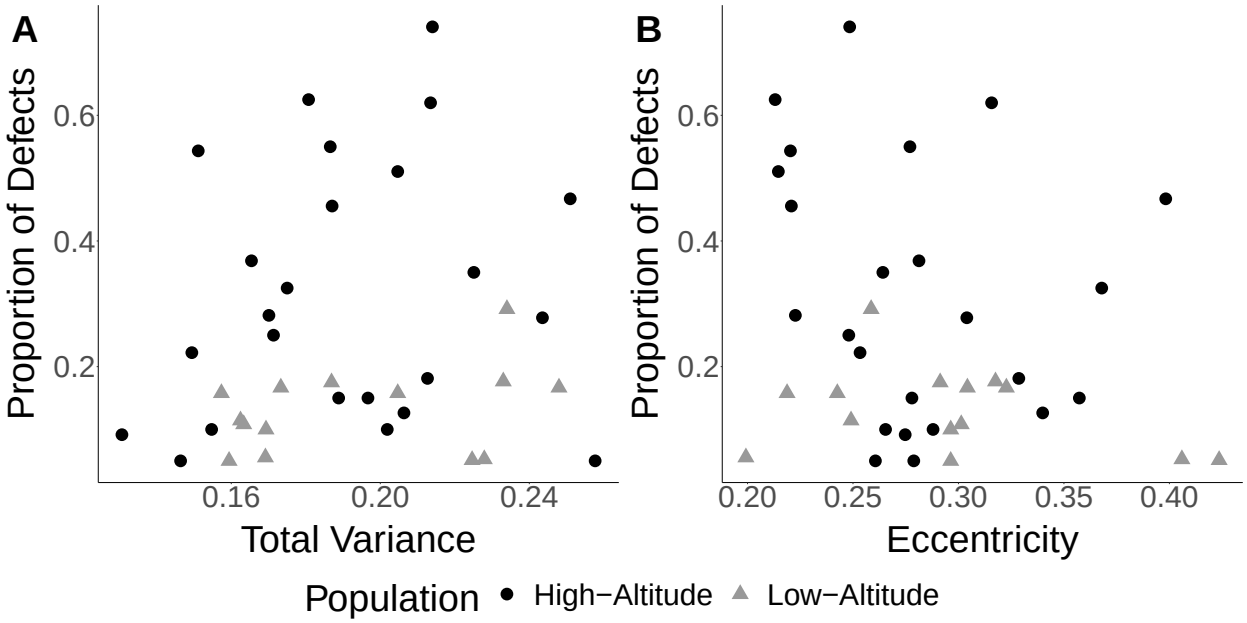

Figure S2: Measures of variability for wing size and shape are not correlated with proportion of wing defects. Total variance (A) and Eccentricity (B) are not correlated with the proportion of defects (Total variance: r = 0.16 95% CI −0.26 - 0.53; eccentricity: r = −0.24 95% CI −0.59 - 0.18) or the low-altitude population (Total variance: r = 0.34 95% CI −0.24 - 0.74; eccentricity: r = −0.33 95% CI −0.73 - 0.24)

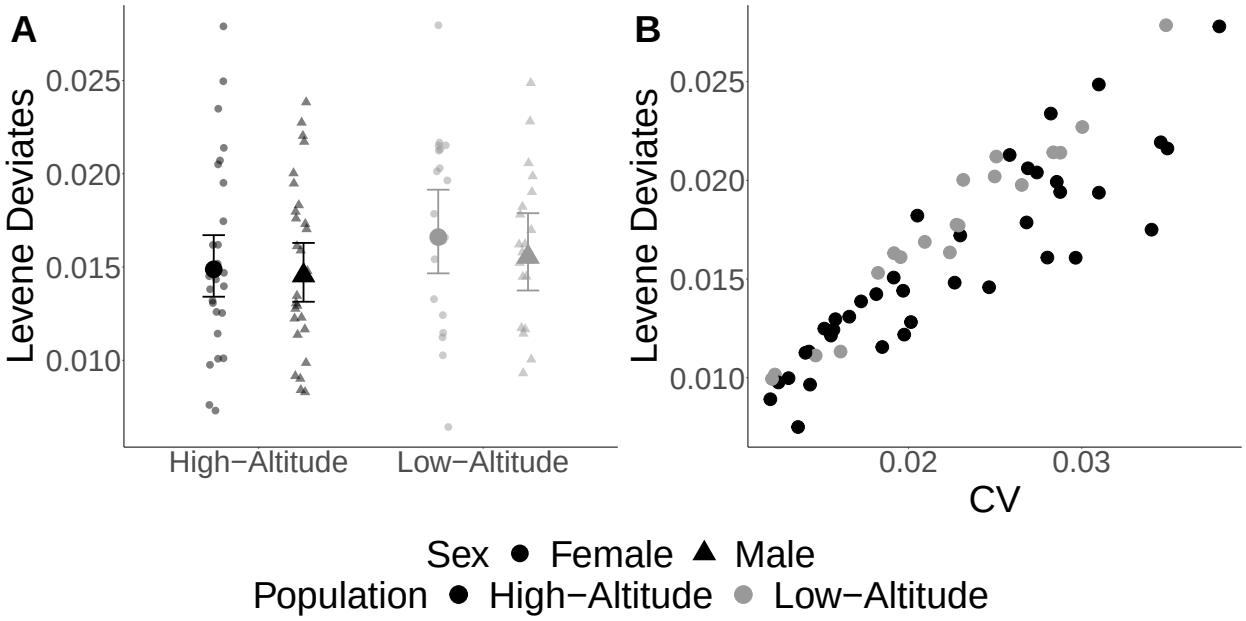

Figure S3: Levene's deviates as a measure of variability of wing size. (A) Consistent with the results using the coefficient of variation, Levene's deviates are similar for the high- and low-altitude populations is similar. (B) Levene's Deviates and CV are highly correlated (high-altitude r = 0.89 95% CIs 0.80 - 0.94; low-altitude r = 0.98 95% CIs 0.94 - 0.99)

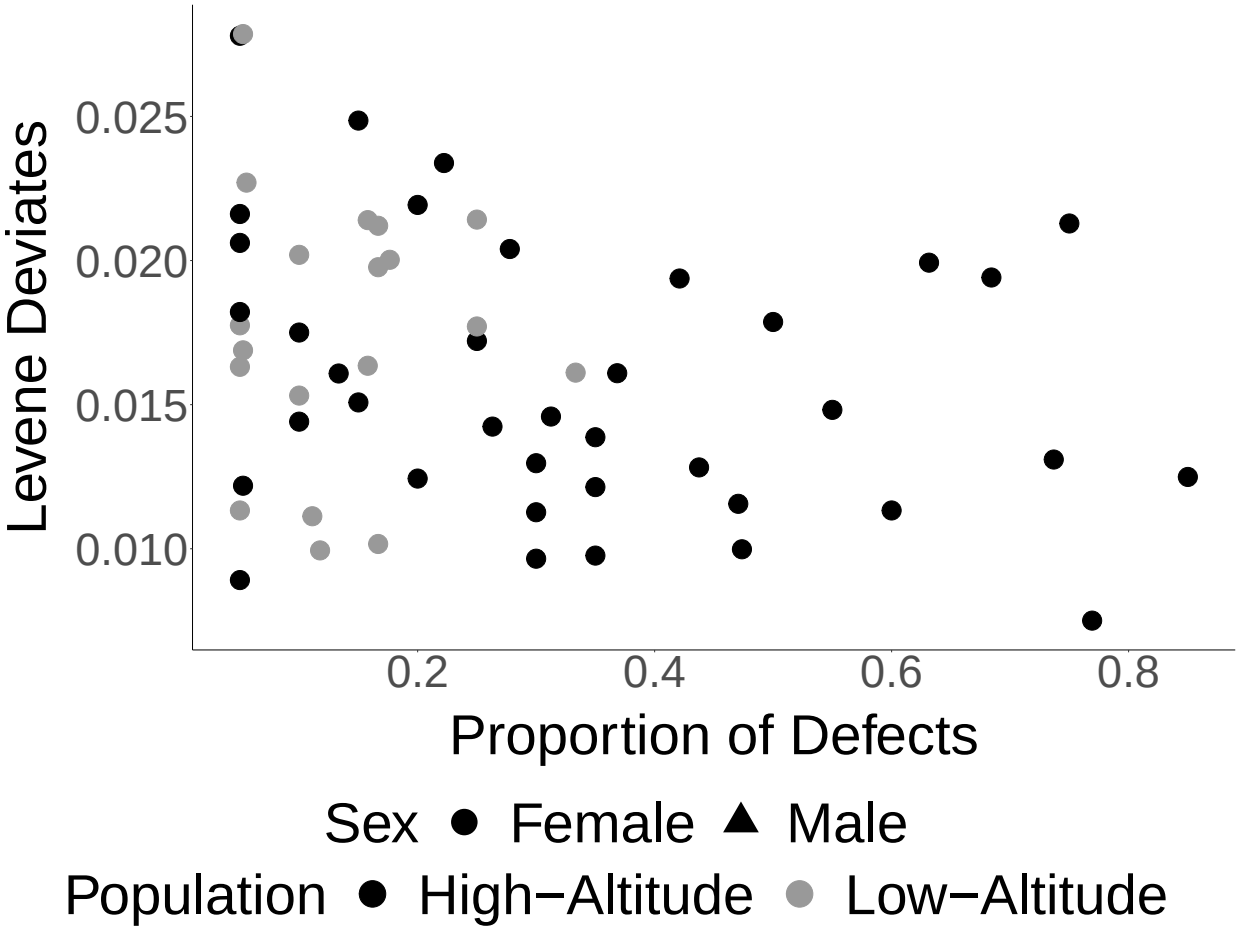

Figure S4: Consistent with the results using the coefficient of variation, (A) within-line Levene's deviates are not correlated with within-line proportion of defects in either the high-altitude population(r = −0.27 95% CIs −0.54 - 0.055) and the low-altitude population (r = −0.0032 95% CIs −0.46 - 0.45).

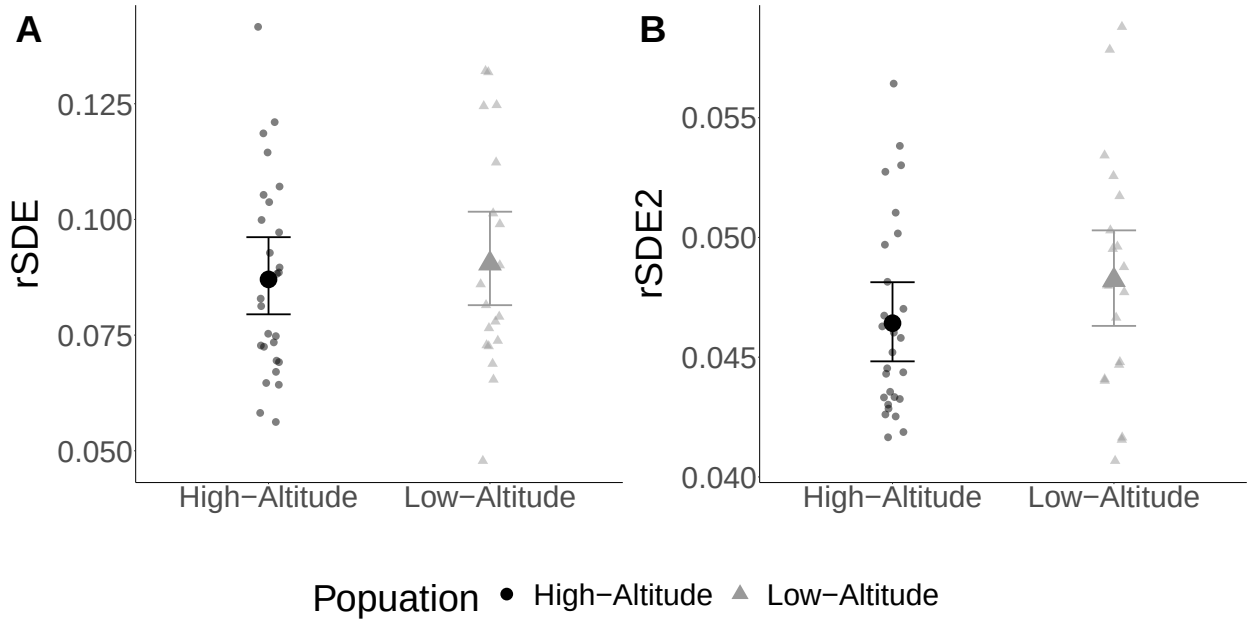

Figure S5: Measures of matrix shape for the phenotypic covariance matrix for wing shape estimated for each strain (A) rSDE and (B) rSDE2 are similar in the high- and low-altitude populations

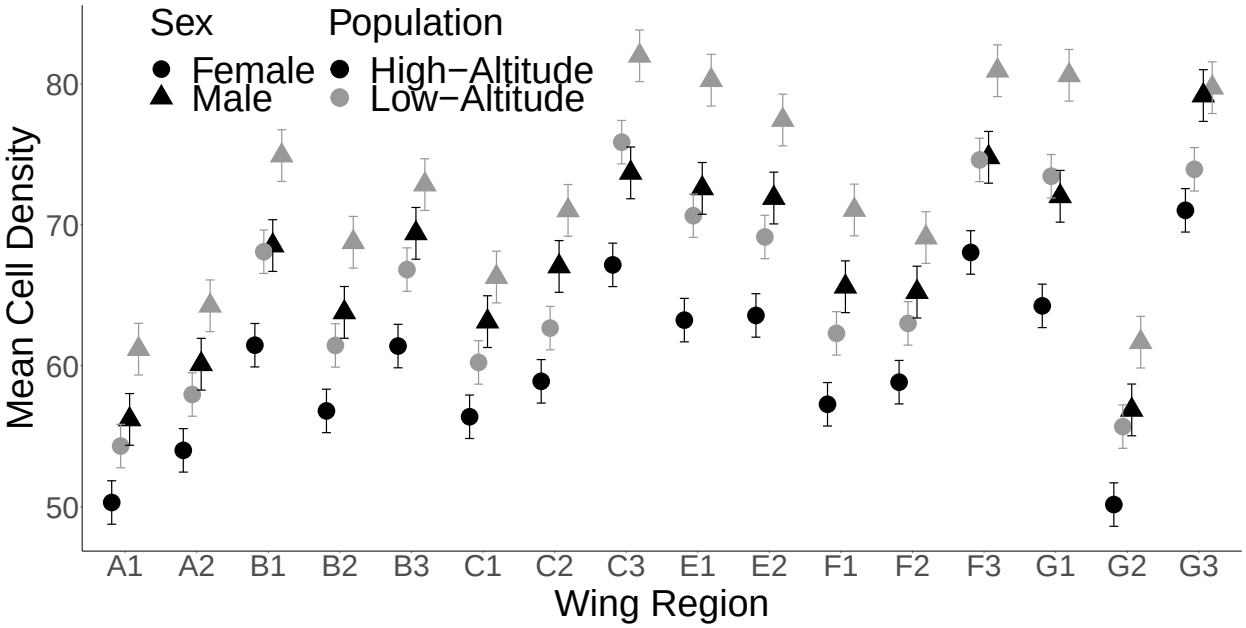

Figure S6: Mean cell density across 16 wing regions. Wing density varies by wing region, sex and population

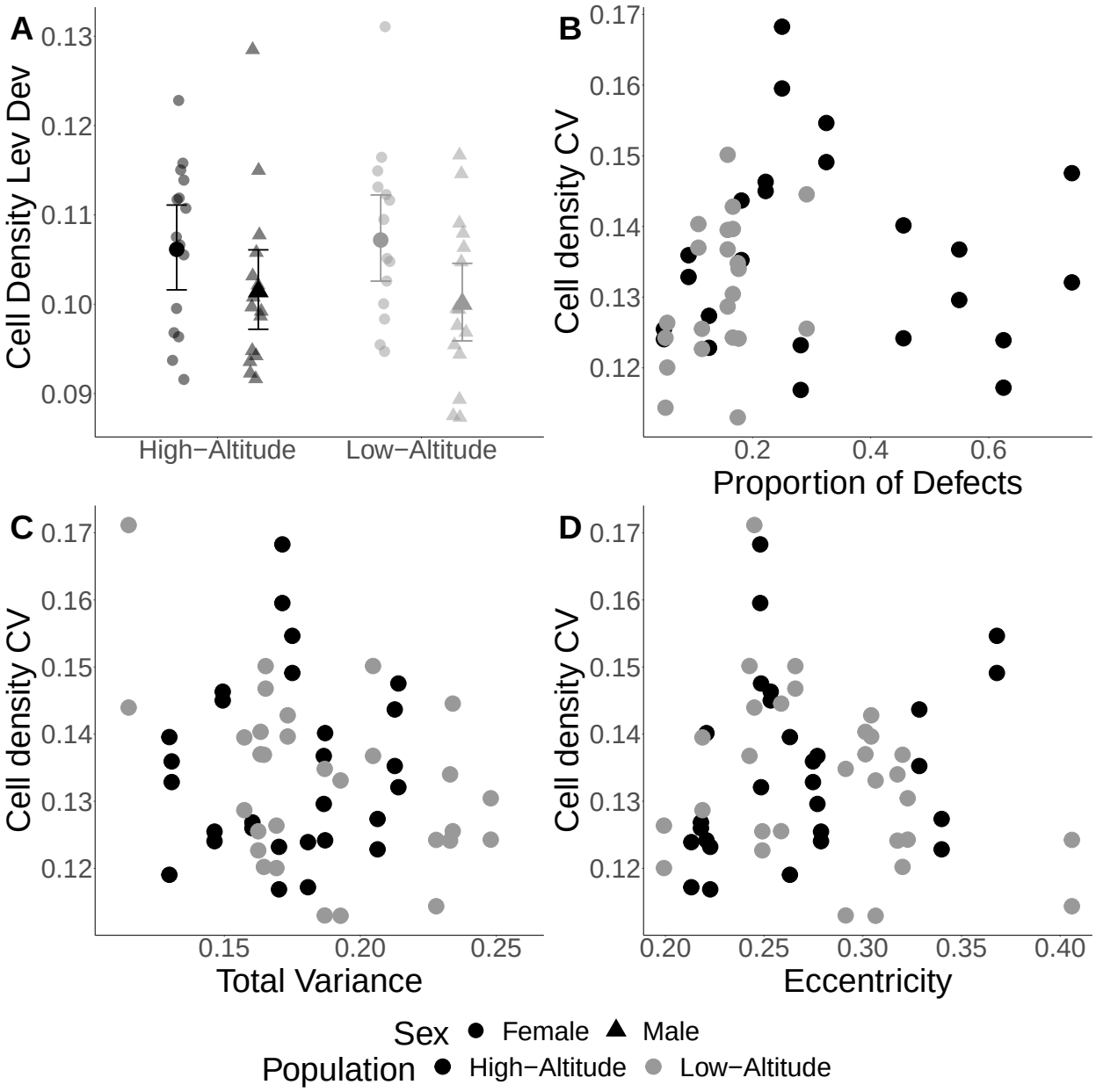

Figure S7: Cell density Levene's Deviates and association of cell density CV with within-line wing defects and measures of matrix shape for the covariance matrix for wing shape. (A) As seen using CV, Levene's deviates for cell density are similar between high- and low-altitude populations. (B) There is little association between cell density CV and proportion of defects in either the high-altitude (r = −0.043 95% −0.44 - 0.37) or low-altitude populations (r = 0.36 95% −0.076 - 0.68). (C) total variance and (D) eccentricity of wing shape are not correlated with within-line cell density CV (Total variance $r = -0.22$ 95% CIs −0.45 - 0.04) , Eccentricity $r = -0.04$ 95% CIs −0.29 - 0.22)

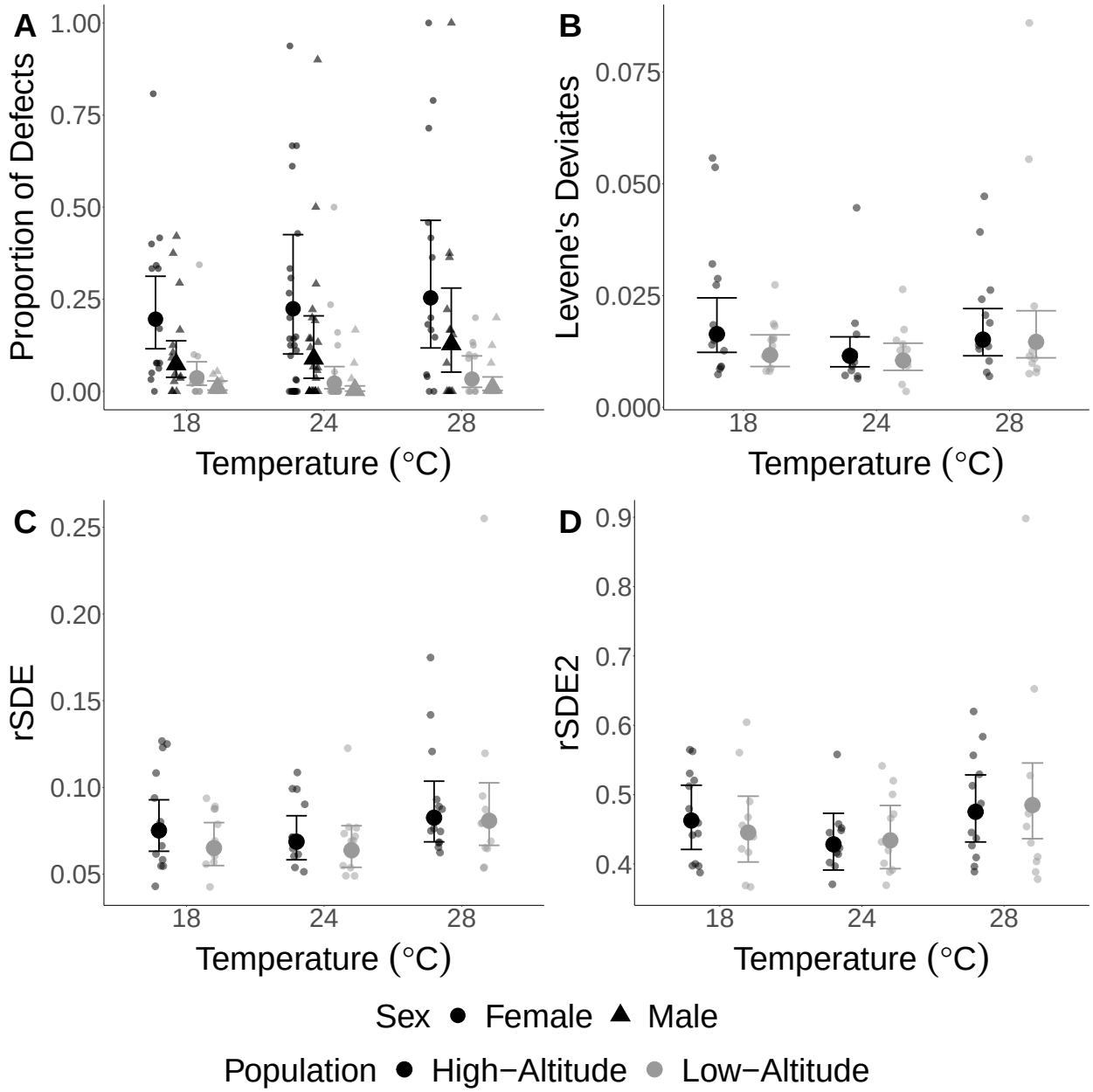

Figure S8: Proportion of wing defects and alternative measures of variation for wing size and shape for the high- and low-altitude population at different developmental temperatures. (A) Proportion of defects are similar across the different developmental temperatures with the high-altitude population having consistently greater proportion of defects than the low-altitude population.(B) Within-line Levene's Deviates for wing size and (C) within-line rSDE (multiplied by a factor of 10000) and (D) rSDE2 (multiplied by a factor of 10) for wing shape are similar in the high- and low-altitude populations across developmental temperatures.

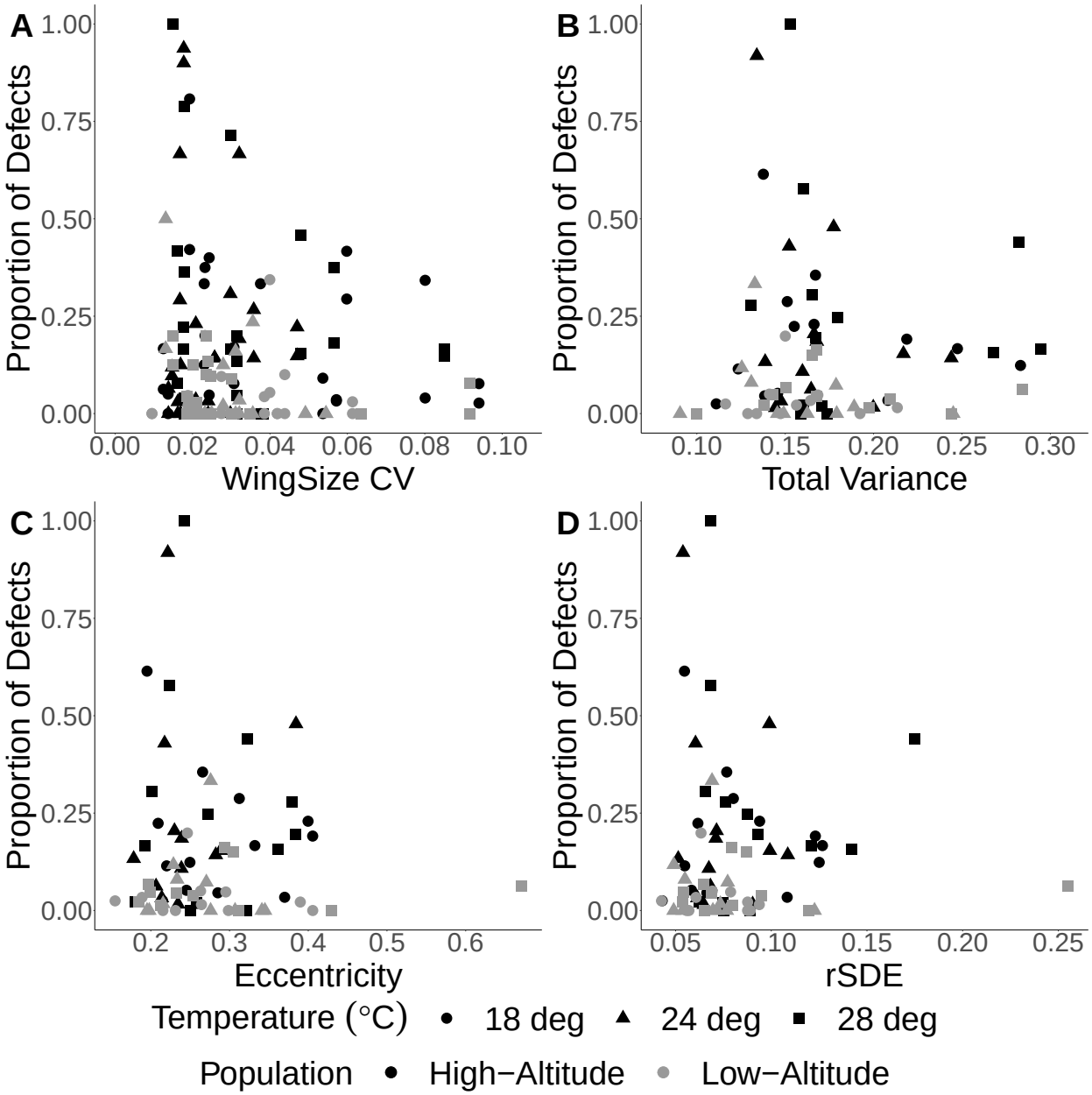

Figure S9: Little evidence for correlation between measures of variability for (A) wing size (CV) and (B-D) wing shape (total variance, eccentricity and rSDE) with proportion of defects for the high- and low-altitude populations at different developmental temperatures

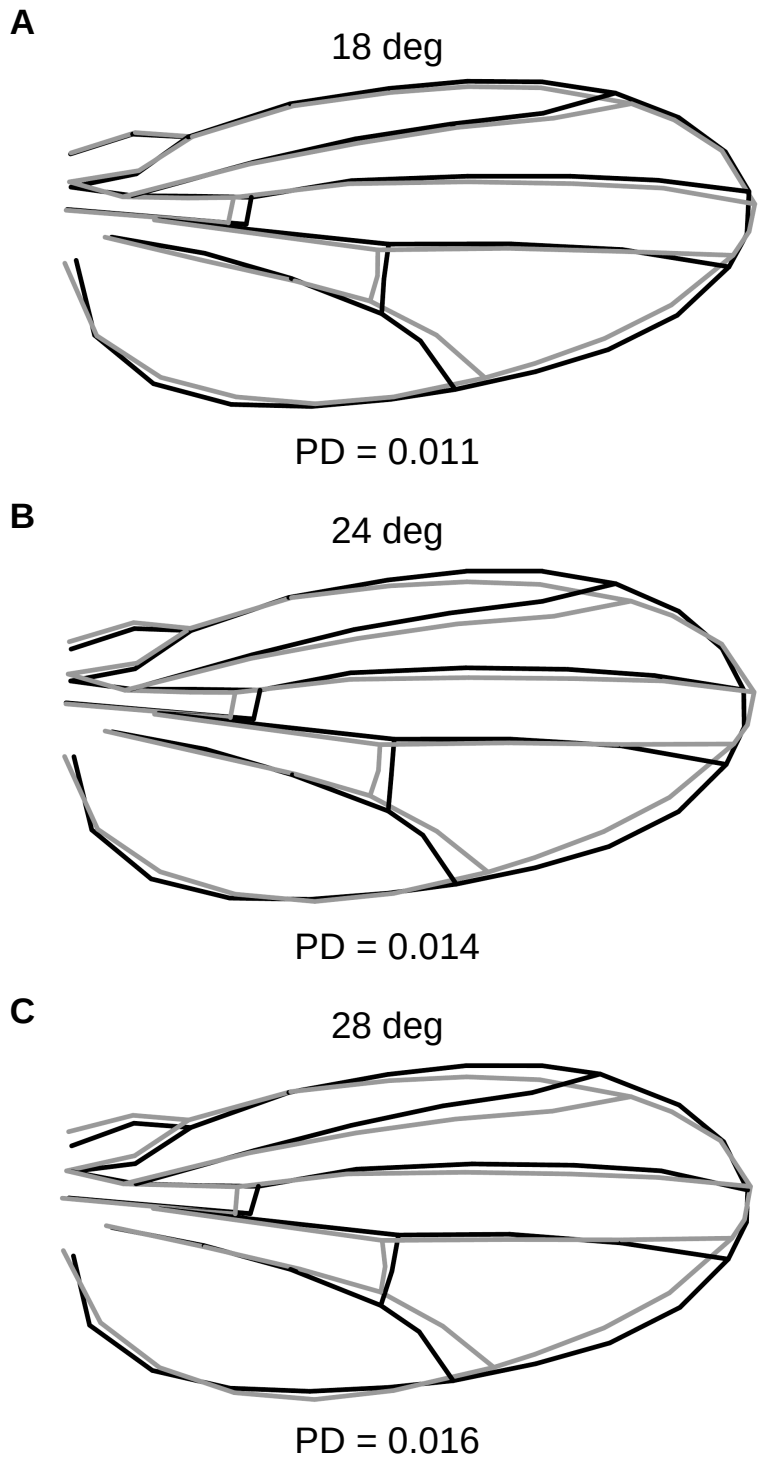

Figure S10: Wing shape differences between the high- and low-altitude populations at different temperatures (females). The difference in shape measured as PD (Procrustes distance) is increasing with temperature

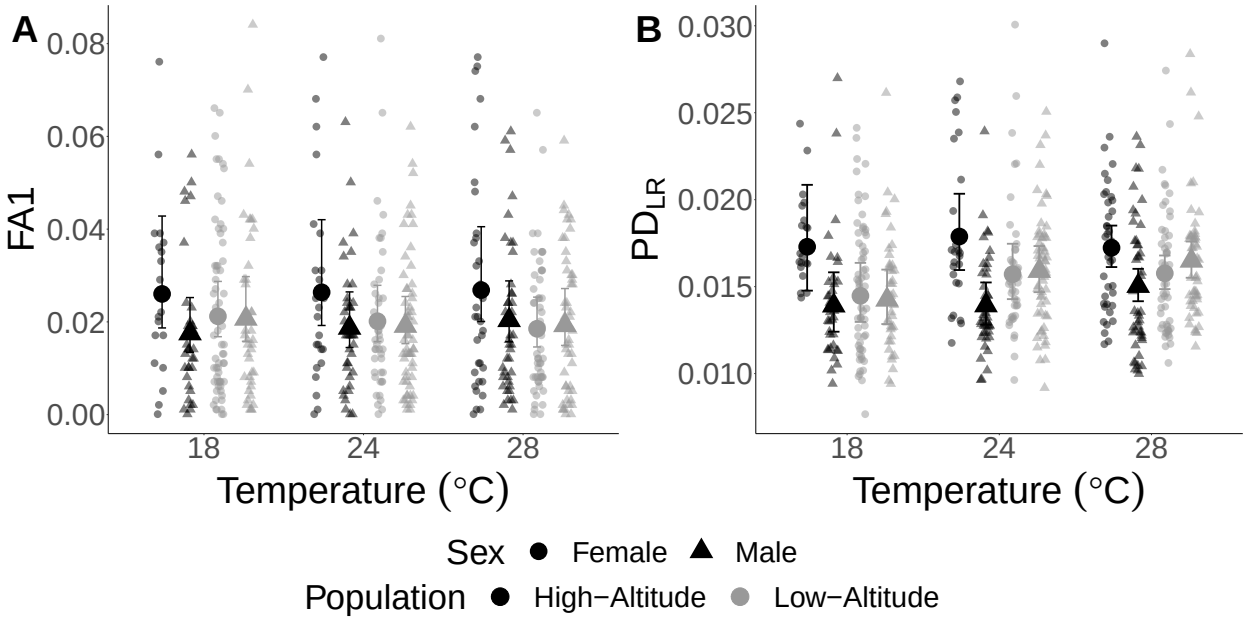

Figure S11: Fluctuating Asymmetry for wing size and shape represented by (A) FA1 and (B) PD_{LR}. High-altitude females have consistently greater FA1 and PD_{LR} across the three developmental temperature treatments. Large symbols represent population means, small symbols represent individuals, error bars represent 95% CIs

Supplemental Tables

Table S1: Sample Sizes of fly strains for micro-environmental canalization

Line	Population	N_F	N_M	Line	Population	N_F	N_M
ef101	High-Altitude	19	20	zi124	Low-Altitude	20	18
ef112	High-Altitude	20	20	zi159	Low-Altitude	18	15
ef115	High-Altitude	19	20	zi160	Low-Altitude	17	12
ef117	High-Altitude	20	19	zi186	Low-Altitude	17	17
ef119	High-Altitude	18	14	zi197	Low-Altitude	19	19
ef122	High-Altitude	18	20	zi216	Low-Altitude	16	17
ef131	High-Altitude	20	20	zi217	Low-Altitude	19	20
ef135	High-Altitude	19	20	zi251	Low-Altitude	19	14
ef136	High-Altitude	16	20	zi254	Low-Altitude	12	17
ef15	High-Altitude	15	20	zi311	Low-Altitude	17	19
ef16	High-Altitude	20	19	zi322	Low-Altitude	23	14
ef19	High-Altitude	20	19	zi337	Low-Altitude	19	20
ef1	High-Altitude	20	17	zi357	Low-Altitude	20	20
ef25	High-Altitude	20	20	zi360	Low-Altitude	18	14
ef31	High-Altitude	20	19	zi366	Low-Altitude	18	20
ef39	High-Altitude	20	20	zi383	Low-Altitude	19	19
ef3	High-Altitude	18	14	zi403	Low-Altitude	17	16
ef54	High-Altitude	20	19	zi455	Low-Altitude	20	19
ef59	High-Altitude	19	13	zi461	Low-Altitude	14	18
ef65	High-Altitude	19	19	zi507	Low-Altitude	20	18
ef67	High-Altitude	11	17				
ef73	High-Altitude	18	16				
ef75	High-Altitude	8	18				
ef7	High-Altitude	14	19				
ef86	High-Altitude	19	20				
ef98	High-Altitude	14	20				
ef9	High-Altitude	17	20				

Table S2: Sample Sizes of fly strains for Temperature Plasticity and Macro-environmental canalization

Line	Population	$N_{18^\circ CF}$	$N_{24^\circ CF}$	$N_{28^\circ CF}$	$N_{18^\circ CM}$	$N_{24^\circ CM}$	$N_{28^\circ CM}$
ef112	High-Altitude	28	10	22	18	23	9
ef117	High-Altitude	7	22	8	9	5	8
ef119	High-Altitude	20	24	11	16	9	8
ef122	High-Altitude	8	14	6	11	19	18
ef131	High-Altitude	26	27	41	35	39	22
ef136	High-Altitude	12	20	16	14	20	10
ef15	High-Altitude	13	19	21	19	24	14
ef19	High-Altitude	7	12	9	6	13	10
ef39	High-Altitude	7	13	17	12	21	23
ef65	High-Altitude	13	20	4	8	14	6
ef73	High-Altitude	36	21	12	23	26	8
ef16	High-Altitude	42	27	43	36	26	30
ef98	High-Altitude	25	16	13	19	20	8
zi124	Low-Altitude	47	20	31	39	21	25
zi159	Low-Altitude	41	72	42	39	60	31
zi186	Low-Altitude	26	8	21	37	8	23
zi216	Low-Altitude	26	47	10	29	56	4
zi251	Low-Altitude	20	32	17	23	47	28
zi254	Low-Altitude	40	40	40	38	49	36
zi311	Low-Altitude	22	22	15	13	16	12
zi360	Low-Altitude	9	17	5	20	9	4
zi366	Low-Altitude	4	12	8	14	6	2
zi367	Low-Altitude	8	35	23	8	23	10
zi455	Low-Altitude	32	32	15	35	37	16

Table S3: Recipe for 1.5:1 Protein:Sugar Food used in Temperature Plasticity/Macro-environmental canalization and FA experiments

Ingrident	Amount
Water	4250 ml
Black strap molasses	20 g
Fancy table molasses	20 g
Cornmeal	20 g
Carageenan	27 g
Yeast	280 g
Sucrose	75 g
Propionic acid	12 ml
Methylparaben	2.5 g
Ethanol	25 ml

Table S4: linear mixed estimates for the contributions of sex, population and their interaction on wing size

Effect	Chisq	Df	Pr(>Chisq)
Sex	4.6×10^3	1	$< 2.2 \times 10^{-16}$
Pop	9.5×10^2	1	$< 2.2 \times 10^{-16}$
Sex:Pop	1.0×10^2	1	$< 2.2 \times 10^{-16}$

Table S5: Results from Multivariate Procrustes ANOVA testing the effects of wing size, population, sex and all interactions on wing shape

Effect	Df	SS	MS	Rsq	F	Z	p (>F)
Size	1	5.7×10^{-2}	5.7×10^{-2}	5.8×10^{-2}	2.7×10^2	11	5.0×10^{-4}
Pop	1	1.1×10^{-1}	1.1×10^{-1}	1.1×10^{-1}	11	4.7	5.0×10^{-4}
Sex	1	1.0×10^{-2}	1.0×10^{-2}	1.0×10^{-2}	47	9	5.0×10^{-4}
Size:Pop	1	3.3×10^{-3}	3.3×10^{-3}	3.3×10^{-3}	16	6.4	5.0×10^{-4}
Size:Sex	1	1.3×10^{-3}	1.3×10^{-3}	1.3×10^{-3}	6.0	4.5	5.0×10^{-4}
Pop:Sex	1	5.5×10^{-4}	5.5×10^{-4}	5.6×10^{-4}	2.6	2.7	5.0×10^{-4}
Pop:Line	45	4.5×10^{-1}	1.0×10^{-2}	4.6×10^{-1}	47	35	5.0×10^{-4}
Size:Pop:Sex	1	4.0×10^{-4}	3.9×10^{-4}	4.0×10^{-4}	1.8	4.0	5.0×10^{-4}
Residuals	1.6×10^3	3.5×10^{-1}	2.1×10^{-4}	3.5×10^{-1}			
Total	1.7×10^3	9.8×10^{-1}					

Table S6: Results from generalized mixed effects model testing the effects of population and sex and their interaction on within-line among-individual wing size CV

Effect	Chisq	Df	Pr(>Chisq)
Pop	0.13	1	0.72
Sex	1.0×10^{-3}	1	0.98
Pop:Sex	0.41	1	0.52

Table S7: Results from generalized mixed effects model testing the effects of sex and population and their interaction on within-line among-individual wing size Levene's deviates

Effect	Chisq	Df	Pr(>Chisq)
Sex	0.84	1	0.36
Pop	1.5	1	0.22
Sex:Pop	0.20	1	0.66

Table S8: Results from generalized mixed effects model testing the effects of population and sex and their interaction on within-line proportion of wing defects

Effect	LR Chisq	Df	Pr(>Chisq)
Pop	44	1	3.6×10^{-11}
Sex	2.5	1	0.11
Pop:Sex	1.7	1	0.20

Table S9: Results from generalized mixed effects model testing the effects of population on within-line total variance for wing shape

Effect	LR Chisq	Df	Pr(>Chisq)
Pop	5.6×10^{3}	1	0.94

Table S10: Results from generalized mixed effects model testing the effects of population on within-line eccentricity for wing shape

Effect	LR Chisq	Df	Pr(>Chisq)
Pop	1.3	1	0.25

Table S11: Results from linear mixed effects model testing the effects of wing region, sex, population and their interactions on cell density across 16 different regions of the wing

Effect	Chisq	Df	Pr(>Chisq)
WingReg	1.9×10^4	15	$< 2.2 \times 10^{-16}$
Sex	5.6×10^2	1	$< 2.2 \times 10^{-16}$
Pop	31	1	2.0×10^{-8}
WingReg:Sex	1.1×10^2	15	$< 2.2 \times 10^{-16}$
WingReg:Pop	4.0×10^2	15	$< 2.2 \times 10^{-16}$
Sex:Pop	0.31	1	0.58
WingReg:Sex:Pop	20	15	0.16

Table S12: Results from generalized mixed effects model testing the effects of population and sex and their interaction on within-line among-individual cell density CV

Effect	Chisq	Df	Pr(>Chisq)
Pop	4.2×10^{-2}	1	0.84
Sex	10	1	1.0×10^{-3}
Pop:Sex	0.27	1	0.60

Table S13: Results from generalized mixed effects model testing the effects of population and sex and their interaction on within-line among-individual cell density Levene's deviates

Effect	Chisq	Df	Pr(>Chisq)
Sex	14	1	2.0×10^{-4}
Pop	3.0×10^{-3}	1	0.95
Sex:Pop	0.63	1	0.43

Table S14: Results from linear mixed effects model testing the effects of temperature, sex, population and their interactions on wing size

Effect	Chisq	Df	Pr(>Chisq)
Temp^2	6.5×10^2	2	$< 2.2 \times 10^{-16}$
Sex	1.8×10^3	1	$< 2.2 \times 10^{-16}$
Pop	1.1×10^2	1	$< 2.2 \times 10^{-16}$
Sex:Temp^2	16	2	3.0×10^{-4}
Pop:Temp^2	20	2	1.0×10^{-4}
Sex:Pop	22	1	4.0×10^{-6}
Sex:Pop:Temp^2 :	8.7	2	1.3×10^{-2}

Table S15: Results from Multivariate Procrustes ANOVA testing the effects of wing size, sex, temperature, population and all interactions on wing shape

Effect	Df	SS	MS	Rsq	F	Z	p (>F)
Size	1	7.6×10^{-2}	7.6×10^{-2}	3.6×10^{-2}	3.4×10^{2}	12	5.0×10^{-4}
Sex	1	0.14	0.14	6.7×10^{-2}	6.3×10^{2}	13	5.0×10^{-4}
Temp	1	0.13	0.13	6.1×10^{-2}	5.8×10^{2}	13	5.0×10^{-4}
$Temp^2$	1	2.0×10^{-2}	2.0×10^{-2}	9.4×10^{-3}	89	9.6	5.0×10^{-4}
Pop	1	5.90×10^{-2}	5.9×10^{-2}	2.8×10^{-2}	1.7	1.1	0.14
Size:Sex	1	2.6×10^{-3}	2.6×10^{-3}	1.5×10^{-3}	12	5.8	5.0×10^{-4}
Size:Temp	1	3.8×10^{-3}	3.8×10^{-3}	1.8×10^{-3}	17	6.4	5.0×10^{-4}
Size:$Temp^2$	1	1.4×10^{-3}	1.4×10^{-3}	6.6×10^{-4}	6.2	4.4	5.0×10^{-4}
Size:Pop	1	2.0×10^{-2}	2.0×10^{-2}	9.4×10^{-3}	89	10	5.0×10^{-4}
Sex:Temp	1	3.9×10^{-3}	3.9×10^{-3}	1.9×10^{-3}	18	6.5	5.0×10^{-4}
Sex:$Temp^2$	1	1.5×10^{-3}	1.5×10^{-3}	7.3×10^{-4}	6.9	4.7	5.0×10^{-4}
Sex:Pop	1	6.9×10^{-3}	6.9×10^{-3}	3.3×10^{-3}	31	7.5	5.0×10^{-4}
Pop:Temp	1	1.3×10^{-2}	1.3×10^{-2}	6.2×10^{-3}	59	9.0	5.0×10^{-4}
Pop:$Temp^2$	1	4.4×10^{-3}	4.4×10^{-3}	2.1×10^{-3}	20	6.8	5.0×10^{-4}
Pop:Line	25	0.89	3.5×10^{-2}	0.42	1.6×10^{2}	42	5.0×10^{-4}
Size:Sex:Temp	1	5.0×10^{-4}	5.0×10^{-4}	2.4×10^{-4}	2.0	4.4	5.0×10^{-4}
Size:Sex:$Temp^2$	1	4.5×10^{-4}	4.5×10^{-4}	2.1×10^{-4}	2.2	4.7	5.0×10^{-4}
Size:Sex:Pop	1	1.4×10^{-3}	1.4×10^{-3}	6.8×10^{-4}	6.4	7.2	5.0×10^{-4}
Size:Pop:Temp	1	6.0×10^{-4}	6.0×10^{-4}	2.9×10^{-4}	2.7	5.2	5.0×10^{-4}
Size:Pop:$Temp^2$	1	1.7×10^{-3}	1.7×10^{-3}	8.1×10^{-4}	7.6	7.6	5.0×10^{-4}
Sex:Pop:Temp	1	9.8×10^{-4}	9.8×10^{-4}	4.7×10^{-4}	4.4	6.4	5.0×10^{-4}
Sex:Pop:$Temp^2$	1	3.2×10^{-4}	3.2×10^{-4}	1.5×10^{-4}	1.4	3.7	5.0×10^{-4}
Residuals	3.3×10^{3}	0.73	2.2×10^{-4}	0.35			
Total	3.3×10^{3}	2.1					

Table S16: Results from generalized mixed effects model testing the effects population, temperature and their interaction on within-line among-individual CV for wing size

	Chisq	Df	Pr(>Chisq)
Pop	5.0×10^{-4}	1	0.99
Temp	8.1	2	1.8×10^{-2}
Pop:Temp	5.2	2	7.5×10^{-2}

Table S17: Results from generalized mixed effects model testing the effects of population, temperature and thier interaction on within-line among-individual Levene's deviates for wing size

	Chisq	Df	Pr(>Chisq)
Pop	0.67	1	0.41
Temp	14	2	1.0×10^{-3}
Pop:Temp	4.4	2	0.11

Table S18: Results from generalized mixed effects model testing the effects of population, temperature and their interaction on within-line among-individual total variance for wing shape

Effect	Chisq	Df	Pr(>Chisq)
Pop	0.44	1	0.51
Temp	10	2	6.0×10^{-3}
Pop:Temp	0.85	2	0.65

Table S19: Results from generalized mixed effects model testing the effects of population, temperature and their interaction on within-line among-individual eccentricity for wing shape

Effect	Chisq	Df	Pr(>Chisq)
Pop	4.0×10^{-3}	1	0.95
Temp	3.7	2	0.15
Pop:Temp	0.91	2	0.63

Table S20: Results from generalized mixed effects model testing the effects of temperature, populations sex and all interactions on within-line proportion of wing defects

Effect	Chisq	Df	Pr(>Chisq)
Temp	1.2	2	0.55
Sex	70	1	$< 2.2 \times 10^{-16}$
Pop	15	1	1.0×10^{-4}
Temp:Sex	0.94	2	0.62
Temp:Pop	2.4	2	0.30
Sex:Pop	1.8	1	0.19
Temp:Sex:Pop	0.52	2	0.77

Table S21: Results from generalized mixed effects model testing the effects of sex, population, temperature and all interactions on FA1

Effect	Chisq	Df	Pr(>Chisq)
Sex	3.2	1	7.2×10^{-2}
Pop	1.1	1	0.28
Temp	0.11	2	0.94
Sex:Pop	3.9	1	4.7×10^{-2}
Sex:Temp	0.34	2	0.84
Pop:Temp	1.0	2	0.60
Sex:Pop:Temp	9.4×10^{-2}	2	0.95

Table S22: Results from generalized mixed effects model testing the effects of sex, population, temperature and all interactions on FA8

Effect	Chisq	Df	Pr(>Chisq)
Sex	9.0×10^{-3}	1	0.92
Pop	1.0×10^{-2}	1	0.92
Temp	1.0	2	0.60
Sex:Pop	10	1	1.0×10^{-3}
Sex:Temp	0.80	2	0.67
Pop:Temp	2.4	2	0.29
Sex:Pop:Temp	0.84	2	0.66

Table S23: Results from generalized mixed effects model testing the effects of sex, population, temperature and all interactions on PD_{LR}

Effect	Chisq	Df	Pr(>Chisq)
Sex	12	1	4.0×10^{-4}
Pop	0.77	1	0.38
Temp	5.2	2	7.5×10^{-2}
Sex:Pop	28	1	1.2×10^{-7}
Sex:Temp	3.2	2	0.20
Pop:Temp	2.2	2	0.34
Sex:Pop:Temp	0.91	2	0.64

Table S24: Results from generalized model testing the effects of wing size, sex, population, temperature and all interactions on wing shape FA after correcting for DA

	LR Chisq	Df	Pr(>Chisq)
Pop	1.4	1	0.24
Sex	4.1	1	0.043
Temp	0.90	2	0.64
Size	3.2	1	0.075
Pop:Sex	1.3	1	0.25
Pop:Temp	2.0	2	0.37
Pop:Size	1.3	1	0.26
Sex:Temp	1.6	2	0.44
Sex:Size	0.070	1	0.79
Temp:Size	1.0	2	0.60
Pop:Sex:Temp	0.15	2	0.93
Pop:Sex:Size	2.7	1	0.10
Pop:Temp:Size	8.8	2	0.012
Sex:Temp:Size	2.1	2	0.36

Table S25: Results from wing size ANOVA on repeated measurements using side, individual and their interaction as effects in order to estimate measurement error

Effect	Sum Sq	Df	F value	Pr(>F)
Side	4.0×10^{-4}	1	14	3.0×10^{-4}
Inds	1.0×10^{2}	76	4.2×10^{4}	$< 2.2 \times 10^{-16}$
Side:Inds	0.31	76	1.3×10^{2}	$< 2.2 \times 10^{-16}$
Residuals	5.0×10^{-3}	154		

Table S26: Results from wing shape multivariate procrustes ANOVA on repeated measurements using side, individual and their interaction as effects in order to estimate Wing shape Measurement Error ANOVA

Effect	Df	SS	MS	Rsq	F	Z	Pr(>F)
Side	1	8.7×10^{-4}	8.7×10^{-4}	5.0×10^{-3}	15	4.8	1.0×10^{-3}
Inds	76	0.15	1.9×10^{-3}	0.86	33	25	1.0×10^{-3}
Side:Inds	76	1.4×10^{-2}	1.9×10^{-4}	8.4×10^{-2}	3.2	28	1.0×10^{-3}
Residuals	154	9.1×10^{-3}	5.9×10^{-5}	5.3×10^{-2}			
Total	307	0.17					