ORIGINAL ARTICLE

doi:10.1111/evo.14

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]*E-mail: dworkin@mcmaster.ca*

Received July 26, 2019

Accepted May 30, 2020

Organisms are exposed to environmental and mutational effects influencing both mean and variance of phenotypes. Potentially deleterious effects arising from this variation can be reduced by the evolution of buffering (canalizing) mechanisms, ultimately reducing phenotypic variability. There has been interest regarding the conditions enabling the evolution of canalization. Under some models, the circumstances under which genetic canalization evolves are limited despite apparent empirical evidence for it. It has been argued that genetic canalization evolves as a correlated response to environmental canalization (congruence model). Yet, empirical evidence has not consistently supported predictions of a correlation between genetic and environmental canalization. In a recent study, a population of *Drosophila* adapted to high altitude showed evidence of genetic decanalization relative to those from low altitudes. Using strains derived from these populations, we tested if they varied for multiple aspects of environmental canalization We observed the expected differences in wing size, shape, cell (trichome) density and mutational defects between high- and low-altitude populations. However, we observed little evidence for a relationship between measures of environmental canalization with population or with defect frequency. Our results do not support the predicted association between genetic and environmental canalization.

KEY WORDS: **Adaptation, body size, canalization, cell size, *Drosophila melanogaster*, geometric morphometrics, high altitude, phenotypic integration, wing shape.**

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; Gibson and Wagner 2000; Visser et al. 2003; Flatt 2005; Pélabon et al. 2010; Siegal and Leu 2014; Félix and Barkoulas 2015; Mayer and Hansen 2017; Klingenberg 2019). 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 1983, 2003). The sensitivity of a given genotype in its response to mutational or environmental influences can vary *among* genotypes. It has been empirically 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 (Scharloo 1991; Gibson and van Helden 1997; Dworkin 2005a; Paaby and Rockman 2014; Paaby et al. 2015). Theory suggests that robustness to environmental variation—environmental canalization—can readily evolve as organisms are constantly exposed to the influence of environmental effects (Wagner et al. 1997; Eshel and Matessi 1998). Yet, as deleterious mutations are often purged by natural selection, this can result in weak selection for genetic canalization (assuming stabilizing selection on the trait), making it potentially less likely to evolve (Wagner et al. 1997; Gibson and Wagner 2000; Visser et al. 2003; Proulx and Phillips 2005).

The congruence hypothesis was proposed as a solution to the 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 (Stearns and Kawecki 1994; Stearns et al. 1995; Szöllősi and Derényi 2009; Lehner 2010). However, some explicit tests for the congruence model did not find support for it (Dworkin 2005a, 2005c; Borenstein and Ruppin 2006). The most likely explanation is that the evolution of genetic and environmental canalization is 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. First, the conditions in which the release of cryptic genetic variation can be used to infer genetic canalization may be more limited than once thought (Hermisson and Wagner 2004; Geiler-Samerotte et al. 2019). Rather, mutation accumulation or mutagenesis experiments are likely to be more fruitful for investigating genetic canalization. Second, 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 (Camara and Pigliucci 1999; Dworkin 2005a, 2005c; Chari and Dworkin 2013; 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 (Levy and Siegal 2008; Debat et al. 2009; Haber and Dworkin 2017). This suggests that 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 influence of both lab adaptation (domestication) and the use of lab-induced mutations. Most studies of canalization and robustness use lineages that have likely undergone some degree of adaption to lab environments. Furthermore, many studies often use lab-induced mutations as a source for genetic perturbations (Gibson and van Helden 1997; Dworkin 2005a, 2005c; Hallgrímsson et al. 2006; Levy and Siegal 2008; Debat et al. 2011; Paaby et al. 2015; Haber and Dworkin 2017). However, lab domestication and induced mutations may be unrepresentative of natural populations (Rockman 2008; Orgogozo et al. 2015; Dittmar et al. 2016). Lab-induced mutations may not reflect the spectrum of mutational effects experienced by natural populations. This may bias inferences regarding the ability of genotypes to buffer the effects of mutations that 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 that have heterogeneous geographical origins, and/or have been maintained in the lab for long periods of time are used (i.e., Dworkin 2005a, 2005c). 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 that examine properties of phenotypic robustness and canalization is that they have examined variation with a univariate perspective (Dworkin 2005a, 2005c), even when examining many traits (Levy and Siegal 2008; Takahashi et al. 2010, 2011; 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 1979; Lande and Arnold 1983; Schluter 1996; Mcguigan and Blows 2007; Hansen and Houle 2008; Agrawal and Stinchcombe 2009; Walsh and Blows 2009; Blows and Mcguigan 2014; Pitchers et al. 2014; Houle et al. 2017). Yet, this perspective has only been considered in a modest number of studies examining variational properties of phenotypes (Cheverud et al. 1983; Breuker et al. 2006; Debat et al. 2006; Hallgrímsson et al. 2006; Debat et al. 2009; Hallgrímsson et al. 2009; Pavlicev et al. 2009; Debat et al. 2011; Green et al. 2017). When considering properties of trait (co)variation in this perspective, it is not just the magnitude of variation (matrix size), but direction (of major axes of variation) and the shape of the variance-covariance matrix (a proxy for trait integration) that need to be considered as well. In a recent study examining variation in $\mathbf{E}$ (from $\mathbf{P} = \mathbf{G} + \mathbf{E}$) across naturally derived strains and lab-induced mutations, it was demonstrated that changes to trait means and relative orientation (directions of major axes of variation) of phenotypic (co)variances matrices were more variable than phenotypic integration (Haber and Dworkin 2017). This suggests that a multivariate perspective needs to be consistently applied to studies examining trait (co)variation (Klingenberg 2019).

Thus, what has been lacking for empirical studies testing evolutionary models of canalization is a system with the relevant natural history that can be studied with a multivariate approach. Lack et al. (2016a), among other recent studies, demonstrated that populations of *Drosophila melanogaster* from sub-Saharan Africa have recently adapted to a high-altitude environment. As is common for small insects evolving to

high-altitude environments (Dillon et al. 2006), the high-altitude Ethiopian population has evolved increase in cold tolerance (Pool et al. 2016) and melanism (Bastide et al. 2016). They have also adapted via increased body size, wing size, and shape (Pitchers et al. 2013; Klepsatel et al. 2014; Fabian et al. 2015; Lack et al. 2016a, 2016b; Bastide et al. 2016), likely to deal with changes in flight response in cold, thin air (reviewed in Dillon et al. 2006). Intriguingly, there is a substantial increase in the frequency of qualitative mutational defects of wing morphology in the high-altitude population (Lack et al. 2016a). Partially inbred strains derived from a high-altitude Ethiopian 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 influence 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) that have been artificially selected for larger wing size. Currently, it is inferred that the increase in mutational sensitivity in the high-altitude population may have been a result of strong directional selection on size, leading to rapid adaptation, with negative pleiotropic consequences. Compared with the ancestral low-altitude population that likely experienced a long history of stabilizing selection (and thus potentially promoting the evolution of canalization), the increase in size due to adaptation to conditions at high altitude, or due to strong artificial selection (Groth et al. 2018) have resulted in the evolutionary loss of canalization. This may represent a situation similar to that envisioned by Waddington (1942), where the population has not yet reevolved 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. Although there is considerable genetic variation within populations, strains derived from a high-altitude population in Ethiopia from an elevation of ∼3000 m are much larger in body size and wing size, have distinct wing shapes, and have a greater frequency of qualitative "mutant" phenotypes than low-altitude populations from Zambia from an elevation of ∼500 m (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 (Klingenberg and Zaklan 2000; Breuker et al. 2006; Debat et al. 2006; Pélabon et al. 2006; Soto et al. 2008; Debat et al. 2011; Haber and Dworkin 2017). In this study, we compared environmental canalization between the high- and low-altitude populations. We used different measures of environmental canalization: within-line, among-individual variation (microenvironmental variation), within-individual variation (fluctuating asymmetry) and phenotypic plasticity across a temperature gradient (macroenvironmental variation). Further, we examined associations between different measures of environmental canalization and mutational perturbation to test the congruence hypothesis and determine whether strains with greater proportion of defects are also more variable (also known as 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 microenvironmental and macroenvironmental 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, 2016b). 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. These strains underwent inbreeding in the lab, which is expected to substantially reduce the effects of lab adaptation (because of the small N_e within each strain), but this does result in substantial genetic drift within lines (but should not substantially alter allele frequencies among lines).

The flies for the microenvironmental 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. This experiment was performed in June 2013.

A subset of the strains described above were used for the temperature plasticity (macroenvironmental variation) and fluctuating asymmetry experiments. These strains 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. 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. This experiment

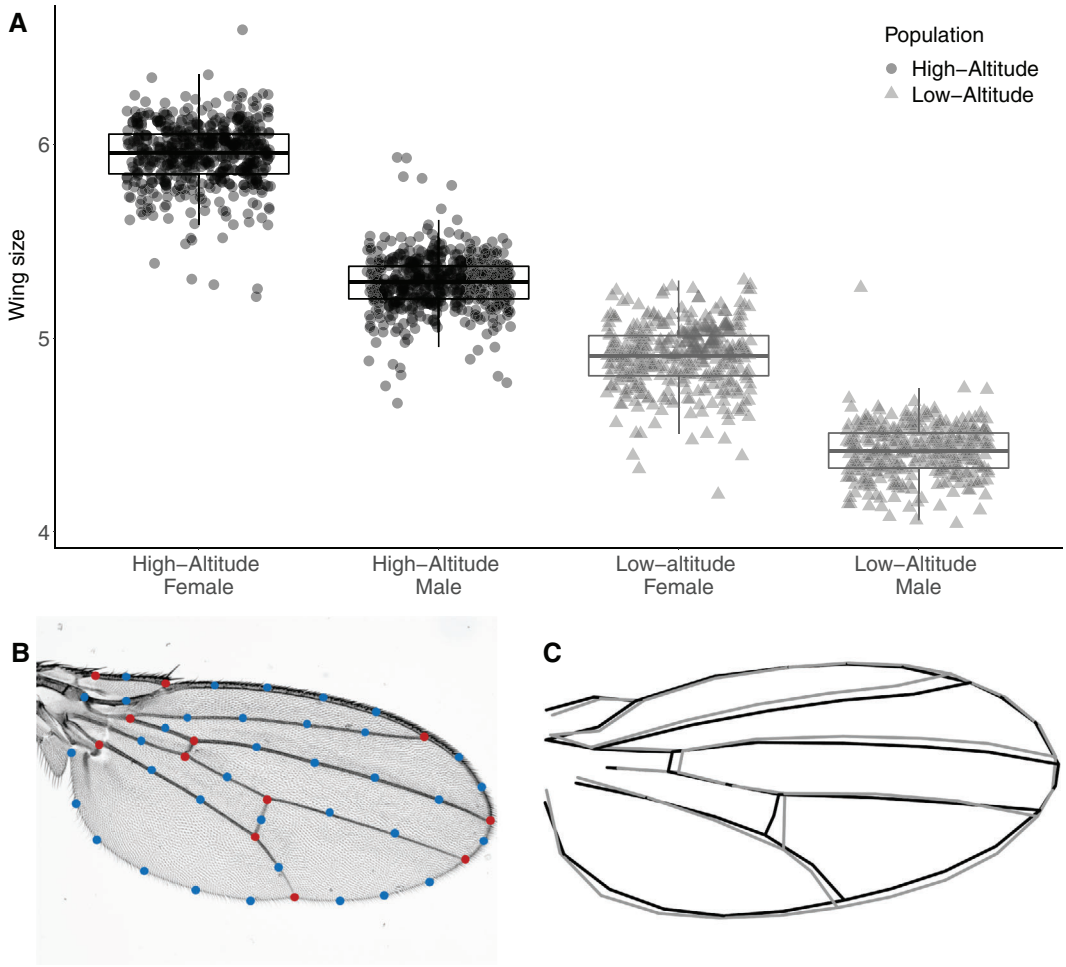

Figure 1. **Wing size (centroid size) and shape variation within and among high- and low-altitude populations. (A) High-altitude population has larger wings than the low-altitude population (B) Landmarks (red) and semi-landmarks (blue) used in the analysis of wing shape. (C) Mean difference in wing shape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

was performed in 2017. Although these strains are partially inbred, as a check, we confirmed that the phenotypic effects of the size-related traits from these lineages remained correlated with the low-altitude populations (where variation for size is considerable), and also showed the same overall patterns (for mean size, shape, and defect frequencies).

PHENOTYPING

Wing size and shape—Microenvironmental canalization

The right wing of each fly was dissected and imaged using an Olympus DP30B camera mounted on an Olympus BX51 microscope (Olympus software version 3,1,1208) using a 2× objective (20× total magnification). Landmark and semilandmark data were 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 (version 2.16). 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 (version 3.7), reviewed and manually adjusted, if necessary. Landmark and semilandmark positions were extracted and the shape information after adjusting for size, position, and rotation information using CPReader software (version 1.12r). This produced data composed of 12 landmarks and 36 semilandmarks (Fig. 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 initial size and shape analysis were re-imaged with a higher resolution camera. We chose 15 strains from each population as follows: five strains each were chosen with the highest and lowest within-line coefficient of variation (CV) for wing size. Additionally, five strains 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 strain. Wings were imaged using an Olympus DP80 camera mounted on an Olympus BX43 microscope, using a 4× objective (total 40× magnification). Images were captured with cellSens Standard (version 1.14) software. Cell density was quantified by counting trichomes on the surface of the wing using the ImageJ FijiWings macro (version 2.2) (Dobens and Dobens 2013). Each trichome represents a single cell (Dobzhansky 1929). We used a 0.0065 mm^2 (75 × 75 px) measurement area in each of 16 different locations in the wing (Fig. 4A).

Wing size and shape—Macroenvironmental 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 microenvironmental canalization experiment. We used three replicate vials per strain 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 individuals: 509) to assess 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 microenvironmental canalization experiment.

Quantification of wing defects

Each wing image was manually scored for venation defects. For the microenvironmental 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 macroenvironmental 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 were analyzed using R (v3.5.1) (R Core Team 2018) in RStudio (version 1.1.456) on a MacBook Pro, running macOS Mojave (version 10.14.2). Mixed models were run using *lmer* and *glmer* from the package *lme4* (version 1.1.19) (Bates et al. 2015), glmmTMB (Brooks et al. 2017), and *procD.lm* from the package *geomorph* (version 3.0.7) (Adams et al. 2018). Generalized linear models were run using *glm* from the *stats* package (version 3.5.1) (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 linear mixed model (binomial distribution, with a logit link) was fit with the wing defects data using population, sex, rearing temperature (for macroenvironmental canalization experiment), and their interactions as fixed effects. Where possible, the intercept and sex effects were allowed to vary as random effects of line (strain). 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 the temperature plasticity experiment, we used a model similar to the one described above, but allowing temperature effects to vary according to linear and quadratic effects (using second-degree orthogonal polynomials), including interaction effects with temperature and within the random effects of line nested within population.

For wing shape, we fit a multivariate linear model using *procD.lm* estimating the contributing effects of centroid 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 CV ($CV = \frac{\sigma}{\mu}$) and the median form of Levene's deviates (used for all formal statistical inference) (Van Valen 2005; Dworkin 2005b). For the macroenvironmental canalization experiment, sex effects were first modeled out and then CV and Levene's deviates were calculated for each line at each temperature.

To capture some of the multivariate variational properties for wing shape, we focused on two measures estimated for each strain. First we used matrix size (total variance), which is the trace of variance-covariance matrix for the strain. This is meant to capture overall variation. This is equivalent to the sum of the eigenvalues (Van Valen 2005). Total variance estimates were multiplied by a factor of 1000.

We also examined trait integration of wing shape using two measures derived from the within-line covariance matrix. Specifically we used matrix eccentricity as well as the standard deviation of its eigenvalues (scaled), (Van Valen 1974; Jones et al. 2003; Kirkpatrick 2009; Pavlicev et al. 2009; Haber 2011). The

standard deviation of eigenvalues of the covariance matrix has been used extensively as a proxy for integration (Cheverud et al. 1983; Pavlicev et al. 2009; Haber 2011). We calculated the relative standard deviation of eigenvalues (rSDE) and the rSDE scaled by total variance (rSDE2) (Van Valen 1974; Pavlicev et al. 2009; Haber 2011). rSDE estimates were multiplied by a factor of 10,000 and rSDE2 estimates were multiplied by a factor of 10. The shape of the VCV matrix can also be quantified using matrix eccentricity. Although typically defined as the ratio between the first two eigenvalues (Jones et al. 2003; Kirkpatrick 2009), we used a generalization that was the ratio between the largest eigenvalue and the total variance (Haber and Dworkin 2017), which has been shown to be proportional to rSDE2. Variation due to sex and size was modeled out prior to estimating total variance, eccentricity, and rSDE for each strain.

The registration process (Procrustes superimposition) influences covariation within and among landmarks. As such, the use of Procrustes residuals for 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.

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 CV for line. Alternatively, cell density CV was calculated 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 article. Associations between the measures calculated above across strains was performed using a Pearson correlation coefficient.

We 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 linear mixed models were fit, sex (excluded for macroenvironmental experiment), population, temperature (for macroenvironmental 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 were taken for left and right wing shape for 77 individuals. Wing size measurement error was estimated using an analysis of variance (ANOVA) and wing shape measurement error 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 (DA), the individual effect represents variation among individuals and the side:individual interaction term represents fluctuating asymmetry The residual variance in this model estimates measurement error (Palmer and Strobeck 1986; Palmer 1994; Klingenberg and McIntyre 1998; Palmer and Strobeck 2003; Debat et al. 2009). We compared the variation of the side:individual interaction term with the residual variation to determine whether the measurement error was negligible with respect to fluctuating asymmetry (Tables S29 and S30).

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

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. Random effects of the intercept, sex, and temperature were allowed to vary according to line nested within population. A Gamma distribution and an inverse link function for the response were used. The FA component for shape was extracted for each specimen using the *bilat.symmetry* function from *geomorph* to remove DA. As a confirmation of this analysis, wing shape FA was calculated as the Procrustes distance (PD) between the left and right wing for each individual PD_{RL}. We fit a generalized linear mixed 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 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

We first confirmed differences in trait means across populations. Consistent with previous findings (Pitchers et al. 2013; Fabian et al. 2015; Lack et al. 2016a), the high-altitude population has substantially larger wing size compared to the low-altitude population (Fig. 1A and Table S4). Wing shape also varies in a manner consistent with previous studies (PD of 0.013 between populations) (Pitchers et al. 2013) shown in Figure 1C and Table S5.

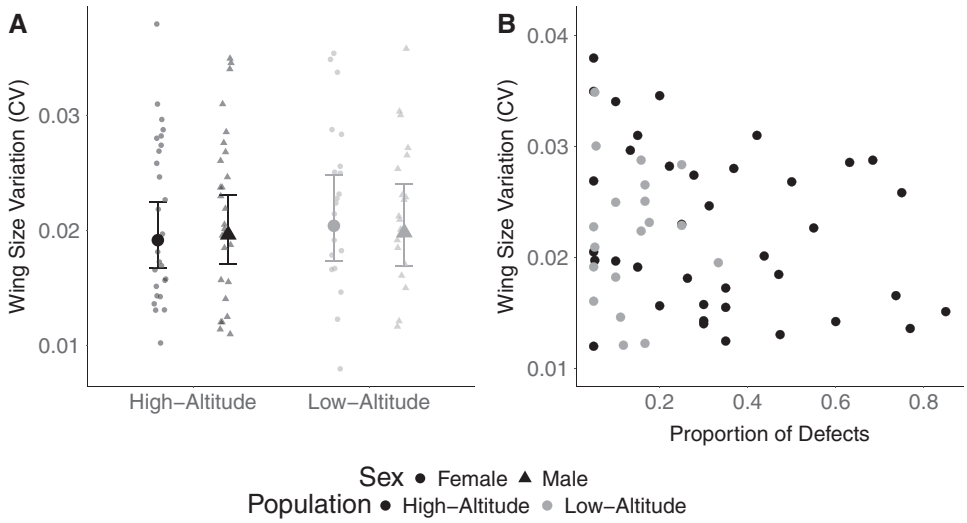

Figure 2. Within-line, among-individual variation for wing size is similar across high- and low-altitude populations. (A) Within-line coefficient of variation for wing size is similar in high- and low-altitude populations. Large symbols represent fitted values, small symbols represent coefficient of variation by line. Error bars are 95% CI (B); within-line variation for wing size is not correlated with proportion of defects in both 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).

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 (Fig. S1 and Table S8). On average, 26.7% (95% 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.

MICROENVIRONMENTAL VARIATION FOR WING SIZE IS SIMILAR BETWEEN HIGH- AND LOW-ALTITUDE POPULATIONS

Although 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 CV and Levene's deviates to measure among-individual, within-line variation. CV was plotted for ease of interpretation, but all statistical analyses were performed using Levene's deviates (Dworkin 2005b; Van Valen 2005). 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]) (Fig. S3B). As shown in Figures 2A and S3A, and Tables S6 and S7 , measures of among-individual, within-strain variability are similar between the high- 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) (Figs. 2B, S4A), although confidence intervals included zero for both populations. Similar results were observed when comparing within-line Levene's deviates with proportion of defects (Fig. S4A).

MICROENVIRONMENTAL VARIATION FOR WING SHAPE IS SIMILAR BETWEEN HIGH- AND LOW-ALTITUDE POPULATIONS

Strains derived from high- and low-altitude populations have similar levels of wing shape variation measured as the total variance (matrix size). This is also true for measures of integration (Fig. 3 and Tables S9– S12). We observed this using both rSDE and eccentricity of the covariance matrices (Fig. S6). Similar to the patterns for wing shape among populations, there is little evidence that total variance, eccentricity, rSDE and rSDE2 are correlated with frequency of wing defects (Figs. S2, S5) in either the high-altitude (total variance: $r = 0.16$, CI −0.26–0.53; eccentricity: $r = -0.24$, CI −0.59–0.18; rSDE: $r = 7.6 \times 10^{-3}$, 95% CI −0.40–0.41; rSDE2: $r = -0.29$, 95% CI −0.62–0.12) or low-altitude populations (total variance: $r = 0.34$, CI −0.24 to 0.74; eccentricity: $r = -0.33$, CI −0.73–0.24; rSDE: $r = 0.12$, 95% CI −0.44–0.61; rSDE2: $r = -0.17$, 95% CI −0.65–0.39).

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 (Fig. 4B)

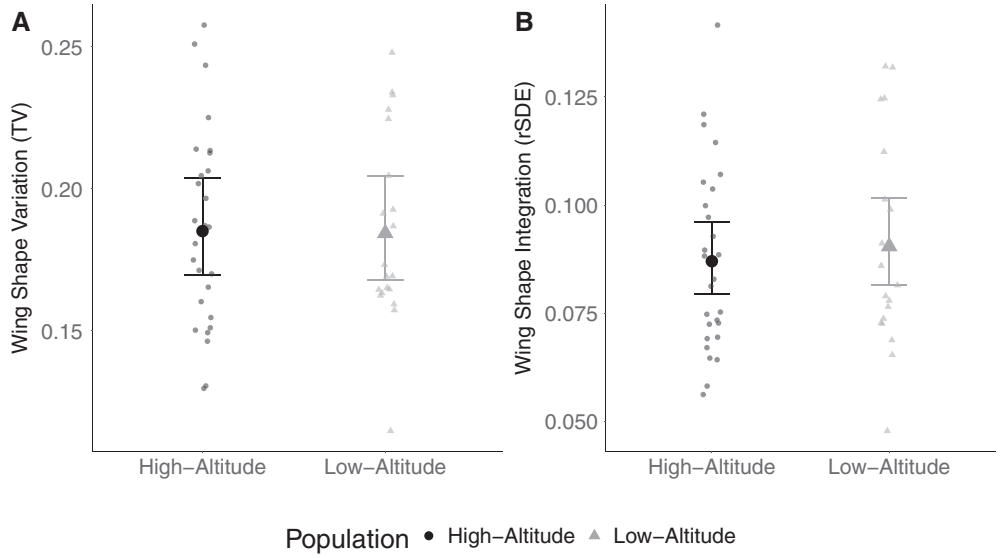

Figure 3. Similar patterns of within-line measures of variability for wing shape between high- and low-altitude populations using (A) a measure of wing shape variation, total variance of the VCV matrix (values multiplied by 1000), and (B) measure of wing shape integration using rSDE of the VCV matrix (rSDE; values multiplied by 10000). Error bars are 95% CIs.

(Fabian et al. 2015; Lack et al. 2016b) as well as for females versus males (Dobzhansky 1929; Alpatov 1930). 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 7.9 μm to greater than 11.3 μm in diameter (González-Gaitán et al. 1994). To account for local effects, we measured cell density in 16 regions across the wing (Figs. 4A and S7). Although 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 (Fabian et al. 2015; Lack et al. 2016b) remains (Fig. S7 and Table S13).

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 variation for wing size and shape also show the greatest variation for cell density within and between individuals? We calculated the within-line CV for cell density across the wing 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 (Fig. 4C and Table S14). Similarly, we did not observe differences in within-line cell density Levene's deviates between the high- and low-altitude populations (Fig. S8A and Table S15).

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) (Fig. 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 to 0.59; low-altitude, $r = -0.30$, CIs -0.60–0.063) (Fig. S8C and 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) (Fig. S8B).

TEMPERATURE-INDUCED PLASTICITY

To assess whether patterns of phenotypic plasticity varied among populations (macroenvironmental canalization) we reared strains derived 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

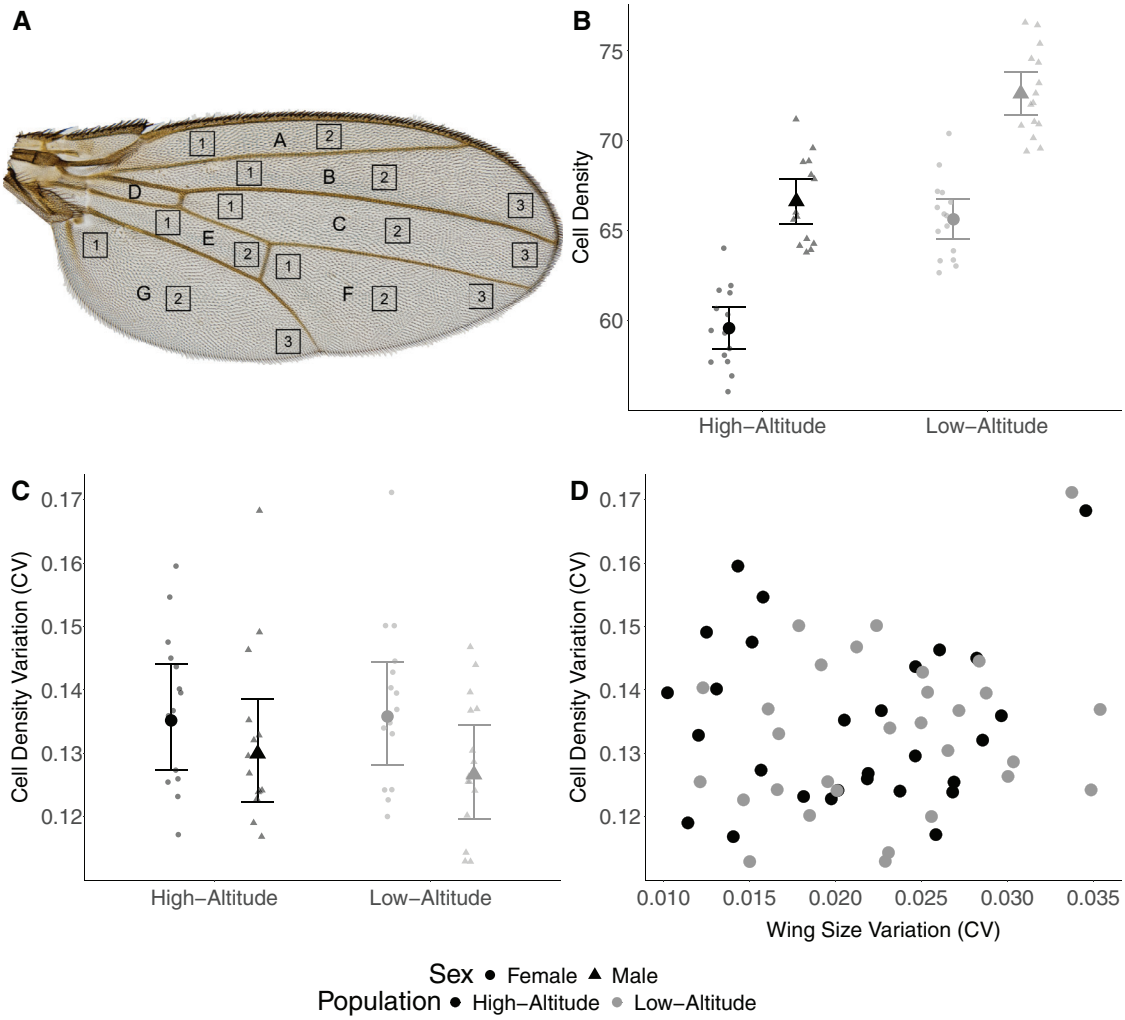

Figure 4. **Mean cell density show population and sex differences, but no differences in variability. (A) Wing regions used for cell density measurement. Squares represent a 0.0065 mm^2 measurement area. (B) Cell density varies between the sexes and between the high- and low-altitude populations. Error bars are 95% CIs. (C) Among-individual, within-line variation measured as CV for cell density is similar between high- and low-altitude populations. (D) Within-line cell density variation and within-line wing size variation (CV) are not strongly correlated in either the high-altitude population ($r = 0.076$ CI $-0.31 - 0.44$), and the low-altitude population ($r = 0.25$, CI $-0.12 - 0.56$).**

evidence of nonlinearity (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 (Fig. 5A and Tables S16, and S17). We observed an increase in the procrustes distance between mean shapes of the high- and low-altitude populations as temperature increases (Fig. S11).

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 (Fig. S9A and Table S24).

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°C and 28°C for high-altitude females and an increase in CV at 18°C but not at 28°C for high-altitude males. Within-line CV for low-altitude males and females is consistent across temperatures (Fig. 5B and Table S18). We observed a similar pattern when using Levene's deviates as we did for CV (Fig. S9B and Table S18), 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 (Fig. 5C and Table S20). Degree of integration of wing shape is similar across populations

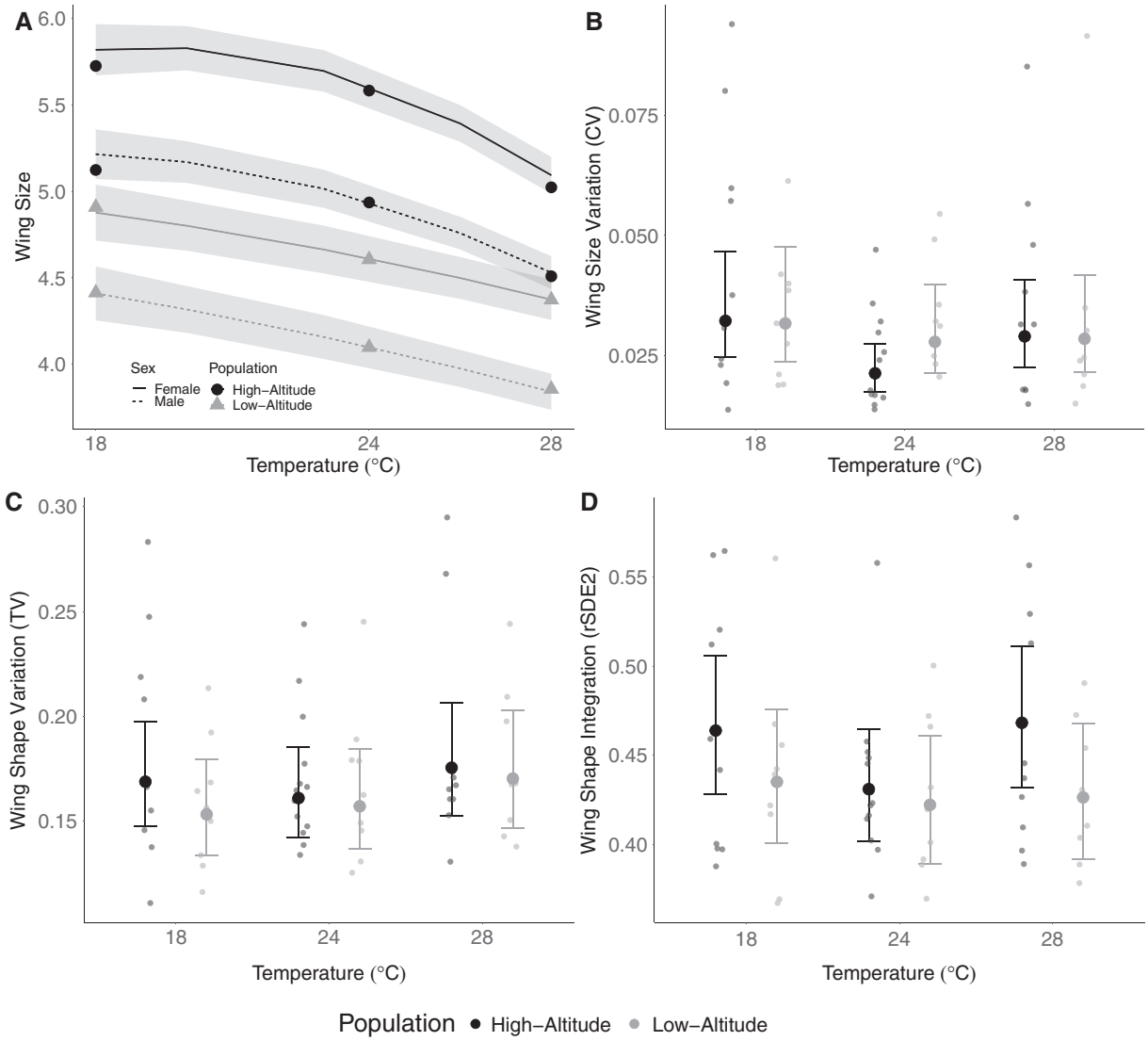

Figure 5. Considerable plasticity for mean wing size under different temperature rearing environments, but with minimal differences in trait variation (A) Mean differences observed in wing size (centroid size) at different temperatures in males and females in the high- and low-altitude populations. (B) Within-line variation for wing size is similar across sexes and populations within each temperature treatment. (C) Within-line variation for wing shape measured as the total variance of the VCV matrix (multiplied by a factor of 1000). (D) Within-line wing shape integration measured as rSDE2 (rSDE scaled by the total variance) are similar between the high- and low-altitude populations across the different rearing temperatures. Gray shading and error bars are 95% CIs.

and temperatures (Fig. 5D and Tables S21– S23). This pattern holds whether examining rSDE or eccentricity of the covariance matrix (Fig. S9C and D).

FLUCTUATING ASYMMETRY FOR WING SIZE AND SHAPE IS SIMILAR BETWEEN HIGH- AND LOW-ALTITUDE POPULATIONS

Although 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 measures do not always agree, with correlations ranging from ∼0.07 to 0.6 for wing size and ∼0.35 to 0.48 for shape (Breuker et al. 2006; Debat et al. 2006, 2009). Thus, we measured asymmetry for both wing size and shape in a subset of high- and low-altitude lines (two 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 (Table S29) although had a larger impact on shape (Table S30). Using FA8 as an index for wing size FA, we compared developmental instability for high- and low-altitude populations at three temperatures (18 °C, 24 °C, and 28 °C). We did not see clear evidence of differences in FA8 between high- and low-altitude populations or across the different rearing temperatures, except for consistently lower FA8 in the high-altitude

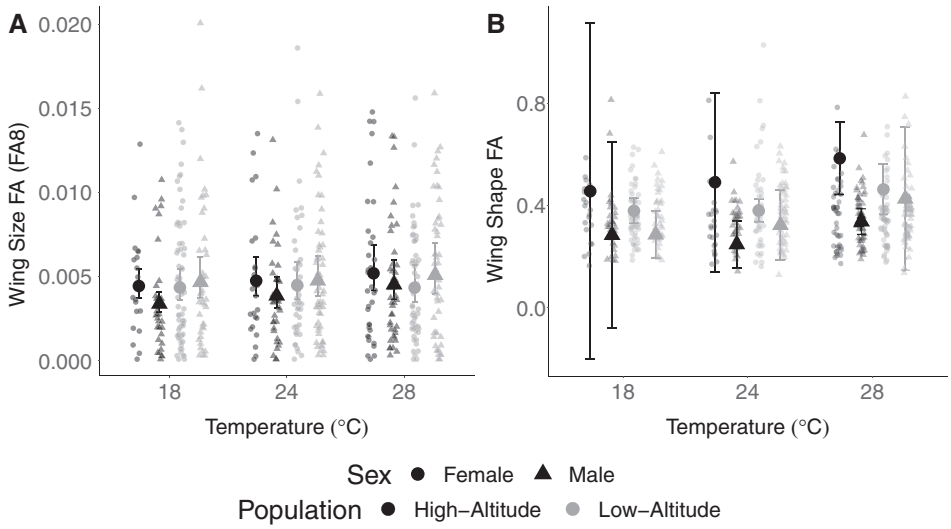

Figure 6. Inconsistent differences in fluctuating asymmetry among the high- and low-altitude populations for (A) wing size fluctuating asymmetry measured as *FA*8 and (B) wing shape fluctuating asymmetry between the high- and low-altitude populations across temperatures. High-altitude males have consistently lower fluctuating asymmetry and high-altitude females have consistently greater wing shape fluctuating asymmetry across temperatures. Large symbols represent population means, small symbols represent individuals, error bars represent 95% CIs.

males at all temperatures (Fig. 6A and Table S26). For comparison we also report FA1 as a measure of developmental instability (Fig. S12 and Table S25), 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 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 (Fig. 6B and Table S28). We also calculated PD between left and right wings (PD_{LR}). We observed a similar pattern as we did using the first method to calculate wing shape FA, where there is consistently greater PD_{LR} in the high-altitude females across temperatures (Fig. S12B and Table S27).

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 micro-RNAs *in vivo* and *in silico*, which may not necessarily representative of multicellular organisms (Szöllősi and Derényi 2009). Studies in other systems have not provided evidence to support the congruence hypothesis (Dworkin 2005a, 2005c). However, most empirical studies were conducted using lab domesticated lineages, lab-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 and shape (Pitchers et al. 2013; Lack et al. 2016a, 2016b) (Figs. 1, 5A, S11), cell size (Figs. 4, S7), and frequencies of mutational defects (Figs. 2B, S1), we did not observe any evidence for associations between genetic and microenvironmental canalization (Figs. 2, 3, 4C and D, 5B and D). Additionally, measures of among-individual, within-line variance for wing size and shape were not correlated with the proportion of mutational defects both in the high- and low-altitude populations (Figs. 2B, 4D, S2, S4, S8B, S10). We did observe greater temperature induced plasticity of mean wing size in the high-altitude population (Fig. 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 (Figs. 5B, S9B), 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 (Fig. 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 variation were not correlated with the proportion of defects at any of the developmental temperatures (S10). 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 approximately 2 years prior to the first experiment and 7 years prior to the temperature manipulation experiment. As such, both drift (due to the initial inbreeding process) and some degree of lab adaptation may have occurred. As we used multiple strains from each population, the impact of drift with respect to allele frequencies should be modest. Additionally lab domestication should be weak as N_e is extremely small within each strain. Although the two populations have modest genetic differentiation ($F_{ST} = 0.15$), lines derived from African populations tend to have high residual heterozygosity even after 8 generations of inbreeding (Lack et al. 2016a). However, given that our results for mean size and shape of the wings and cell densities recapitulate previous findings, the impact of both drift and lab domestication appear to be minor.

Although 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 derived from high altitude, and the increased sensitivity under mutagenesis may reflect one aspect of genetic canalization (i.e., 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 is complex.

Although the work of Lack et al. (2016a, 2016b) 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 size 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 armor plates and 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-altitude population, as has occurred with the modifier variants influencing asymmetry in the blowflies (Davies et al. 1996). Indeed it is not yet clear whether wing size is near its optima for the high-altitude population, and whether that is necessary for the evolution of canalization.

Based on what is known about the genetic architecture of body size, wing size and wing shape the fact that genetic decanalization occurred is surprising. The mutational target size of body size, wing size, and wing shape are quite large (Weber et al. 2005; Carreira et al. 2008, 2011; Houle and Fierst 2012). Similarly, these traits harbor extensive standing genetic variation in populations and are polygenic in nature (Weber 1990a, 1990b; 1999; Mezey et al. 2005; Mezey and Houle 2005). As such the expected modest changes in individual allele frequencies would seem to be unlikely to result in changes in canalization. Yet, that is exactly what has been observed in the high-altitude population which increased its wing (and body) size, and also from the low-altitude population artificially selected for large wing size (Lack et al. 2016a; Groth et al. 2018). Body size and wing size have been a frequent subject of study in *Drosophila*, but until now this pattern has not been previously observed. As such future work both examining additional populations that vary for size and also on identifying 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? Although the results from the current study, and some previous studies (Dworkin 2005a, 2005c; 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. Under this model, if deleterious alleles are purged efficiently enough, adaptive genetic canalization will not have the opportunity to evolve. However, this does not consider that deleterious alleles are not always purged efficiently for numerous reasons (pleiotropy, linkage with beneficial mutations, drift, G $\times$ E, 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 (i.e., 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 (Hermisson et al. 2003; Proulx and Phillips 2005). However, our knowledge of the distribution of this fitness load and the frequency of relevant environmental perturbations is limited. Although 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; Siegal and Bergman 2002; Visser et al. 2003; 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., McKenzie and Clarke 1988; Morris et al. 2019). 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.

AUTHOR CONTRIBUTIONS

Experimental design: MP and ID; ran experiments: MP; analysis: MP and ID; writing and editing: MP and ID.

ACKNOWLEDGMENTS

We wish to thank Dr. John Pool and Dr. Justin Lack for sending strains. We 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. Dr. Michael Collyer and 2 anonymous reviewers provided helpful feedback that improved the clarity of this manuscript. This work was funded by an NSERC (Canada) Discovery and Discovery accelerator grant to ID.

DATA ARCHIVING

Data and scripts are available on DRYAD (https://doi.org/10.5061/dryad.b8gtht79c) and github (https://github.com/DworkinLab/PesevskiDworkin_Evolution_2020). Raw images for the first series of experiments are available on figshare (https://doi.org/10.6084/m9.figshare.1449060).

LITERATURE CITED

Adams, D., M. Collyer, and A. Kaliontzopoulou. 2018. Geomorph: software for geometric morphometric analyses. R package version 3.0.6. Available at https://cran.r-project.org/package=geomorph.

Agrawal, A. F., and J. R. Stinchcombe. 2009. How much do genetic covariances alter the rate of adaptation? Proc. R. Soc. B 276:1183–1191.

Alpatov, W. W., 1930. Phenotypical variation in body and cell size of *Drosophila melanogaster*. Biol. Bull. 58:85–103.

Ancel, L. W., and W. Fontana. 2000. Plasticity, evolvability, and modularity in RNA. J. Exp. Zool. 288:242–283.

Arnold, S. 1983. Morphology, performance and fitness. Am. Zool. 23:347–361.

———. 2003. Performance surfaces and adaptive landscapes. Integr.Comp. Biol. 43:367–375.

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

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using *lme4*. J. Stat. Softw. 67:1–48.

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

Borenstein, E., and E. Ruppin. 2006. Direct evolution of genetic robustness in microRNA. Proc. Natl. Acad. Sci. 103:6593–6598.

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

Brooks, M. E., K. Kristensen, K. J. van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. J. Skaug, M. Maechler, and B. M. Bolker. 2017. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R. J. 9:378–400. https://journal.r-project.org/archive/2017/RJ-2017-066/index.html.

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

Carreira, V. P., J. Mensch, and J. J. Fanara. 2008. Body size in Drosophila: genetic architecture, allometries and sexual dimorphism. Heredity 102:246–256.

Carreira, V. P., I. M. Soto, J. Mensch, and J. J. Fanara. 2011. Genetic basis of wing morphogenesis in Drosophila: sexual dimorphism and non-allometric effects of shape variation. BMC Dev. Biol. 11:32.

Chandler, C. H., S. Chari, A. Kowalski, L. Choi, D. Tack, M. DeNieu, W. Pitchers, A. Sonnenschein, L. Marvin, K. Hummel et al. 2017. How well do you know your mutation? Complex effects of genetic background on expressivity, complementation, and ordering of allelic effects. PLoS Genet. 13:e1007075.

Chari, S., and I. Dworkin. 2013. The conditional nature of genetic interactions: the consequences of wild-type backgrounds on mutational interactions in a genome-wide modifier screen. PLoS Genet. 9:e1003661. http://plos.org/10.1371/journal.pgen.1003661.

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

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

Davies, A. G., A. Y. Game, Z. Chen, T. J. Williams, S. Goodall, J. L. Yen, J. A. McKenzie, and P. Batterham. 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., C. C. Milton, S. Rutherford, C. P. Klingenberg, and A. A. Hoffmann. 2006. Hsp90 and the quantitative variation of wing shape in *Drosophila melanogaster*. Evolution 60:2529–2538.

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

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

Dillon, M. E., M. R. Frazier, and R. Dudley. 2006. Into thin air: physiology and evolution of alpine insects. Integr. Comp. Biol. 46:49–61.

Dittmar, E. L., C. G. Oakley, J. K. Conner, B. A. Gould, and D. W. Schemske, 2016. Factors influencing the effect size distribution of adaptive substitutions. Proc. R. Soc. B 283:20153065.

Dobens, A. C., and L. L. Dobens. 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' Arch. Entwicklungsmechanik Organ. 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.

———. 2005b. Canalization, cryptic variation, and developmental buffering: a critical examination and analytical perspective. Pp. 131–158 *in* B. Hallgrimsson and B. K. Hall, eds. Variation: a central concept in biology. Elsevier, Amsterdam. https://www.sciencedirect.com/science/article/pii/B9780120887774500107.

———. 2005c. Evidence for canalization of distal-less function in the leg of *Drosophila melanogaster*. Evol. Dev. 7:89–100.

Eshel, I., and C. Matessi. 1998. Canalization, genetic assimilation and preadaptation: a quantitative genetic model. Genetics 149:2119–2133.

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

Félix, M.-A., and M. Barkoulas. 2015. Pervasive robustness in biological systems. Nat. Rev. Genet. 16:483–496.

Flatt, T. 2005. The evolutionary genetics of canalization. Q. Rev. Biol. 80:287–316.

Geiler-Samerotte, K., F. M. Sartori, and M. L. Siegal. 2019. Decanalizing thinking on genetic canalization. Semin. Cell Dev. Biol. 88:54–66.

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

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

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

Green, R. M., J. L. Fish, N. M. Young, F. J. Smith, B. Roberts, K. Dolan, I. Choi, C. L. Leach, P. Gordon, J. M. Cheverud et al. 2017. Developmental nonlinearity drives phenotypic robustness. Nat. Commun. 8: 1970.

Groth, B. R., Y. Huang, M. J. Monette, and J. E. Pool. 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. Evol. Biol. 38:476–488.

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

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

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

Hansen, T. F., and D. Houle. 2008. Measuring and comparing evolvability and constraint in multivariate characters. J. Evol. Biol. 21:1201–1219.

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

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

Houle, D., and J. Fierst. 2012. Properties of spontaneous mutational variance and covariance for wing size and shape in *Drosophila melanogaster*. Evolution 67:1116–1130.

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

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

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

Jones, A. G., S. J. Arnold, and R. Bürger. 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., M. Gáliková, C. D. Huber, and T. Flatt. 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. Front. Ecol. Evol. 7:56.

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

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

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

Lack, J. B., A. Yassin, Q. D. Sprengelmeyer, E. J. Johanning, J. R. David, and J. E. Pool, 2016b. Life history evolution and cellular mechanisms associated with increased size in high-altitude *Drosophila*. Ecol. Evol. 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 S. Arnold. 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 M. L. Siegal. 2008. Network hubs buffer environmental variation in *Saccharomyces cerevisiae*. PLoS Biol. 6:e264.

Mayer, C., and T. F. Hansen. 2017. Evolvability and robustness: a paradox restored. J. Theor. Biol. 430:78–85.

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

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

Mezey, J. G., and D. Houle. 2005. The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. Evolution 59:1027–1038.

Mezey, J. G., D. Houle, and S. V. Nuzhdin. 2005. Naturally segregating quantitative trait loci affecting wing shape of *Drosophila melanogaster*. Genetics 169:2101–2113.

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

Orgogozo, V., B. Morizot, and A. Martin. 2015. The differential view of genotype-phenotype relationships. Front. Genet. 6:179.

Paaby, A. B., and M. V. Rockman. 2014. Cryptic genetic variation: evolution's hidden substrate. Nat. Rev. Genet. 15:247–258.

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

Palmer, A. R. 1994. Fluctuating asymmetry analyses: a primer. Pp. 335–364 *in* T. Markow, ed. Developmental instability: its origins and evolutionary implications. Kluwer, Dordrecht.

———. 1986. Fluctuating asymmetry: measurement, analysis, patterns. Annu. Rev. Ecol. Evol. Syst. 17:391–421.

Palmer, A. R., and C. Strobeck. 2003. Fluctuating asymmetry analyses revisited. Pp. 279–319 *in* M. Polak, ed. Developmental instability: causes and consequences. Oxford Univ. Press, Oxford.

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

Pavlicev, M., J. M. Cheverud, and G. P. Wagner. 2009. Measuring morphological integration using eigenvalue variance. Evol. Biol. 36:157–170.

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

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

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

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

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

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

Pool, J. E., D. T. Braun, and J. B. Lack. 2016. Parallel evolution of cold tolerance within *Drosophila melanogaster*. Mol. Biol. Evol. 34: 349–360.

Proulx, S., and P. Phillips. 2005. The opportunity for canalization and the evolution of genetic networks. Am. Nat. 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. Pp. 4–16 *in* Proceedings—National Poultry Breeders.

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. Annu. Rev. Ecol. Evol. Syst. 22:65–93.

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

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

Siegal, M. L., and A. Bergman. 2002. Waddington's canalization revisited: developmental stability and evolution. Proc. Natl. Acad. Sci. 99:10528–10532.

Siegal, M. L., and J.-Y. Leu. 2014. On the nature and evolutionary impact of phenotypic robustness mechanisms. Annu. Rev. Ecol. Evol. Syst. 45:496–517.

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

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

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

Szöllősi, G. J., and I. Derényi. 2009. Congruent evolution of genetic and environmental robustness in micro-RNA. Mol. Biol. Evol. 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., L. Rako, T. Takano-Shimizu, A. A. Hoffmann, and S. F. Lee, 2010. Effects of small Hsp genes on developmental stability and microenvironmental canalization. BMC Evol. Biol. 10:284.

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

Van Valen, L. 1974. Multivariate structural statistics in natural history. J. Theor. Biol. 45:235–247.

———. 2005. The statistics of variation. Pp. 29–47, *in* B. Hallgrimsson and B. K. Hall, eds. Variation: a central concept in biology. Elsevier Academic Press, New York. https://www.sciencedirect.com/science/article/pii/B9780120887774500053.

Visser, J., J. Hermisson, G. Wagner, L. Meyers, H. Bagheri-Chaichian, J. Blanchard, L. Chao, J. Cheverud, S. Elena, W. Fontana et al. 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., G. Booth, and H. Bagheri-Chaichian. 1997. A population genetic theory of canalization. Evolution 51:329–347.

Walsh, B., and M. W. Blows. 2009. Abundant genetic variation plus strong selection—multivariate genetic constraints: a geometric view of adaptation. Annu. Rev. Ecol. Evol. Syst. 40:41–59.

Weber, K. E. 1990a. Increased selection response in larger populations. I. Selection for wing-tip height in *Drosophila melanogaster* at three population sizes. Genetics 125:579–584. https://www.genetics.org/content/125/3/579.short.

———. 1990b. Selection on wing allometry in *Drosophila melanogaster*. Genetics 126:975–989. https://www.genetics.org/content/126/4/975.short.

Weber, K., R. Eisman, L. Morey, A. Patty, J. Sparks, M. Tausek, and Z. B. Zeng, 1999. An analysis of polygenes affecting wing shape on chromosome 3 in *Drosophila melanogaster*. Genetics 153:773–786. https://www.genetics.org/content/153/2/773.short.

Weber, K., N. Johnson, D. Champlin, and A. Patty. 2005. Many P-element insertions affect wing shape in *Drosophila melanogaster*. Genetics 169:1461–1475. https://www.genetics.org/content/169/3/1461.short.

Associate Editor: M. Collyer
Handling Editor: T. Chapman

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. High-altitude population has a higher frequency of wing defects compared to the low-altitude population.

Figure S2. Within-line measures of variability for wing shape are not correlated with proportion of wing defects (TV - Total variance: r = 0.16 95% CI −0.26 - 0.53;Ecc - 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. Within-line wing size variation measured as Levene's Deviates.

Figure S4. Within-line variation for wing size measured as 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. Within-line measures of variability for wing shape measured as wing shape integration using the relative standard deviation of the eigenvalues (rSDE - multiplied by 10000 and rSDE2) are not correlated with proportion of wing defects (rSDE: r = 7.6^{-3} 95% CI −0.40 - 0.41; rSDE2: r = −0.29 95% CI −0.62 - 0.12) or the low-altitude population (rSDE: r = 0.12 95% CI −0.44 - 0.61; rSDE2: r = −0.17 95% CI −0.65 - 0.39).

Figure S6. Within-line variation of wing shape measured as wing shape integration using matrix eccentricity and the relative standard deviation (rSDE2) of the eigenvalues of VCV matrix are similar in the high- and low-altitude populations.

Figure S7. Mean cell density across the 16 measurement wing regions.

Figure S8. Within-line variation for cell density measured as Levene's Deviates and association between within-line variation for cell density with within-line wing defects, and within-line variation for wing shape.

Figure S9. Proportion of wing defects and alternative measures of within-line variation for wing size and shape for the high- and low-altitude population at different developmental temperatures.

Figure S10. Little evidence for correlation between measures of within-line variation 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 S11. Mean wing shape differences between the high- and low-altitude populations at different temperatures (females).

Figure S12. Fluctuating Asymmetry for wing size and shape represented by (A) FA1 and (B) Procrustes distance between the left and right wing (PD_{LR}).

Figure S13. Fluctuating asymmetry for wing size and shape vs. measures of variability for wing size.

Figure S14. Fluctuating asymmetry for wing size and shape vs. measures of variability for wing shape.

Table S1. Sample sizes of fly strains for microenvironmental canalization.

Table S2. Sample sizes of fly strains for Temperature Plasticity and Macroenvironmental canalization.

Table S3. Recipe for 1.5:1 Protein:Sugar Food used in Temperature Plasticity/Macroenvironmental canalization and FA experiments.

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

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

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.

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.

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.

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

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

Table S11. Results from generalized mixed effects model testing the effects of population on within-line eccentricity for wing shape.

Table S12. Results from generalized mixed effects model testing the effects of population on within-line rSDE2 for wing shape.

Table S13. 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.

Table S14. Results from generalized mixed effects model testing the effects of population and sex and their interaction on within-line among-individual cell density CV.

Table S15. 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.
Table S16. Results from linear mixed effects model testing the effects of temperature, sex, population and their interactions on wing size.
Table S17. Results from Multivariate Procrustes ANOVA testing the effects of wing size, sex, temperature, population and all interactions on wing shape.
Table S18. Results from generalized mixed effects model testing the effects population, temperature and their interaction on within-line among-individual CV for wing size.
Table S19. 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.
Table S20. 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.
Table S21. Results from generalized mixed effects model testing the effects of population, temperature and their interaction on within-line among-individual eccentricity for wing shape.
Table S22. Results from generalized mixed effects model testing the effects of population, temperature and their interaction on within-line among-individual rSDE for wing shape.
Table S23. Results from generalized mixed effects model testing the effects of population, temperature and their interaction on within-line among-individual rSDE2 for wing shape.
Table S24. Results from generalized mixed effects model testing the effects of temperature, population, sex and all interactions on within-line proportion of wing defects.
Table S25. Results from generalized mixed effects model testing the effects of sex, population, temperature and all interactions on FA1.
Table S26. Results from generalized mixed effects model testing the effects of sex, population, temperature and all interactions on FA8.
Table S27. Results from generalized mixed effects model testing the effects of sex, population, temperature and all interactions on PD_{LR}.
Table S28. Results from generalized model testing the effects of wing size, sex, population, temperature and all interactions on wing shape FA after correcting for D.
Table S29. Results from wing size ANOVA on repeated measurements using side, individual and their interaction as effects to estimate measurement error.
Table S30. Results from wing shape multivariate procrustes ANOVA on repeated measurements using side, individual and their interaction as effects to estimate Wing shape Measurement Error ANOVA.