doi:10.1111/j.1558-5646.2010.01075.x

# NO EFFECT OF ENVIRONMENTAL HETEROGENEITY ON THE MAINTENANCE OF GENETIC VARIATION IN WING SHAPE IN *DROSOPHILA MELANOGASTER*

**Sam Yeaman,[1,2] Yukon Chen,[1] and Michael C. Whitlock[1]**

[1]*Department of Zoology, University of British Columbia, 6270 University Blvd., Vancouver, British Columbia V6T 1Z4, Canada*

[2]*E-mail: yeaman@zoology.ubc.ca*

Received December 11, 2009
Accepted June 18, 2010

**Theory suggests that heterogeneous environments should maintain more genetic variation within populations than homogeneous environments, yet experimental evidence for this effect in quantitative traits has been inconsistent. To examine the effect of heterogeneity on quantitative genetic variation, we maintained replicate populations of *Drosophila melanogaster* under treatments with constant temperatures, temporally variable temperature, or spatially variable temperature with either panmictic or limited migration. Despite observing differences in fitness and divergence in several wing traits between the environments, we did not find any differences in the additive genetic variance for any wing traits among any of the treatments. Although we found an effect of gene flow constraining adaptive divergence between cages in the limited migration treatment, it did not tend to increase within-population genetic variance relative to any of the other treatments. The lack of any clear and repeatable patterns of response to heterogeneous versus homogeneous environments across several empirical studies suggests that a single general mechanism for the maintenance of standing genetic variation is unlikely; rather, the relative importance of putative mechanisms likely varies considerably from one trait and ecological context to another.**

**KEY WORDS: Environmental heterogeneity, fluctuating allelic effects, gene flow, genetic diversity, laboratory natural selection, maintenance of variation.**

In 1981, Mackay published the results of a landmark experiment examining the effect of heterogeneous environments on the maintenance of quantitative genetic variation in bristle traits and body size. Interestingly, she found that both long- and short-period temporally heterogeneous environments (ethanol vs. standard fly medium) maintained more additive genetic variance ($V_A$) in sternopleural bristles than spatially varying environments, which is not predicted from classical single-locus theory (Felsenstein 1976). While her study presented strong evidence for an effect of heterogeneity on variation, finding up to 3.5 times more $V_A$ in the heterogeneous treatments than the homogenous control, other similar studies have yielded inconsistent results. Of two studies with *Drosophila* spp. that preceded Mackay's, one found either an increase in genetic variance for sternopleural bristles of ∼50% under diurnal temperature fluctuations (Beardmore 1961), while the other found a small and nonsignificant relationship between $V_G$ and heterogeneity on a 1-month or 3-month cycle (Long 1970). A third study that varied the type of medium given to successive generations of *Tribolium castaneum* found no increase in genetic variance relative to constant controls (Riddle et al. 1986). Two subsequent studies that inferred genetic variation from heritability or response to selection also did not find

consistent trends (Verdonck 1987; Garcia-Dorado et al. 1991). By contrast, studies of the maintenance of allozyme heterozygosity found comparatively consistent evidence of modest increases (usually under 50%) under heterogeneous environments (Powell 1971; McDonald and Ayala 1974; Powell and Wistrand 1978; but see Haley and Birley 1983), which may be due to the relatively less stringent conditions for maintenance of polymorphism in single-locus traits (Spichtig and Kawecki 2004). Given the variation in the results of these studies and the range of study designs they have employed (often with limited replication), it is unclear how broadly Mackay's results would generalize to other traits and environmental stresses.

Over a quarter of a century after Mackay's paper, we are still uncertain about the relative importance of the various evolutionary processes for the maintenance of genetic variation within populations. Predictions from models of mutation-selection balance are still thought to be inconsistent with empirical measurements of the strength of selection and rate of mutation (Johnson and Barton 2005; but see Zhang et al. 2004; Zhang and Hill 2005), and while a range of models have shown that heterogeneous environments have the potential to maintain more variation within populations (Via and Lande 1987; Gillespie and Turelli 1989; Bürger and Gimelfarb 2002; Spichtig and Kawecki 2004; Turelli and Barton 2004), there is little formal quantification of the likely magnitude of this effect under biologically reasonable parameters. On the other hand, there is considerable evidence that environmental heterogeneity often maintains variance between populations by promoting local adaptation (e.g., Hedrick et al. 1976; Linhart and Grant 1996; Mitchell-Olds et al. 2007). While there is some evidence that gene flow between populations in heterogeneous environments can cause a detectable increase in genetic variance within populations (Yeaman and Jarvis 2006), the generality of this effect is unclear. Given the difficulties in parameterizing models of mutation-selection balance in quantitative traits (Johnson and Barton 2005), further experimental studies may yield the most productive evidence to illuminate our understanding of the maintenance of variation (see Charlesworth et al. 2007 for an interesting approach to this problem).

Here, we test the effect of various patterns of environmental heterogeneity on the maintenance of genetic variation in wing traits in *Drosophila melanogaster*, replicating and extending Mackay's classic study in a different trait and environment, and including a treatment examining the effect of spatial heterogeneity with limited migration (treatments described in Table 1). As differences in temperature are known to cause adaptive responses in *Drosophila* for wing size (Partridge et al. 1994) and shape (Santos et al. 2004), we established 25 replicate populations from a sample of wild-collected *D. melanogaster* and exposed each of them to one of five different heterogeneous or homogeneous temperature treatments for 116 weeks (five replicate populations per treatment with two cages per population). To simulate a spatially heterogeneous environment, the two cages of each replicate were maintained at different temperatures (16°C and 25°C) with either panmictic (S; $m = 0.5$) or limited migration (M; $m \sim 0.001$). The migration in this experiment was forced, meaning that the flies did not have the opportunity to choose a preferred habitat. To simulate a temporally heterogeneous environment, both cages were maintained together at the same temperature, but were moved every 4 weeks from one of the experimental temperatures to the other (T). These experimental treatments were compared to two homogeneous treatments, where both cages of each line were maintained at either the cold or hot temperature for the duration of the experiment (C & H). At the end of the period of adaptation, we assayed genetic variance in 20 wing morphology traits at both experimental temperatures using a parent–offspring breeding design. This design allows us to test whether more variance is maintained under temporal than spatial heterogeneity, as found unexpectedly by Mackay (1981), and to compare the variance maintained by these treatments to a pattern of spatial heterogeneity with limited migration.

**Table 1. Description of experimental treatments. Each treatment was applied to five replicate populations with two cages in each replicate.**

| Code | Name | Description |
|---|---|---|
| C | Cold homogeneous | Two cages at 16°C, 8 bottles per cage, 4 bottles migrate reciprocally every 4 weeks |
| H | Hot homogeneous | Two cages at 25°C, 8 bottles per cage, 4 bottles migrate reciprocally every 4 weeks |
| S | Spatially heterogeneous with panmixia | One cage at 25°C, one at 16°C, 8 bottles per cage, 4 bottles migrate reciprocally every 4 weeks |
| T | Temporally heterogeneous | Two cages maintained together, 8 bottles per cage, 4 bottles migrate reciprocally, cages move between 25°C and 16°C every 4 weeks |
| M | Spatially heterogeneous with limited migration | One cage at 25°C (MH), one at 16°C (MC), 8 bottles per cage, 2 mated females migrate reciprocally every 4 weeks |

# Methods

## ESTABLISHMENT AND MAINTENANCE OF EXPERIMENTAL LINES

All lines used in this experiment were established from a large sample of *D. melanogaster* captured on September 25, 2005,

using a sweep net over a permanent compost pile from a certified organic orchard near Cawston in the Similkameen Valley, British Columbia, Canada. From an initial capture of ~2000 adult males and females, 400 mated females were placed individually into vials and allowed to lay eggs. Virgin females were collected from these vials and used to establish lines by making reciprocal crosses between the numbered isofemale lines. All isofemale lines were test-crossed with males from a known *Drosophila simulans* strain from the Tucson Stock Centre to confirm species identity (see Supporting Information for details), yielding 149 reciprocally crossed lines with known species identity (representing the genetic variation from 298 isofemale lines).

The population started from these remaining lines was allowed to expand in size over the next six generations, with bottles randomly reassigned among cages (see Supporting Information for culture medium and cage design). By February 26, 2006, after nine generations in the laboratory, the population had grown in size to fill 32 cages with eight bottles each (approximately 64,000 adults). To establish the experimental lines, 400 bottles were added to the cages of this population, then removed 2 days later and redistributed among 50 new cages. Two cages were then randomly assigned to each replicate population of each treatment and transferred to either 16°C or 25°C according to their experimental treatment (see Table 1).

In cages maintained at 25°C (or 16°C, in parentheses), a new generation began every 2 (4) weeks, with fresh bottles added on Wednesday evening (Tuesday morning) and transferred to new cages on Friday morning. Flies in the cold room thus had approximately twice the laying time of hot room flies to account for their approximately twofold slower metabolism. Migration occurred every 4 weeks, such that cages in the cold room migrated every generation, while cages in the hot room migrated every second generation. Migration for the "M" treatment was performed on Tuesday morning before feeding; a random sample of ~15–30 flies was aspirated from the top of each cage, lightly anaesthetized using $CO_2$, and sorted by sex. Two migrant females from each sample were retained and moved to the corresponding cage in the opposite temperature (excess flies from these samples were not returned to the cages). Migration in treatments S, T, H, and C was performed on Friday morning during the transfer of the bottles; half of the bottles from each cage were randomly selected and swapped between cages, yielding an enforced random migration rate of $m \sim 0.5$. The S and T treatments therefore reproduce a pattern of heterogeneity similar to the design of Mackay (1981) while the M treatment represents the effect of heterogeneity with limited migration. We note, however, that by enforcing random migration in the S and T treatments, our design did not permit habitat choice, which may have occurred in Mackay's experiment, where bottles with different types of medium were maintained in the same cage.

Population size at the end of each generation typically ranged from 2000 to 4000 adults per cage but on rare occasions was observed to be as low as ~800 in some cages, due to natural variability of the populations. There were consistent differences between treatments in mean population size due to differences in temperature and ecology (from census data collected over nine 2-week intervals, estimated mean number of adults per cage was: C = 2077, H = 3615, MC = 2351, MH = 2936, S = 1684/2818 (cold/hot chambers), T = 2375). We regard these differences in density as a (potentially) important part of the ecology causing trait divergence between the warm and cold chambers (e.g., Santos et al. 2004), but as genetic drift is more pronounced in smaller populations, it is also possible these differences in size could have caused changes in genetic variance for reasons other than the environmental heterogeneity of the treatments. We note that while cages in the warm room had twice as many generations for selection, mutation, and drift to act upon them as those in the cold room, we control for any bias this could cause by comparing the heterogeneous treatments to both the H and the C treatments (i.e., the treatments with the most and fewest generations). A minor mite infestation occurred midway through the experimental evolution phase of this project, but did not seem to affect the results (see Supporting Information).

## EXPERIMENTAL CROSSES

Experimental crosses to assay genetic variance in wing traits were begun in early May, 2009, following 116 weeks of the experiment, which corresponds to 29 generations in the cold room or 58 generations in the hot room. Assays were conducted under both experimental temperatures by initiating two assay lines from each experimental line and rearing one under each temperature. Each assay line was established with 10 vials of 30 randomly picked eggs taken from laying plates left overnight in both cages of each experimental line. To allow for comparisons between the two subpopulations of each replicate of the limited migration treatment (MC/MH), we treated the subpopulations from the hot and cold chambers as separate replicate lines, establishing hot and cold assay lines from each cage. This yielded a total of 30 replicate lines in each assay temperature. Because the procedure for maintaining the assay lines was labor-intensive, we staggered the establishment of assay lines by replicate number, starting one line per treatment every working day over a 1-week period. Following this initial staggering, a strict schedule was observed such that each replicate block was transferred for mating and laying at the same time relative to its initiation (see Supporting Information for details). Except where specifically stated, all activities associated with the establishment and maintenance of assay lines were conducted at the assay temperature. To control for maternal effects, we reared three generations of flies from each replicate line at controlled densities under both experimental temperatures,

using the final two generations as parents and offspring to assay genetic variation. Productivity trials were performed with the C and H treatment parents at each assay temperature by allowing them to lay for either 2 (hot assay) or 4 (cold assay) days and counting the number of offspring that emerge from each vial (see Supporting Information for details).

## WING TRAITS AND STATISTICAL ANALYSIS

Images of the left wings of flies were captured using the approach of Houle et al. (2003), with B-splines fit to the wing veins using FINDWING (see Supporting Information for details), yielding 12 landmarks which were used for all subsequent analysis (Fig. 1). Wing data were included in the analysis for all families with suitable images of both parents and at least one offspring of each sex, with typical families having either two males and two females or three males and three females (after digitizing several thousand wings from families of six offspring, it was determined that little power would be lost by reducing family size). To explore the effect of temperature on the genetic variation, we chose 20 wing traits a priori: centroid size, length to centroid ratio, four allometries (from Weber et al. 2005), five angles (from Whitlock and Fowler 1999), and an additional seven line segments and two angles (described in Table 2). Generalized Procrustes analysis in the statistical package R (procGPA, "shapes" library; Dryden 2007) was used to calculate the centroid size of each wing and to scale and rotate the landmarks to a common orientation, such that the rotated values represent displacement of each landmark from the centroid. Following Mezey and Houle (2005), we ran procGPA separately on eight groups of wings (sires, dams, male offspring, female offspring in the cold vs. hot assay) to avoid introducing error into the Procrustes analysis due to differences in trait variance among these groups. The scaled and rotated landmarks from these analyses were combined back into a single dataset and then used to calculate the length of all of the line segments described in Table 2, while the allometries and angles were calculated from the raw landmark data. We checked all results reported below by running the same analyses on the data that had been scaled and rotated with a single application of procGPA on all pooled data and found no substantial differences.

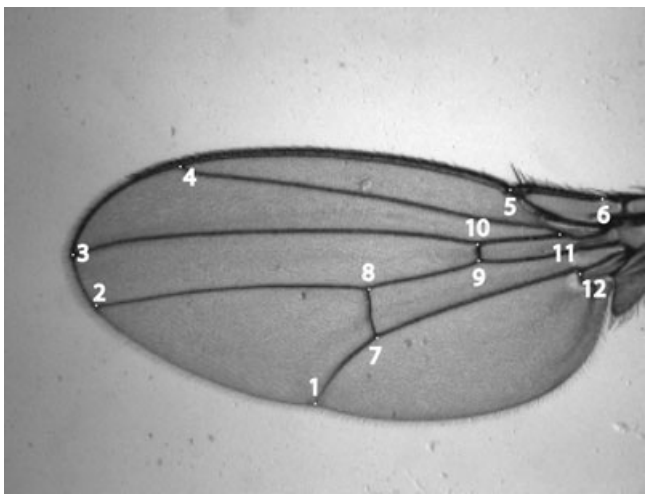


**Figure 1.** Landmarks produced by FINDWING spline fitting.

By choosing a wide range of traits, we maximized the chances of detecting divergence between the homogeneous populations (C and H) for any biologically relevant dimensions of the wing. To identify traits that had likely been under divergent selection between the conditions in the cold and hot chambers, we used maximum likelihood methods to fit a linear mixed effects model to the mean trait values from each replicate population at each assay temperature (each of which was calculated by averaging over the mean values for sires, dams, sons, and daughters). The analysis is a split-plot design, with each replicate being affected by only one treatment but measured for each assay condition. We used the following formula notation to fit this model using the nlme library in R (Pinheiro et al. 2008): lme (trait_value $\sim$ treatment + assay, random = $\sim$1 | replicate/assay, method = "ML"), which treats *treatment* and *assay* as fixed effects and *replicate* as a random effect (with *assay* nested within *replicate*). The average magnitude of trait divergence, $(\bar{Z}_C - \bar{Z}_H)$, was calculated as the mean difference between C and H treatments standardized by the mean trait standard deviation, both of which were averaged across the eight possible combinations of the levels of "sex," "generation," and "assay." We calculated the average between-treatment component of variance ($V_B$) for the C and H treatments alone by fitting the following model: lme (trait_value $\sim$ 1, random = $\sim$1 | *treatment*/*replicate*) to the data for sires and dams in each assay. Phenotypic correlations were analyzed between all pairwise combinations of the traits with evidence for significant divergence by calculating the average correlation across sons, daughters, sires, and dams. For any pairs of traits that yielded $r^2 > 0.1$, we excluded the trait that had a lower average divergence between the C and the H treatments from further analyses. We compared genetic variation among the full set of experimental treatments using this subset of divergent and weakly correlated traits.

When males and females have equal phenotypic variance, narrow sense heritability ($h^2$) can be calculated by regressing the mid-offspring values on the mid-parent values for each trait, with the slope of this relation = $h^2$ (Falconer and Mackay 1996). When males and females have unequal variance, separate heritabilities must be calculated for each sex, with male heritability equal to twice the average of the slopes yielded by regressing sons on sires and daughters on sires, after multiplying the daughter-sire regression by $\sigma_m/\sigma_f$ (with the reverse procedure for female $h^2$, multiplying the son-dam coefficient by $\sigma_f/\sigma_m$; Falconer and Mackay 1996). We used Levene's test ("car" library in R) to compare phenotypic variance between sires and dams for both the hot and cold assay, bulked across all populations. We implemented both of these

**Table 2.** Description of the wing traits used in the preliminary analysis and evidence for divergence in trait means in C versus H and MC versus MH. Divergence between the C and H treatments is represented by the difference in mean trait values scaled by the phenotypic standard deviation,$(\bar{Z}_C - \bar{Z}_H)/\sigma_p$, and by the between-treatment component of variance relative to the $V_A$ maintained within them ($V_B/V_A$; from Table 3). Positive values of ($D_{C-H} - D_{MC-MH}$) indicate that the average phenotypic distance between subpopulations of the M treatment was less than that of the C and H treatments, suggesting constraint due to migration. *P*-values for the comparisons between MC versus MH and C versus H are shown for the test of the null hypothesis of no divergence (df = 24). Rows shaded gray indicate the four divergent and minimally correlated traits used in the analysis of $V_A$; (*) indicates significance following sequential Holm–Bonferroni correction.

| Trait | Description | $(\bar{Z}_C - \bar{Z}_H)/\sigma_p$ | $V_B/V_A$ | $D_{C-H} - D_{MC-MH}$ | $p$ [C vs. H] | $p$ [MC vs. MH] |
|---|---|---|---|---|---|---|
| centroid | centroid | 0.19 | 0.010 | 0.17 | 0.2564 | 0.7949 |
| line7–8 | line between 7 and 8 | −0.16 | 0.006 | −0.16 | 0.2125 | 0.0698 |
| line9–10 | line between 9 and 10 | −0.34 | 0.087 | 0.11 | 0.0036 | 0.1686 |
| line2–12 | line between 2 and 12 | 0.09 | 0.000 | 0.16 | 0.5813 | 0.537 |
| line1–4 | line between 1 and 4 | −0.16 | 0.003 | 0.01 | 0.1827 | 0.3526 |
| line1–5 | line between 1 and 5 | −0.11 | 0.003 | 0.00 | 0.3214 | 0.7944 |
| line2–8 | line between 2 and 8 | 0.07 | 0.037 | 0.04 | 0.4263 | 0.1977 |
| line3–10 | line between 3 and 10 | 0.20 | 0.121 | 0.04 | 0.0556 | 0.2061 |
| angle1–2–4 | angle between 1,2,4 | −0.10 | 0.000 | −0.03 | 0.5260 | 0.4272 |
| angle1–5–4 | angle between 1,5,4 | −0.12 | 0.004 | 0.00 | 0.3296 | 0.1171 |
| angle7–8–9 | angle between 7,8,9 | −0.46 | 0.148 | 0.18 | *0.0002 | 0.0272 |
| angle3–10–4 | angle between 3,10,4 | −0.43 | 0.126 | 0.08 | *0.0025 | 0.0122 |
| angle1–8–2 | angle between 1,8,2 | 0.08 | 0.008 | 0.03 | 0.4972 | 0.3071 |
| angle2–4–9 | angle between 2,4,9 | 0.34 | 0.059 | 0.07 | 0.0347 | 0.7873 |
| angle2–4–8 | angle between 2,4,8 | 0.41 | 0.103 | 0.16 | 0.0116 | 0.8300 |
| allometry1 | line1–2/line4–5 | 0.07 | 0.000 | −0.06 | 0.5806 | 0.2057 |
| allometry2 | line2–4/line1–5 | −0.19 | 0.017 | 0.02 | 0.2348 | 0.1198 |
| allometry3 | line1–4/line5–12 | −0.07 | 0.049 | 0.12 | 0.3786 | 0.9111 |
| allometry4 | line4–5/line1–12 | 0.004 | 0.024 | 0.17 | 0.8738 | 0.5839 |
| centroid/length | centroid/line2–12 | −0.09 | 0.000 | 0.16 | 0.5777 | 0.5178 |

methods of estimating heritability using "lm" in R, calculating the phenotypic standard deviations across all individuals of a given sex for each line assay. Additive genetic variance ($V_A$) can then be estimated from: $V_A = h^2\ (V_{P,sires} + V_{P,dams})/2$ (for $\sigma_m = \sigma_f$), or $V_{A,m} = h_m^2 \cdot V_{P,sires}$ and $V_{A,f} = h_f^2 \cdot V_{P,dams}$ (for $\sigma_m \neq \sigma_f$), where $V_P$ is the phenotypic variance. Additive genetic variance maintained within each line was then compared among treatments using a linear mixed effects model based on a split plot design. The model was fit in R using the following formula: lme ($V_A \sim$ treatment × assay, random = $\sim 1 \mid$ replicate/assay, method = "ML"), while 95% confidence intervals around the mean $V_A$ in each treatment were calculated using "ci" in the "gmodels" library in R, which uses MCMC simulation methods to compute confidence intervals (Warnes et al. 2005).

# Results

## FITNESS DIFFERENCES BETWEEN HOMOGENEOUS LINES

The hot and cold homogeneous populations (C and H) produced more offspring that survived to adulthood when assayed in the temperature at which they had evolved over the course of the experiment (relative to the opposite-temperature homogeneous populations). In the cold assay, the flies from the C and H treatments produced 67.0 and 59.4 offspring on average, respectively, over a 4-day laying period, while in the hot assay the C and H treatment flies produced an average of 84.3 and 86.4 offspring, respectively, over a 2-day laying period. Variance in productivity among replicate populations was lower in the cold assay than the hot assay, so to statistically analyze this pattern we log-transformed all productivity measures. To test the significance of the interaction between treatment, assay, and productivity, we subtracted the log-transformed mean productivity of each replicate in the cold assay from the same quantity in the hot assay (Fig. 2), and compared the differences in the resulting values between the C and H populations using a Wilcoxon rank sum test ($W = 1$, $P = 0.016$) and Welch's test for unequal variances ($t = 2.37$; df = 4.13, $P = 0.075$). The differences in productivity are statistically significant by the former test and marginally nonsignificant by the latter, suggesting that these lines have adapted to their respective environments. A more accurate measure of fitness accounting for the effects of density and development time would likely reveal

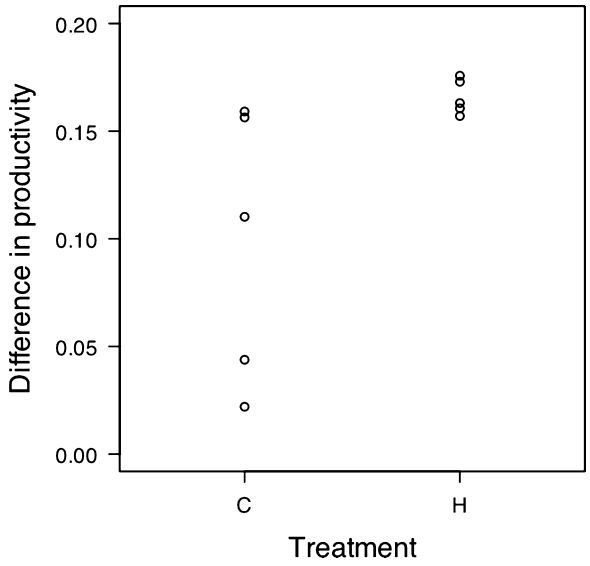


**Figure 2. Difference in mean log-transformed productivity between hot assay and cold assay for each replicate of the homogeneous treatments.**

more pronounced differences between the homogeneous populations. We also examined selection differentials for the 20 traits by regressing productivity on dam phenotype but found no consistent significant associations (either when grouped by replicate or bulked across all replicates; results not shown).

## TRAIT DIVERGENCE BETWEEN C AND H LINES

Of the 20 traits that we included in our initial survey, five were found to have $P$-values $< 0.05$ for the test of different trait means between the C and H treatments (Table 2). Of these five traits, only ANGLE2–4–9 and ANGLE2–4–8 were highly correlated with each other ($r^2 > 0.1$). Based on the average trait divergence between the homogeneous populations (Table 2), we excluded ANGLE2–4–9 from further study, as it was less divergent than ANGLE2–4–8. Of the remaining four traits, only ANGLE7–8–9 and ANGLE3–10–4 had $P$-values that were significant following sequential Holm–Bonferroni correction for multiple comparisons and are thus the only traits for which we have strong evidence of divergence. Nevertheless, we used all four remaining traits that were divergent and weakly correlated to explore the effect of the experimental heterogeneity treatments on the maintenance of genetic variation (referred to hereafter as "divergent traits"; shaded gray in Table 2). We also examined the effect of the treatments on $V_A$ for CENTROID, as size is commonly reported to diverge with temperature, although we did not see significant differences in size between the homogeneous treatments (see Supporting Information for a discussion of the possible reasons for this). We refer to these five traits as "biologically motivated," as we used criteria defined before analyzing changes in variance to identify them as biologically interesting to avoid the loss of power associated with post hoc corrections for multiple comparisons.

## TRAIT DIVERGENCE UNDER LIMITED MIGRATION

The populations in the "M" treatment consisted of one cage in each chamber, with two randomly selected gravid females moved reciprocally between the cages every 4 weeks to simulate limited migration rate of approximately $m = 0.001$. To ask whether this low rate of migration was sufficient to constrain the adaptation of the subpopulations in each chamber, we compared the average divergence between the hot and cold cages within each replicate of the M treatment to the average divergence between all possible pairs of replicate lines between the C and H treatments for each of the divergent traits. For all four traits, the average value of the absolute difference between MC and MH cages within a replicate ($D_{MC-MH} = (\sum_i^{n=5} |\bar{Z}_{MC,i} - \bar{Z}_{MH,i}|/n)/\sigma_P$) was less than the average value of the absolute difference between the C and the H cages among all replicates ($D_{C-H} = (\sum_i^{n=5} \sum_j^{n=5} |\bar{Z}_{C,i} - \bar{Z}_{H,j}|/n^2)/\sigma_p$; Table 2). For two out of the four traits, the mean trait values averaged over all five replicates for MC and MH fell between the values for the C and H traits and diverge in the expected direction, such that H > MH > MC > C or the reverse (e.g., Fig. 3). The two exceptions to this pattern nearly conformed with the expectation, but differed in the sign of one comparison (ANGLE3–10–4 had MH > H > MC > C while LINE9–10 had H > MC > MH > C).

A problem with drawing inference from this evidence, however, is due to ascertainment bias, as these four traits were selected

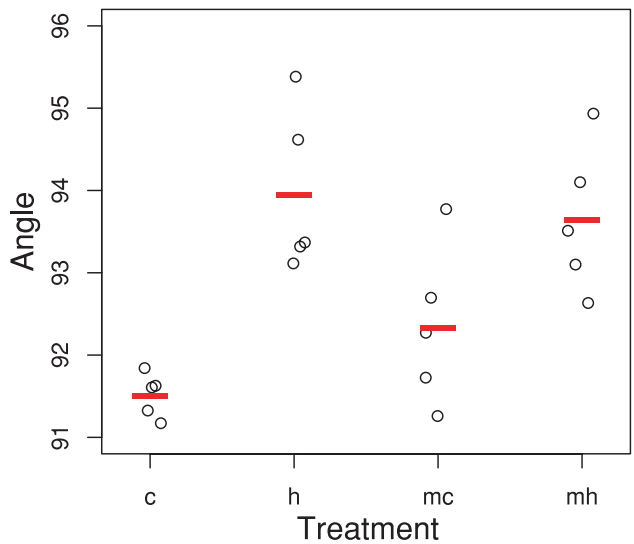


**Figure 3. Reduced divergence between cages in the limited migration treatment. Mean trait values for ANGLE7–8–9 averaged across both assay temperatures for the five replicates in each treatment for the two homogeneous treatments and the hot and cold cages of the limited migration treatment. Grand means across all replicates shown as gray bars.**

because they were known to be divergent between populations in the C and H treatments. As such, the finding of less divergence between MC and MH could simply be due to the trait selection criteria. If ascertainment bias is an issue, we should find the reverse pattern (more divergence between MC and MH than C and H) if we select traits based on divergence between MC and MH. To check for ascertainment bias, we used the above methodology to examine divergence between MC and MH, and found that only two of the 20 traits showed any evidence of divergence (ANGLE7–8–9). Neither of the *P*-values for these traits was statistically significant after adjusting for multiple comparisons, and both of the traits showed more divergence between C and H than between MC and MH (Table 2). This general pattern therefore suggests that migration is constraining adaptation and divergence between MC and MH in at least some of the traits, although we do not have strong support for any individual comparisons (we did not have enough replication to detect significant differences between MH and H or MC and C for the four divergent traits).

### DIFFERENCES IN $V_A$ AMONG TREATMENTS

Narrow-sense heritabilities for the five biologically motivated traits were high and significantly different from zero for most replicate populations and traits using the equal variance heritability estimate (Table 3). Despite these high heritabilities, we did not find any statistically significant differences in either the additive genetic variance or heritability among any of the experimental treatments for any of the biologically motivated traits identified above, using the equal variance method for estimating heritability ($P > 0.3$ in all cases; df = 24; see Fig. S1 for plots of the point estimates for each of these traits). While Levene's test suggests that phenotypic variance differed between the sexes for at least one of the five biologically motivated traits, the ratio of $\sigma_{P,m}/\sigma_{P,f}$ was not very different from one in all cases (Table 4). In any case, we still found no effect of treatment on $V_A$ when calculated using the unequal variance method for estimating heritability. Statistical nonsignificance aside, we also did not observe any obvious trends in the rank-order of $V_A$ by treatment across the five traits (Table 3). This lack of effect of experimental treatment on standing genetic variation was also insensitive to whether we excluded populations with nonsignificant slopes for the offspring–parent regression from the analysis, analyzed variation in $V_A$ for the males and females assays separately, or averaged measures of $V_A$ for each replicate across both assays.

Approaching the limited migration treatment as a pair of connected subpopulations and assaying the genetic variance within each cage (i.e., MC and MH) focuses on the effect of migration on the maintenance of variation within populations. If we instead pool the families from both cages within each replicate of the limited migration treatment ($M_{pooled}$), we can ask whether this treatment maintains more genetic variance across the two environments. Taking this approach, we see much the same results as when we consider the subpopulations individually (Table 3), with no significant increase in $V_A$ in any trait relative to the lines in either the homogeneous or the other heterogeneous treatments. This further suggests that this type of heterogeneity does not substantially influence the maintenance of standing genetic variation.

Despite using the most powerful analysis available by testing the null hypothesis of equal $V_A$ across all treatments, there was no evidence of any effect of any of the heterogeneous treatments on genetic variation. We then expanded the above analysis to include all traits, not just those that displayed evidence of divergence, but still found no evidence for an effect of the heterogeneous treatments on genetic variance (even without Bonferroni corrections; see Supporting Information).

It is notoriously difficult to estimate variance with high precision. In any analysis that fails to detect an effect when one was expected, it is important to consider the power of the methods employed. Namely, if there was a true effect of heterogeneity on genetic variance that we failed to detect, can we place an upper bound on the maximum effect that would be likely to have gone undetected by the study design we employed? The 95% confidence intervals around the estimates of mean $V_A$ for each treatment are reasonably small; in almost all cases the upper bounds of intervals for the heterogeneous treatments do not exceed the lower bounds of the intervals of the homogeneous treatment by more than a factor of two (Table 3). We also used a simulation method that tested the average factor by which $V_A$ within one treatment would have to increase to be detected by our analysis, finding similar results to the confidence intervals (results not shown). Thus, although it is impossible to rule out an effect of the heterogeneous treatments on genetic variance, we may be reasonably confident that they could not have caused an increase of genetic variance of more than approximately twofold, relative to the homogeneous treatments. If our experiment had resulted in a real effect of heterogeneity on variance of the magnitude observed by Mackay (up to 3.5 fold; 1981), we should have had the statistical power to detect it.

## Discussion

Population genetic theory generally suggests that heterogeneous environments have the potential to maintain more genetic variation within populations than homogeneous environments (Felsenstein 1976; Bürger and Gimelfarb 2002; Spichtig and Kawecki 2004). Despite high trait heritabilities (Table 3), divergence in trait values (Table 2), and evidence for environment-dependent differences in productivity between homogeneous lines (Fig. 2), we found no evidence that any of the environmentally heterogeneous treatments maintained more variation than the homogeneous treatments (Table 3). Although limited migration in the M treatment

**Table 3.** Trait heritabilities ($h^2$), additive genetic variation averaged across replicate and assay ($V_A$), and 95% confidence intervals (95% CI) for these values in each treatment. Numbers in brackets adjacent to each heritability estimate represent the number of replicate/assay combinations with $h^2$ significantly different from 0 (out of 10). Treatment symbols indicate: cold homogeneous (C); hot homogeneous (H); spatial heterogeneity with limited migration, cold and hot subpopulations (MC and MH) and from samples pooled from both subpopulations (POOLED); spatial heterogeneity with panmictic migration (S); temporal heterogeneity (T).

| | | Treatment | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | C | H | MC | MH | $M_{POOLED}$ | S | T |
| LINE9–10 | $h^2$ | 0.54(10) | 0.56(10) | 0.49(9) | 0.54(10) | 0.53(10) | 0.46(8) | 0.57(10) |
| | $V_A$ (x$10^{-5}$) | 0.173 | 0.190 | 0.170 | 0.184 | 0.192 | 0.142 | 0.203 |
| | 95% CI | (0.116–0.204) | (0.133–0.221) | (0.114–0.202) | (0.128–0.216) | (0.140–0.223) | (0.085–0.173) | (0.146–0.235) |
| ANGLE7–8–9 | $h^2$ | 0.63(10) | 0.64(10) | 0.55(10) | 0.61(9) | 0.59(10) | 0.67(10) | 0.6(10) |
| | $V_A$ | 19.11 | 17.98 | 16.27 | 19.19 | 18.54 | 17.92 | 18.29 |
| | 95% CI | (17.35–25.5) | (16.23–24.38) | (14.51–22.66) | (17.43–25.58) | (16.8–24.9) | (16.16–24.32) | (16.53–24.68) |
| ANGLE3–10–4 | $h^2$ | 0.62(10) | 0.61(10) | 0.66(10) | 0.65(10) | 0.69(10) | 0.62(10) | 0.7(10) |
| | $V_A$ | 0.333 | 0.362 | 0.328 | 0.358 | 0.330 | 0.313 | 0.350 |
| | 95% CI | (0.273–0.397) | (0.302–0.426) | (0.268–0.392) | (0.298–0.422) | (0.320–0.436) | (0.253–0.377) | (0.29–0.414) |
| ANGLE2–4–8 | $h^2$ | 0.67(10) | 0.67(10) | 0.69(10) | 0.64(10) | 0.67(10) | 0.73(10) | 0.65(10) |
| | $V_A$ | 1.645 | 1.442 | 1.526 | 1.591 | 1.650 | 1.717 | 1.574 |
| | 95% CI | (1.579–2.159) | (1.375–1.956) | (1.459–2.04) | (1.525–2.105) | (1.519–2.150) | (1.651–2.231) | (1.507–2.088) |
| CENTROID | $h^2$ | 0.43(8) | 0.43(7) | 0.59(9) | 0.41(9) | 0.50(10) | 0.54(10) | 0.46(9) |
| | $V_A$ | 26.82 | 24.81 | 30.63 | 25.66 | 26.82 | 28.15 | 26.37 |
| | 95% CI | (23.68–38.77) | (21.66–36.75) | (27.48–42.57) | (22.51–37.6) | (26.97–40.27) | (25.01–40.09) | (23.23–38.31) |

**Table 4. Test of the differences in phenotypic variance between males and females from the parental generation using Levene's test (df = 1) for the five biologically motivated traits.**

| | $\sigma_{P,m}/\sigma_{P,f}$ | | Levene's test $P$-value | |
|---|---|---|---|---|
| Trait | Hot assay | Cold assay | Hot assay | Cold assay |
| LINE9–10 | 0.96 | 1.06 | 0.0672 | 0.0681 |
| ANGLE7–8–9 | 1.03 | 0.96 | 0.5809 | 0.0535 |
| ANGLE3–10–4 | 1.05 | 1.05 | 0.1506 | 0.3977 |
| ANGLE2–4–8 | 0.95 | 0.94 | 0.0439 | 0.0052 |
| CENTROID | 1.05 | 1.03 | 0.2364 | 0.7443 |

seems to have constrained local adaptation and trait divergence between the hot and cold cages (Table 2), it does not seem to have had an effect on the maintenance of variance within the subpopulations (Table 3). Thus, although our study has replicated the conditions under which one would expect to see an effect of heterogeneity on $V_A$, we witnessed no such effect.

Before discussing the interpretation of these results, it is helpful to briefly review theory about the maintenance of polymorphism in heterogeneous environments. We know of three main mechanisms through which spatial environmental heterogeneity can favor the maintenance of genetic variation within populations: (1) local adaptation and genetic divergence between environments, with the potential for a subsequent increase in variation within environments due to gene flow (Spichtig and Kawecki 2004; Yeaman and Jarvis 2006); (2) mean heterozygote overdominance (Levene 1953); and (3) environmentally induced variation in the phenotypic effects of alleles, causing balancing selection (Gillespie and Turelli 1989; Turelli and Barton 2004). The second and third mechanisms can also apply to temporally heterogeneous environments (Felsenstein 1976; Turelli and Barton 2004), while temporal fluctuations in the direction of natural selection can also maintain more variation than homogeneous conditions, provided the magnitude of the change in optimum is large relative to the width of the fitness function and the period is sufficiently long (Bürger and Gimelfarb 2002). While in nature each of these mechanisms may influence observed patterns of genetic variation to some extent, we did not find evidence that the heterogeneous treatments maintained much more variance than the homogeneous treatments, suggesting that these mechanisms had little effect under our experimental conditions.

Explanations based on either temporally or spatially fluctuating selection depend upon the strength of this selection and the difference between optima, which in this experiment is indirectly represented by the divergence in trait means between the C and H treatments. In our experiment, the mean phenotypes of two constant environments diverged in response to selection, but not by a large amount. The between-treatment variance components for C and H were never larger than ~15% of the magnitude of the standing $V_A$ for any trait (Table 2). This perhaps represents the upper bound for an increase in variance within populations caused by local adaptation, which would have been difficult to measure given the power of this study (and in any case, would be much lower than the 3.5-fold difference observed by Mackay 1981). While the T treatment was set up as a potential test of the effects of temporal environmental variation, the divergent selection induced in the hot and cold treatments was unlikely to have been strong enough to have reproduced conditions in which we would expect to see an effect based on fluctuating selection (as per Bürger and Gimelfarb 2002). Given that Bürger and Gimelfarb (2002) only found increases in genetic variance of more than twofold under large variations in optimum that changed over more than four generations, the negative result in the T treatment does not constitute strong evidence against the importance of this mechanism. By contrast, models of overdominance or environmental variation in allelic effects make no strong predictions about the relationship between variance within populations and the variance among populations. The results of this study thus demonstrate that neither the mean overdominance model (Levene 1953; Felsenstein 1976) nor the genotype × environment interaction model (Gillespie and Turelli 1989; Turelli and Barton 2004) led to a large change in variance for these populations.

Unfortunately, as far as we know, no study that has documented an increase in variance or heterozygosity under heterogeneous conditions has also explored whether the environmental variable used in the experiment caused divergence in trait values or allele frequencies between populations exposed to constant conditions of one extreme or the other (see Introduction). As such, it is difficult to infer whether the positive results in other heterogeneity-variance studies were due to fluctuating selection, overdominance, or genotype × environment interaction. The rates of divergence observed in our study are roughly similar to others that have studied response to laboratory natural selection in *Drosophila*. Santos et al. (2004) found that an overall wing shape metric diverged at a rate of ~0.01 $\sigma_p$ per generation (Haldanes) in response to divergent temperature treatments, which is only twice as fast as the rate for ANGLE7–8–9 found here. Cavicchi and colleagues (1989) found that wing length diverged in response to temperature by approximately 2.25 phenotypic standard deviations after 208 generations of divergent laboratory natural selection, which yields a rate of divergence similar to that observed in our experiment if constant gradual change is assumed (two populations were founded from a long-term stock that had been maintained at 18°C, and were subsequently maintained at 28°C). As such, it is reasonable to assume that the strength of divergent selection resulting from our experimental design was on the same order of magnitude as the other heterogeneity-genetic variance studies, but may have been considerably weaker than is commonly found

in nature. Assuming the heterogeneity-variance studies used designs that resulted in a similar strength of divergent selection, it seems unlikely that the positive results they observed could have been due to the fluctuating selection mechanism; future studies should measure trait divergence between control populations to test the possible mechanisms involved more directly.

Interestingly, while the conditions of our experiment did not result in an increase in variance in the M treatment, the low migration rate ($m \sim 0.001$) appears to have been high enough to have constrained divergence between the MC and MH subpopulations for traits that diverged between the C and H treatments (Table 3). Several observational studies have shown patterns consistent with constrained adaptive divergence due to gene flow (e.g., Moore et al. 2007; reviewed in Räsänen and Hendry 2008), while others have found increased response to selection following manipulations that reduced gene flow (Riechert 1993; Nosil 2009). Our experiment adds to this growing body of work; to our knowledge, it provides the first experimental evidence of constraint due to a controlled level of gene flow. It would be interesting to know how commonly variance within populations is increased within populations whose divergence is constrained by gene flow.

If mutation provides a sufficient explanation for the maintenance of variation, then evolvability is largely determined by processes that can operate within a single population, irrespective of the environment. If instead, the maintenance of variation also depends upon the spatial and temporal qualities of the environment that a species inhabits, then evolvability may depend upon processes that operate at the meta-population level. If migration between populations maintains a considerable fraction of the standing variation within populations, then habitat loss and fragmentation may have long-term impacts on evolvability and species survival, especially in the face of environmental and climatic change. The results of our study suggest that neither the overdominance (Levene 1953) nor the genotype $\times$ environment interactions (Gillespie and Turelli 1989; Turelli and Barton 2004) result in the maintenance of much more variation for the traits, species, and laboratory ecology that we studied. In light of the variability in genetic responses to heterogeneity among other similar experimental studies, we conclude that it is unlikely that a single mechanism will provide a sufficient general explanation for the maintenance of variation. Rather, we suspect that the importance of different mechanisms will vary widely among traits, species, and environments. Further empirical studies, especially employing manipulations in natural populations, are likely to provide the best evidence for the relative importance of these mechanisms in nature.

## ACKNOWLEDGMENTS

We wish to thank L. Correia for help with the assays, I. Chen, B. Glover, M. Wood, and R. Best for help with the maintenance of experimental lineages, R. Mottus and his assistants for organizing the cooking of the fly food, and T.F.C Mackay, S. Otto, D. Schluter, S. Aitken, D. Irwin, M. Kirkpatrick, J. Whitton, and the SOWD lab group for help discussing the experiment and results. These people and anonymous reviewers also provided very helpful comments on earlier drafts of this article. Rick White provided helpful advice on statistical models in R. This work was supported by an NSERC Discovery grant to MCW and an NSERC PGS-D to SY.

## LITERATURE CITED

Beardmore, J. A. 1961. Diurnal temperature fluctuation and genetic variance in *Drosophila* populations. Nature 189:162–163.

Bürger, R., and A. Gimelfarb. 2002. Fluctuating environments and the role of mutation in maintaining quantitative genetic variation. Genet. Res. 80:31–46.

Cavicchi, S., D. Guerra, V. Natali, C. Pezzoli, and G. Giorgi. 1989. Temperature-related divergence in experimental populations of *Drosophila melanogaster*. II. Correlation between fitness and body dimensions. J. Evol. Biol. 2:235–251.

Charlesworth, B., T. Miyo, and H. Borthwick. 2007. Selection responses of means and inbreeding depression for female fecundity in *Drosophila melanogaster* suggest contributions from intermediate-frequency alleles to quantitative trait variation. Genet. Res. 89:85–91.

Dryden, I. L. 2007. The shapes package: statistical shape analysis in R. Version 1.1–1.

Falconer, D., and T. F. C. Mackay. 1996. An introduction to quantitative genetics. 4th ed. Prentice Hall, Essex, England.

Felsenstein, J. 1976. The theoretical population genetics of variable selection and migration. Annu. Rev. Genet. 10:253–280.

Garcia-Dorado, A., P. Martin, and N. Garcia. 1991. Soft selection and quantitative genetic variation: a laboratory experiment. Heredity 66:313–323.

Gillespie, J. H., and M. Turelli. 1989. Genotype-environment interactions and the maintenance of polygenic variation. Genetics 121:129–138.

Haley, C. S., and A. J. Birley. 1983. The genetical response to natural selection by varied environments. II. Observations on replicate populations in spatially varied laboratory environments. Heredity 51:581–606.

Hedrick, W. P., M. W. Ginevan, and P. E. Ewing. 1976. Genetic polymorphism in heterogeneous environments. Annu. Rev. Ecol. Syst. 7:1–32.

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

Johnson, T., and N. H. Barton. 2005. Theoretical models of selection and mutation on quantitative traits. Phil. Trans. R. Soc. B. 360:1411–1425.

Levene, H. 1953. Genetic equilibrium when more than one niche is available. Am. Nat. 87:331–333.

Linhart, Y. B., and M. C. Grant. 1996. Evolutionary significance of local genetic differentiation in plants. Annu. Rev. Ecol. Syst. 27:237–277.

Long, T. 1970. Genetic effects of fluctuating temperature in populations of *Drosophila melanogaster*. Genetics 66:401–416.

McDonald, J. F., and F. J. Ayala. 1974. Genetic response to environmental heterogeneity. Nature 250:572–574.

Mackay, T. F. C. 1981. Genetic variation in varying environments. Genet. Res. 37:79–93.

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

Mitchell-Olds, T., J. H. Willis, and D. B. Goldstein. 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? Nat. Rev. Genet. 8:845–856.

Moore, J.-S., J. L. Gow, E. B. Taylor, and A. P. Hendry. 2007. Quantifying the constraining influence of gene flow on adaptive divergence

in the lake-stream threespine stickleback system. Evolution 61:2015–2026.

Nosil, P. 2009. Adaptive population divergence in cryptic color-pattern following a reduction in gene flow. Evolution 63:1902–1912.

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.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and the R Core Team. 2008. nlme: linear and nonlinear mixed effects models. R package version 3.1–90.

Powell, J. R. 1971. Genetic polymorphisms in varied environments. Science 174:1035–1036.

Powell, J. R., and H. Wistrand. 1978. The effect of heterogeneous environments and a competitor on genetic variation in *Drosophila*. Am. Nat. 112:935–947.

Räsänen, K., and A. P. Hendry. 2008. Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. Ecol. Lett. 11:624–636.

Riddle, R. A., P. S. Dawson, and D. F. Zirkle. 1986. An experimental test of the relationship between genetic variation and environmental variation in tribolium flour beetles. Genetics. 113:391–404.

Riechert, S. E. 1993. Investigation of potential gene flow limitation of behavioral adaptation in an aridlands spider. Behav. Ecol. Sociobiol. 32:355–363.

Santos, M., P. F. Iriarte, W. Céspedes, J. Balanyà, A. Fontdevila, and L. Serra. 2004. Swift laboratory thermal evolution of wing shape (but not size) in *Drosophila subobscura* and its relationship with chromosomal inversion polymorphism. J. Evol. Biol. 17:841–855.

Spichtig, M., and T. J. Kawecki. 2004. The maintenance (or not) of polygenic variation by soft selection in heterogeneous environments. Am. Nat. 164:70–84.

Turelli, M., and N. H. Barton. 2004. Polygenic variation maintained by balancing selection: pleiotropy, sex-dependent allelic effects and G × E interactions. Genetics 166:1053–1079.

Verdonck, M. V. 1987. Adaptation to environmental heterogeneity in populations of *Drosophila melanogaster*. Genet. Res. 49:1–10.

Via, S., and R. Lande. 1987. Evolution of genetic variability in a spatially heterogeneous environment: effects of genotype-environment interactions. Genet. Res. 49:147–156.

Warnes, G. R., B. Bolker, T. Lumley, and R. C. Johnson. 2005. Gmodels: various R programming tools to model fitting. R package version 2.15.0. Available at: http://CRAN.R-project.org/package=gmodels.

Weber, K., N. Johnson, D. Champlin, and A. Patty. 2005. Many P-element insertions affect wing shape in *Drosophila melanogaster*. Genetics 169:1461–1475.

Whitlock, M. C., and K. Fowler. 1999. The changes in genetic and environmental variance with inbreeding in *Drosophila melanogaster*. Genetics 152:345–353.

Yeaman, S., and A. Jarvis. 2006. Regional heterogeneity and gene flow maintain variance in a quantitative trait within populations of lodgepole pine. Proc. R. Soc. B. 273:1587–1593.

Zhang, X-S., J. Wang, and W. G. Hill. 2004. Influence of dominance, leptokurtosis and pleiotropy of deleterious mutations on quantitative genetic variation at mutation-selection balance. Genetics 166:597–610.

Zhang, X.-S., and W. G. Hill. 2005. Genetic variability under mutation selection balance. Trends Ecol. Evol. 20:468–470.

Associate Editor: J. Kelly

## *Supporting Information*

The following supporting information is available for this article:

**Figure S1.** Additive genetic variance in the five biologically motivated traits for all replicate populations and treatments, measured under the hot assay (red open circles) and the cold assay (black filled circles).

**Table S1.** Mean additive genetic variance for each of the six treatments (applicable multiplication factors shown in brackets beside the trait names), proportion of slopes for offspring–parent regression with significant *P*-values across all replicate populations and both assay conditions, *P*-values for the comparison of mean $V_A$ maintained among treatments (df = 24), and *P*-values for the comparison of $V_A$ between replicates that were infected versus uninfected by mites.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.