doi:10.1111/evo.12771

Quantitative genetic variance in experimental fly populations evolving with or without environmental heterogeneity

Yuheng Huang,[1,2] John R. Stinchcombe,[1] and Aneil F. Agrawal[1]

[1] *Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada*

[2] *E-mail: yheng.huang@utoronto.ca*

Received October 10, 2014
Accepted August 26, 2015

Heterogeneous environments are typically expected to maintain more genetic variation in fitness within populations than homogeneous environments. However, the accuracy of this claim depends on the form of heterogeneity as well as the genetic basis of fitness traits and how similar the assay environment is to the environment of past selection. Here, we measure quantitative genetic (QG) variance for three traits important for fitness using replicated experimental populations of *Drosophila melanogaster* evolving under four selective regimes: constant salt-enriched medium (*Salt*), constant cadmium-enriched medium (*Cad*), and two heterogeneous regimes that vary either temporally (*Temp*) or spatially (*Spatial*). As theory predicts, we found that *Spatial* populations tend to harbor more genetic variation than *Temp* populations or those maintained in a constant environment that is the same as the assay environment. Contrary to expectation, *Salt* populations tend to have more genetic variation than *Cad* populations in both assay environments. We discuss the patterns for QG variances across regimes in relation to previously reported data on genome-wide sequence diversity. For some traits, the QG patterns are similar to the diversity patterns of ecological selected SNPs, whereas the QG patterns for some other traits resembled that of neutral SNPs.

KEY WORDS: **Conditional neutrality, *Drosophila melanogaster*, environmental antagonism, maintenance of genetic variance, spatial and temporal heterogeneity.**

Genetic variation in fitness has important impacts on several aspects of evolution, including the potential to adapt to novel selection pressures, the evolution of life-history traits, and the fitness consequences of different reproductive modes (e.g., asexual vs. sexual; selfing vs. outcrossing). Therefore, it is fundamental to understand the processes maintaining genetic variation in fitness within populations (Mitchell-Olds et al. 2007; Leffler et al. 2012). Although mutation–selection balance undoubtedly contributes to the variation in fitness, it is unable to fully account for empirical observations (Charlesworth and Hughes 2000; Johnson and Barton 2005). The alternative is that variation is maintained by some forms of balancing selection, including negative frequency-dependent selection, overdominance, and environmental heterogeneity. The relative importance of these processes is still unknown. In this study, we examine how different forms of environmental heterogeneity affect quantitative genetic (QG) variance in fitness-related traits in experimental *Drosophila melanogaster* populations.

A common feature of natural environments is that they change over time or space, or both. The predicted effects of different types of heterogeneity on genetic variation in fitness depend on assumptions regarding the genetic architecture of traits under differential ecological selection. If alternative alleles are favored in different environments (environmentally antagonistic selection, EAS), classic theory predicts low levels of variation in populations experiencing constant environments, more variation with temporal heterogeneity, and the highest levels of genetic variation for fitness in populations experiencing spatial heterogeneity (i.e., $V_{Const} < V_{Temp} < V_{Spatial}$; Levene 1953; Dempster 1955; Felsenstein 1976).

Rather than being environmentally antagonistic, allelic variants could be conditionally neutral (CN) whereby the two alleles

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

at a locus are selectively neutral in one environment (e.g., Environment A) but with differential fitness effects in an alternate environment (Environment B) (Kawecki 1994; Fry 1996; Whitlock 1996). The disfavored allele will be eliminated in a population that only experiences Environment B, but the polymorphism could persist for much longer in a population that only experiences Environment A (Hoffmann and Merila 1999). Therefore, the genetic variation for fitness *assayed* in Environment B will be much higher for populations constantly evolved in Environment A than those in Environment B. The reverse would be expected if the fitness was assayed in Environment A, assuming some other loci were neutral in A but selected in B. In other words, we expect that a population that never experiences an environment will express higher genetic variation for fitness than a population adapted to that environment (i.e., $V_{Adapted} < V_{Nonadapted}$). For populations that experience both environments, CN alleles experience half as much directional selection. Therefore, before reaching equilibrium, populations under heterogeneous regimes are still expected to harbor more genetic variation for this locus than the constant populations adapted to the selective environment (i.e., $V_{Adapted} < V_{Heterogeneous}$).

Comparing different forms of environmental heterogeneity, one would predict variation would persist longer with spatial heterogeneity than temporal heterogeneity (i.e., $V_{Temp} < V_{Spatial}$) because the persistence of alleles whose fitness vary over time is determined by their geometric mean fitness over time (ignoring complications from dominance; Felsenstein 1976). It should be noted that EAS and CN are only two of many possible models of gene action. Across the genome, some loci under differential ecological selection may follow one model, whereas other loci follow the alternative. Other loci may follow intermediate models between them (e.g., alternative alleles are favored in different environments with very different strengths of selection), but the amount of fitness variation caused by these loci should follow the predictions shared by the two models (e.g., $V_{Adapted} < V_{Heterogeneous}$ and $V_{Temp} < V_{Spatial}$). Our main goal here is to test hypotheses about the levels of genetic variation in different selective regimes; we do not attempt to identify the mode(s) of gene action underlying this variation.

Some of the predictions above rely on populations being at or near equilibrium. Other sources of variation (other than EAS, CN, or related modes of gene action) could be important in the transitory phases of adaptation. Moreover, genetic drift can cause patterns different from those predicted by deterministic models. The work presented here and earlier experiments discussed below typically involve populations with small to moderate population sizes that have existed in their experimental selective regimes for tens to hundreds of generations. Thus, deviations from a deterministic equilibrium model represent an important set of caveats in interpreting the results. Fortunately, those genes with large effects, which have the inherent potential to make large contributions to the variance, should be less affected by genetic drift and approach their equilibria quickly.

Since the 1960s, the effect of environmental heterogeneity on QG variance has been examined using experimental evolution in *Drosophila.* Of the early studies, mostly on bristle number, comparing homogeneous environments with temporal heterogeneity (Beardmore 1961; Long 1970; Verdonck 1987) or with spatial heterogeneity (Garcia-Dorado et al. 1991), only one found a significant increase in additive genetic variation (V_A) in temporally heterogeneous environment (Beardmore 1961). Also, none of the studies compared the effects of temporal and spatial heterogeneity specifically. An influential study (Mackay 1981; using ethanol vs. standard fly medium) found that both types of heterogeneous environments had two times more V_A in sternopleural bristles and body size than the homogeneous (constant environment) control. Interestingly, the temporal regime maintained even more V_A than the spatial one, which is not predicted by the classic theory (Felsenstein 1976). However, Mackay's influential experiment included only one type of constant population. Heterogeneous populations evolved in both standard and ethanol-enriched medium but the "constant control" was only maintained in standard medium; there was no "constant ethanol-enriched" control, making it challenging to distinguish between effects of maintenance in ethanol vs. heterogeneity per se. More recently, Yeaman et al. (2010) found no effect of temporally or spatially variable temperature on the maintenance of V_A in wing shape. One difficulty with these studies is that it is often not clear what type of selection the traits experience (e.g., stabilizing, directional, disruptive) and the extent to which it differs between environments.

In other systems, allozyme studies have found heterozygosity tends to increase under heterogeneous environments (McDonald and Ayala 1974; Powell and Wistrand 1978; but see Haley and Birley 1983). However, these results are based on a handful enzyme markers, and it is unclear whether these loci experienced differential selection between environments or affected fitness at all. Similar to the studies with quantitative traits, few of them directly examined the difference between temporally and spatially heterogeneous regimes. More recently, Venail et al. (2011) evolved bacteria under constant or heterogeneous carbon substrates and found the genetic diversity on growth rate consistent with the theory (i.e., $V_{Const} < V_{Temp} < V_{Spatial}$). However, the "spatial heterogeneity" in their study involves no migration between habitats and, thus, is fundamentally different than other studies and theory. Further, the lack of recombination in prokaryotes could magnify the effect of directional selection and balancing selection on the overall genetic variation via strong linkage effects.

Here, we reexamine the effects of environmental heterogeneity using 20 replicate experimental populations of *D. melanogaster* divided equally among four regimes: one

constant regime always in salt-enriched medium (*Salt*) and one constant regime always in cadmium-enriched medium (*Cad*), a temporally heterogeneous regime (*Temp*) in which populations switched between salt and cadmium medium in alternating generations, and a spatially heterogeneous regime (*Spatial*) in which the population was split between the two mediums but surviving adults were mixed before producing offspring for next generation. After 45 generations of experimental evolution, we used a half-sib breeding design to estimate the variance among sire families (V_{sire}), which is directly related to the additive genetic variance (Falconer and Mackay 1996). This was done for three fitness-related traits in both assay environments: egg to adult survival, male body mass, and female body mass.

These populations have been studied previously from other perspectives relevant to genetic variance. Inbreeding depression, measured after approximately 20 generations, varied among selective regimes in several ways (Long et al. 2013). Notably, inbreeding depression was higher in populations from heterogeneous regimes than from homogeneous regimes. Although it also depends on other factors (e.g., dominance), inbreeding depression is proportional to levels of polymorphism such that populations with more polymorphism should tend to have more inbreeding depression (Lynch and Walsh 1997). After approximately 40 generations, pooled sequence data were used to compare levels of within-population sequence diversity (Huang et al. 2014). Levels of diversity varied among selective regimes, but patterns differed between putatively selected versus neutral sites (see Discussion). With the data presented below, this is the first system in which the effects of environmental heterogeneity have been considered with respect to genome-wide patterns of molecular variation as well as QG variation in several important traits. The relatively high levels of uncertainty in estimates of QG variation of individual population limits our ability to make quantitative comparisons between QG and sequence diversity, but we discuss the similarities and differences between the major among-treatment patterns in the two data types.

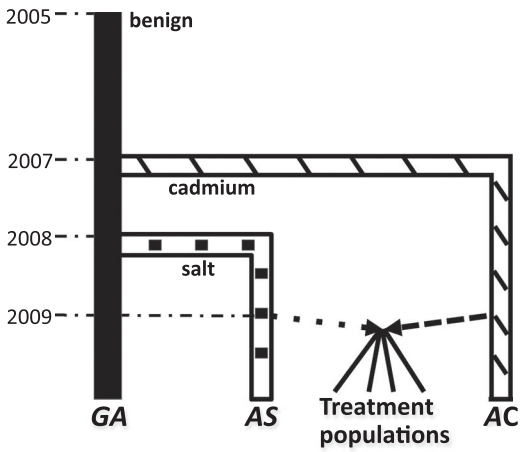

Figure 1. Selection history of the experimental populations. The Grand Ancestor population (*GA*) was maintained in benign laboratory conditions, and was used to initiate populations maintained on salt-enriched media (*AS*) or cadmium-enriched media (*AC*). The treatment populations were produced by crossing two ancestral population *AS* and *AC*. There are five replicate populations of each of the four treatments (not illustrated).

Materials and Methods

HISTORY OF SELECTION POPULATIONS

The selective histories of the ancestral and experimental populations are illustrated in Figure 1. A population of *D. melanogaster* was collected in the Similkameen Valley, British Columbia in 2005 and maintained in regular benign conditions at large size (approximately 2000–4000 adults), referred to as the "*Grand Ancestor*" (*GA*). In July 2007, a subset of flies from the *GA* population was used to initiate a population maintained in a cadmium-enriched medium with population size approximately 1000, referred to as the "*Ancestral Cadmium*" (*AC*) population. In August 2008, a subset of flies from the *GA* population was used to initiate a population maintained in a salt-enriched medium with population size approximately 1000, referred to as the "*Ancestral Salt*" (*AS*) population. During the adaptive history, the concentration of cadmium and salt in the environments was progressively increased, reaching 75 μg/mL and 33 mg/L, respectively, at the starting time of our experiments.

In October 2009, 448 males and 448 virgin females were collected from both the *AC* and *AS* populations and crossed with flies from different populations via mass mating. The offspring from next generation were randomly divided into four selection regimes: constant salt-enriched environment (*Salt*), constant cadmium-enriched environment (*Cad*), alternating each generation between salt- and cadmium-enriched environments (*Temp*), and a "spatial" mix of the two environments each generation (*Spatial*). For the *Spatial* regime, the same number of adult flies produced by the two environments was mixed for each environment for the next generation (i.e., a "soft" selection regime sensu [Wallace 1975]). After 1.5 days for mating, the adult flies were transferred to new food to produce progeny. Twelve days after eggs were laid, the eclosed offspring were collected as new parents. Each selection regime had five replicate populations with population size of 448 adults with equal sex ratio, distributed evenly into 14 vials. Further details of the establishment and history of these populations are described in Long et al. (2013).

BREEDING DESIGN, SURVIVAL, AND BODY MASS MEASUREMENTS

We aimed to study QG variance in traits with strong connections to fitness because the predictions from the simple models

discussed above pertain to variance in fitness. Ideally, we would have measured male mating success and female fecundity rather than body mass, but this was logistically impossible when attempting a half-sib design with 20 populations. Body mass is expected to be important for male mating success and female fecundity (Bangham et al. 2002; Byrne and Rice 2006). Although it is generally believed there is directional selection on female body size in *D. melanogaster*, the evidence for directional selection on male body size is more mixed (Pitnick and Garcia-Gonzalez 2002; Friberg and Arnqvist 2003; Prasad et al. 2007) and caution should be used in interpreting male body size. Furthermore, in our populations, the parents for the next generation need to eclose within 12 days. Therefore, body size may be under conflicting selection pressures because of its presumed negative relationship with development rate but its positive relationship with female fecundity and (putatively) male mating success. Although stabilizing selection on body size is a possibility, it is worth noting that deleterious alleles reduce body size (Sharp and Agrawal 2012; Bonduriansky et al. 2015), as expected if selection on body size is usually positive and directional.

The genetic crosses to assay additive genetic variation started in generation 45 after the establishment of the experimental populations. The 20 populations were divided into five blocks for each of the two environments (75 μg cadmium or 6% salt medium, assayed separately). Each block included one replicate population from each selective history, assayed in one environment. The last block was assayed in generation 57. To control for maternal effects, flies from different populations were reared in the same benign environment for one generation before the assay. The resulting offspring were used to produce approximately 40 half-sib families (one sire mated with four dams). Each mated female was allowed to lay eggs for approximately 20 h in one vial containing salt- or cadmium-enriched medium, as appropriate for the block. After removing the focal female and counting the number of eggs per vial, about 130 competitor eggs were added to each vial via pipetting. The competitor eggs were collected from a population of salt- or cadmium-adapted white-eye flies, used in the different assay environments, respectively. After 12 days, the number of focal adult flies from each assay vials was recorded and divided by the number of focal eggs to calculate the survival (p) for the dam family. To measure the body mass, three focal males and three focal females are sampled randomly from the progeny for each dam family. We measured the combined dry weight of three offspring of each sex. In some cases, fewer than three offspring were weighed. For data analysis, we used the mass per offspring (i.e., dividing the combined mass by the number of offspring weighed).

ANALYSIS

We analyzed each trait separately; we did not attempt any multivariate analyses. In our first set of analyses, each population was analyzed separately. For each trait, the sire variance (V_{sire}) and residual ($V_{residual}$) for each trait within each replicate population in each environment was estimated by the package *MCMCglmm* (version 2.21) in *R* (version 3.1.2). The random effects model was

$$Z_{ij} = \mu + S_j + \varepsilon,$$

where the Z_{ij} is the trait value of the ith dam family within the jth sire and ε is the residual. The μ is the mean for the population. The sire effect S was treated as random effect. Note that no "dam" effect is included in the model because for each trait we only had a single observation per dam. For male and female mass, we used family = "Gaussian," whereas survival was modeled as a binomial trait using family = "multinomial2." (Note *MCMCglmm* assumes additive overdispersion and this is responsible for the residual term in the model.) For male and female body mass, we excluded the outliners beyond 3 SDs for each population. We used priors that were uninformative with respective to variances ($V = 10^{-16}$, nu $= -2$). (The resulting posteriors were consistent with values obtained from non-Bayesian analyses, e.g., using *lmer* or *glmer* for mass and survival, respectively.) Each *MCMCglmm* model was run for 5×10^7 iterations following a burn-in period of 10^4. The samples from the MCMC chain were thinned to every 50 values so that the autocorrelation between stored values was less than 0.1. For survival, values from the MCMC chain (posterior distribution) were back-transformed to the original scale. The most probable estimates (posterior modes) and 95% highest posterior density region of μ, V_{sire}, and $V_{residual}$ were obtained via function posterior.mode and HPDinterval (Hadfield 2010). For survival, the reported values have been back transformed from the link scale to the original (data) scale. The total variance (V_{total}) is the sum of the V_{sire} and $V_{residual}$.

Our main analyses focus on differences among treatments. In addition, we also inspected these patterns while controlling for the differences in phenotypic variance and mean among populations by standardizing V_{sire} by the total variance (V_{sire}/V_{total}) or by the mean μ (I_{sire}) (modified from Houle 1992; see also Hansen et al. 2011; Houle et al. 2011; Garcia-Gonzalez et al. 2012):

$$I_{sire} = V_{sire}/\mu^2.$$

To test our hypotheses, we compared the sire variances between specific pairs of selective histories. Both the EAS and CN models predict that populations from the heterogeneous regimes should have higher genetic variation in fitness than populations from the homogeneous treatment when assayed in its adapted environment, though this prediction is stronger under EAS. Both models also predict $V_{Temp} < V_{Spatial}$. Only the CN model predicts $V_{Adapted} < V_{Nonadapted}$.

To test the predictions from the simple models in the Introduction, we made three types of comparisons with respect to the

amount of sire variance in each environment. First, we compared the two constant regimes with each other to contrast the populations adapted to the assay environment with those not adapted to it (*Adapted* vs. *Nonadapted*). Second, we compared populations from constant regime adapted to the assay environment and to the populations from heterogeneous selection regimes (*Adapted* vs. *Temp* as well as *Adapted* vs. *Spatial*). Third, we contrasted the alternative forms of heterogeneity with each other (*Temp* vs. *Spatial*).

Our comparison between treatments accounts for the uncertainty in the V_{sire} estimate for each population as well as the true differences in the variances among populations from the same treatment. First, we obtained the approximate posterior distribution of V_{sire} (back transformed to original scale for survival) from *MCMCglmm* for each population, which we discretized into 1000 bins. Let $P_{i,j}$ be the (posterior) probability that the true value of V_{sire} for replicate i is within bin j. For selection regime t, we assumed the sire variances follow a gamma distribution with shape k_t and scale θ_t, $G(k_t, \theta_t)$. Our goal was to test whether the means of sire variances significantly differ between selective regimes. Therefore, we used the mean (m) and variance (v) to represent the shape ($k = m^2/v$) and scale ($\theta = v/m$) in the gamma probability function for the V_{sire} in each regime. The likelihood for population i from treatment t is $l_{t,i} = \sum_j P_{i,j} D[j,\ G[k_t, \theta_t]]$ where $D[j, G]$ is the density of the gamma function over the interval spanned by bin j. The log likelihood for treatment t is the sum of $\log(l_{t,i})$ across the replicates $i \in \{1, \ldots, 5\}$. In comparing two treatments, the "full" model allowed the gamma distributions for each treatment to have separate means and variances. The maximum likelihood was calculated by function *optim* (using the "Nelder–Mead" method for rough optimization first and then "BFGS" for fine scale of optimization); 50 different initial starting values were used to help ensure we had arrived had a global maximum. We reran the model with the means of the gamma distributions constrained to be equal between regimes (but allowing each gamma distribution to have a unique variance). The difference in -2log-likelihood values between the full model and the constrained model was compared to a chi-squared distribution with one degree of freedom to obtain the P-value.

COMPARISON WITH THE SEQUENCE DATA OF THESE POPULATIONS

As reported elsewhere (Huang et al. 2014), the genome-wide sequence diversity of these populations was previously surveyed. There are different patterns of within-population diversity for sites that are strongly or weakly differentiated between environments. Here, we examine whether the sequence diversity can predict the QG variance across all 20 experimental populations. The model is

$$V_{sire} = \mu + V_H + V_L + \text{Regime} + \text{Block} + \varepsilon,$$

where V_H is the diversity for the sites that are strongly differentiated between environments (enriched for targets of differential ecological selection) and V_L is the diversity for weakly differentiated sites (enriched for sites that are neutral or under uniform selection; see Huang et al. 2014 for more details). We repeated the analysis for each of the three traits in either cadmium or salt environment. This approach is crude as it ignores the considerable uncertainty in estimates of V_{sire} and relies upon the robustness of the analysis to violations of the assumption that the distribution of sire variances within treatments is Gaussian.

Results

DIFFERENCES IN MEANS AMONG SELECTIVE REGIMES

The homogeneous populations (*Cad* or *Salt*) had higher mean survival as well as higher female and male body mass when assayed in the environment of their selective history compared to the homogeneous populations from the alternative environment (Fig. 2). All comparisons between *Cad* and *Salt* are significant at $P < 0.05$ (t-test), except male mass assayed in salt and this is also in the expected direction. These results suggest that the homogeneous populations have adapted to their respective environments and that some of the alleles affecting these three traits are under differential selection between environments.

As shown in Figure 2, the heterogeneous populations (*Temp* and *Spatial*) had considerably higher trait values than the nonadapted homogeneous populations, indicating the heterogeneous populations are reasonably well adapted to both environments. For example, heterogeneous populations have similar means to homogeneous adapted populations for survival in the cadmium environment and male body mass in the salt environment, possibly suggesting an important role for conditional neutrality. The survival results are qualitatively similar to those reported by Long et al. (2013) from an earlier time point (approximately generation 20).

AMONG-SIRE VARIATION IN ALTERNATIVE SELECTION REGIMES

The point estimates for the among-sire variances from *MCMCglmm* models are shown in Figure 3. It is well known that measurement error on variance components is large unless a very large number of families are measured (Robertson 1959; Falconer and Mackay 1996). Because we were estimating variances in a large number of populations, we measured only a moderate number of families per population, and consequently none of the individual population estimates are highly precise. In many cases, the sire variances were not significantly different from zero and the 95% highest posterior density regions are broadly overlapping with each other (Fig. 3). Nonetheless, patterns emerge when comparing

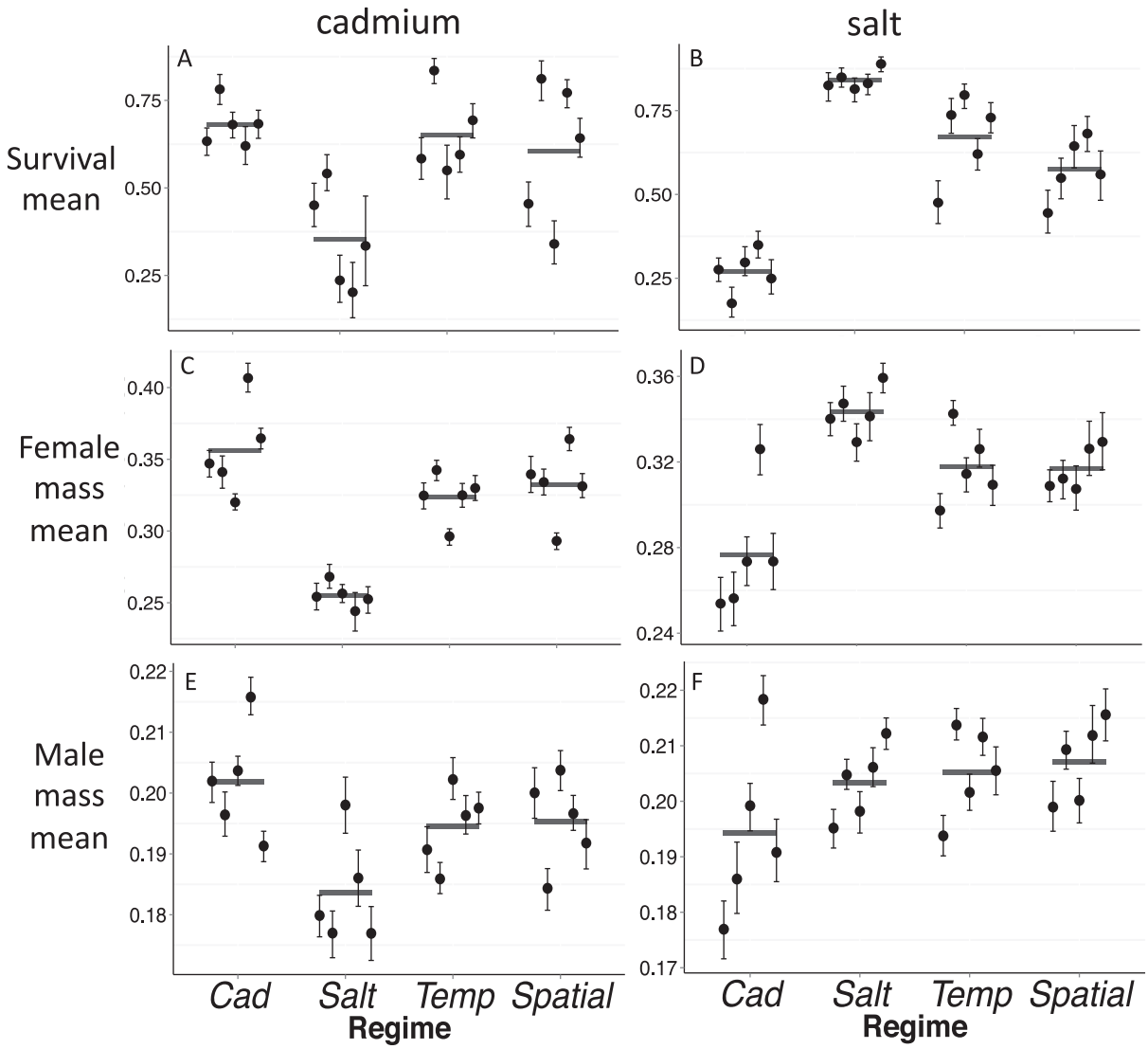

Figure 2. **The mean trait values for each replicate population (represented as points) in each assay environment. The top panel is survival, the middle panel is female body mass and the bottom panel is male body mass. The left column shows results from cadmium assay environment and the right column shows results from salt assay environment. Within each regime, replicate populations 1–5 are shown from left to right. Each dot represents the posterior mode and the error bar represents the 95% highest posterior density region. The gray bar represents the grand mean of survival across the five replicates.**

variances among replicate populations from alternative selective regimes.

For each fitness-related trait in each environment, we statistically tested the following four major predictions. (1) Between the two homogeneous selection regimes, the populations adapted to the assay environment should have less variance than the non-adapted ones (*Adapted* < *Nonadapted*). (2) The adapted homogeneous populations should have less variation than the *Temp* heterogeneous populations (*Adapted* < *Temp*). (3) The adapted populations should have less variation than the *Spatial* populations (*Adapted* < *Spatial*). (4) Between the two heterogeneous regimes, the *Temp* populations should have less variation than the *Spatial* populations (*Temp* < *Spatial*). The results are shown in Figure 3 and a summary of the tests is shown in Table 1. The patterns remain quantitatively similar after standardization by the total variance (Fig. S1) or by the mean (I_{sire}) (Fig. S2).

Visual inspection of Figure 3 with respect to the *Adapted* < *Nonadapted* hypothesis reveals that three of six cases (three traits in two assay environments) trend in the direction of the hypothesis, while the remaining trend in the opposite direction. Two of these tests are significant; both are in the predicted direction but both involve assays in cadmium so that the *Adapted* < *Nonadapted* prediction can be recast as *Cad* < *Salt* ($P = 0.00011$ for survival; $P = 0.0347$ for female mass). In fact, *Cad* tends to have less

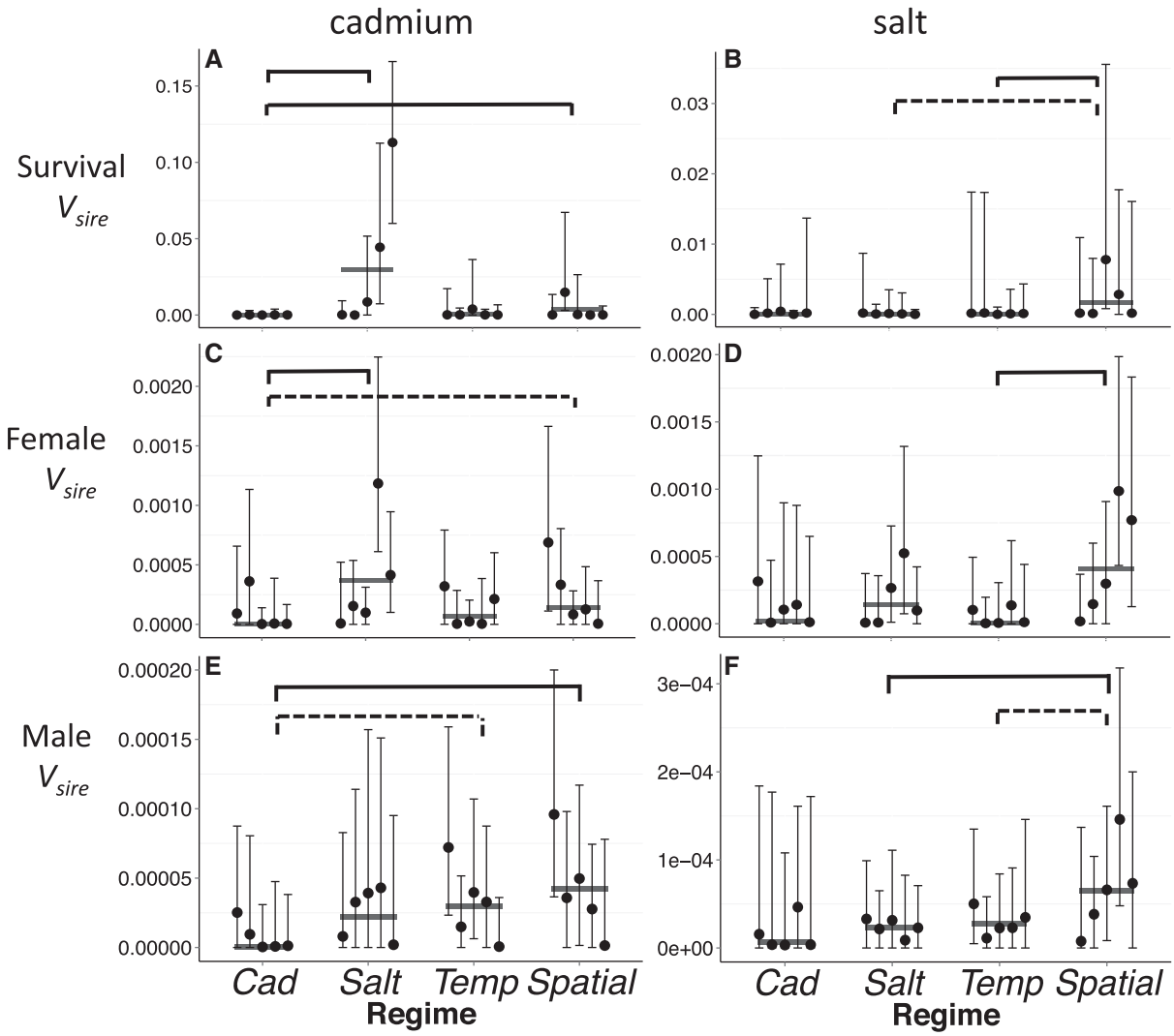

Figure 3. Sire variance (V_{sire}) for survival (A and B), female body mass (C and D), and male body mass (E and F) in cadmium and salt assay environments. Within each regime, replicate populations 1–5 are shown from left to right. The error bars indicate the 95% highest posterior density region. The gray bar presents the most likely mean V_{sire} value for the regime, obtained from the maximum likelihood model. Horizontal solid lines linked across regimes indicate the V_{sire} are significantly different between the two regimes (P-value < 0.05, likelihood ratio test). The dashed line indicates marginally nonsignificant (0.05 < P < 0.09).

Table 1. The numbers of test results that support (S), oppose (O) the hypothesis or are ambiguous (A) across the three traits in each environment.

	Cadmium	Salt	Total
Hypothesis	S, O, A	S, O, A	S, O, A
Adapted < Nonadapted	2, 0, 1	0, 0, 3	2, 0, 4
Adapted < Temp	0, 0, 3	0, 0, 3	0, 0, 6
Adapted < Spatial	2, 0, 1	1, 0, 2	3, 0, 3
Temp < Spatial	0, 0, 3	2, 0, 1	2, 0, 4

"S" (O) means that the two selection histories show a significant difference in sire variance in same (opposite) direction to that predicted (P < 0.05). "A" (ambiguous) means there is no significant difference. The specific comparisons that have significant differences can be seen in Figure 3.

variation than *Salt*, regardless of the assay environment (all six cases). Although the data offer more support for *Adapted < Nonadapted* than they do for the opposite (*Adapted > Nonadapted*), a simpler interpretation is *Cad < Salt*.

For the *Adapted < Heterogeneity* prediction, among the total 12 comparisons (two types of heterogeneity in both environments), 10 cases are in the predicted direction (three of which are significant; Fig. 3, Table 1). However, there is much stronger support for *Adapted < Spatial* than *Adapted < Temp*. For *Adapted < Spatial*, all six cases follow the expected direction, with three tests being significant and two more are only marginally nonsignificant (Fig. 3, *Cad* vs. *Spatial* in cadmium, $P = 0.019$ for survival, $P = 0.043$ for male mass, and $P = 0.083$ for female mass; *Salt* vs.

Spatial in salt, $P = 0.039$ for male mass and $P = 0.062$ for survival). In contrast, the results comparing *Adapted* and *Temp* are mixed: four cases are in the predicted direction, but the other two are opposite; none of the contrasts are significant (though for male mass in cadmium, the comparison is marginally nonsignificant in the predicted direction: *Cad* vs. *Temp*: $P = 0.066$).

Theory predicts that $V_{Temp} < V_{Spatial}$ and our data are generally consistent with this. All six cases follow the predicted direction and two tests are significant and one other is only marginally nonsignificant ($P = 0.026$ for survival in salt; $P = 0.033$ for female mass, and $P = 0.08$ for male mass in salt). Furthermore, for female mass in salt, there is a striking difference in the amount of variation between the two heterogeneities, with the *Temp* regime being lowest among the four regimes and the *Spatial* being highest. This pattern is unexpected from the simple models we introduced earlier and we will discuss the potential reasons for this observation later. In total, these results confirm the distinct effects of temporal and spatial heterogeneity on genetic variation in fitness.

The results above encompass 24 comparisons and the reported *P* values are not adjusted for this. Applying a sequential Bonferroni correction across all 24 comparisons (see Table S1 for full list of raw and adjusted *P* values), leaves only one test significant (*Cad* vs. *Salt* for viability assayed in cadmium). However, we believe this type of correction is inappropriate here (see Discussion).

Because we previously performed whole genome resequencing on these populations (Huang et al. 2014), we attempted to investigate whether the molecular diversity predicts the amount of QG variation. However, there were no significant relationships between molecular diversity and QG variance after accounting for treatment and block. This negative result is not surprising given the high degree of measurement error in individual estimates of V_{sire} (Fig. 3), making it difficult to detect significant relationships at the population level (i.e., within regimes). In the discussion below, we qualitatively compare the patterns for sequence data and the QG results at the regime level.

Discussion

Theory predicts that heterogeneous environments can maintain more genetic variation in fitness than homogeneous environments (Felsenstein 1976; Burger and Gimelfarb 2002; Spichtig and Kawecki 2004; Turelli and Barton 2004). However, the effects of heterogeneity depend on the type of heterogeneity (e.g., temporal or spatial) and the nature of the genetic architecture underlying the traits (e.g., EAS or CN). However, previous attempts using experimental evolution have yielded mixed results and often have not found a significant difference in genetic variation between selection regimes. Further, a number of the previous studies did not compare the effects of different forms of heterogeneity (Beardmore 1961; Long 1970; Riddle et al. 1986; Verdonck 1987; Garcia-Dorado et al. 1991). In other cases, it was unclear whether the different environments were relevant to the measured trait (McDonald and Ayala 1974; Powell and Wistrand 1978; Haley and Birley 1983). In this study, we first showed phenotypic divergences in trait means between the populations in two constant environments in the expected directions. We then compared the QG variation among different selection regimes for these traits, but the considerable measurement error limited our power. Spatial heterogeneity tended to harbor more genetic variation in fitness traits than those maintained with temporal heterogeneity or "adapted" populations from constant environments. However, our results are mixed and many of the comparisons are not statistically significant.

Further, our examination of genetic variances involves 24 comparisons and it is unclear how to deal with multiple testing. Seven of the individual tests are significant even though only approximately one false positive (not seven) is expected when performing 24 tests if there were no true effects. Notably, the seven significant tests do not appear randomly distributed with respect to which of the four tested hypotheses they fall under or the direction of differences, as expected with false positives (Table 1). Yet, a sequential Bonferroni correction renders six of the seven tests nonsignificant, perhaps reflecting the known property of Bonferroni corrections to be overly conservative. Moreover, the standard philosophy motivating multiple comparison corrections does not seem to apply here. Correction for multiple testing is most often performed when numerous variables are being "tried" against a single response variable (e.g., relating SNPs to a trait). In contrast, our tests span four different hypotheses based on theoretical predictions regarding alternative selective regimes (Quinn and Keough 2002). Within each of these four hypotheses, there are six tests considering three traits in each of two environments. For example, two of six tests of the "*Spatial* > *Temp*" hypothesis are individually significant, the other four are not, but all go in the predicted direction. Should these nonsignificant comparisons, which also go in the predicted direction, make us more suspicious of the two comparisons significantly supporting the hypothesis (as implicitly occurs under a correction for multiple tests) or, conversely, strengthen our confidence in the hypothesis (as would occur using a "combined *P*-value" approach)? For conservative readers, we have noted how limited our statistical support is after a sequential Bonferroni correction, though we do not believe this is the most appropriate way to view the results.

COMPARISON IN PATTERNS OF GENETIC VARIATION ACROSS REGIMES WITH SEQUENCE DIVERSITY

We have also applied whole genome resequencing to these same experimental populations (Huang et al. 2014). To our knowledge, this is the first case applying both classic QGs and population

genomics to study genetic variation in the same set of experimental populations. Both approaches have strengths and provide different insights into this long-standing problem. Sequencing provides a relatively "unbiased" survey across the genome; it does not rely on assaying only those traits that are feasible to measure in a large QG experiment in certain environments. Second, it is possible to plausibly distinguish between sites under differential ecological selection (or those linked to such sites) and those that are not, allowing a comparison of how alternative regimes affect qualitatively different classes of sites (e.g., selected vs. neutral). The "selected sites" capture those sites that affected total fitness (directly or indirectly), not just those sites affecting measurable fitness components, enabling a better test of selection-based hypotheses. The "neutral sites" reveal the effect of genetic drift, enabling us to distinguish the effects of ecological differential selection on diversity from the neutral processes.

The classic QG approach allows us to estimate the genetic variance for traits of known biological interest (e.g., life-history traits, morphological traits) and/or fitness components directly. In contrast, fitness effects for SNPs are inferred indirectly and this process is noisy. Sets of SNPs enriched for selective targets are readily found from sequence data, but identifying which of these SNPs truly affect fitness or specific traits can be difficult, if not impossible (Kofler and Schlötterer 2014; Baldwin-Brown et al. 2014). In our data (Huang et al. 2014), many SNPs among the set enriched for targets of differential ecological selection are expected to be due to linkage effects and not true targets of selection. Further, the set of sites that we treat as reflecting patterns of neutrality will include some sites that affect fitness (both those under uniform selection across environments and those under differential ecological selection but with small effects). In addition, the additive genetic variance provided by the QG approach may be more meaningful than sequence variation because the former reflects the effects of the alleles in a specified environment, whereas typical sequencing approaches do not. Therefore, the additive genetic variance is a better measurement of evolvability for a given trait in a specified environment than can be obtained from typical polymorphism data. However, estimating QG variances is laborious, difficult, and subject to relatively high levels of measurement error.

At generation 42, the 20 populations studied here as well as the two environmentally specialized ancestral populations (*AS* and *AC,* Fig. 1) and their laboratory source population (*GA*) were pool-sequenced to survey genome-wide SNP diversity. For sites likely under differential ecological selection (and those linked to them), the pattern of within-population diversity was *Cad*, *Salt* < *Temp* < *Spatial*. However, sites likely to be neutral (or under uniform selection between environments) showed a different pattern: *Temp* < *Cad*, *Salt* < *Spatial*. Recent theory can explain these patterns. Environmental heterogeneity could result in balancing selection, which is expected to increase neutral diversity at sites closely linked to the targets of balancing selection (Hudson and Kaplan 1988; Charlesworth 2006). However, Barton (2000) showed that temporally fluctuating selection is expected to reduce neutral variation elsewhere across the genome because unlinked neutral sites experience increased genetic drift due to variance in fitness from the loci under fluctuating selection (see similar theoretical results by Gillespie 1997; Taylor 2013).

As we mentioned in the Results, we could not detect any significant relationships between molecular diversity estimates and QG variation at the population level, though this may be due to the high measurement error in QG estimates. The experiment was designed to examine patterns among regimes (rather than populations) and we have greater confidence in some of those patterns. Here, we qualitatively compare patterns at the regime level in the sequence data relative to those from the QG data. Some of the QG patterns reported here match the diversity of the "selected sites," whereas others match those of the "neural sites" in the sequence data. For male body mass in both environments (Fig. 3E and F), the patterns of variation are consistent with the ecologically selected sites: *Cad* < *Salt* < *Temp* < *Spatial*. Populations from heterogeneous regimes tend to express more genetic variance than populations from the homogeneous regimes, regardless of whether the latter are adapted to the assay environment or not. This pattern suggests that loci underlying variation in male body mass are under antagonistic selection between the two environments. However, the QG patterns are much weaker than those in the sequence data (e.g., in the QG data there are no significance differences between *Temp* and *Adapted* and only three of six comparisons are significant for *Spatial* and *Adapted*).

In contrast, the pattern for QG variance in female body mass assayed in salt (Fig. 3D) matched the diversity of "neutral sites": V_{sire} was lowest in the *Temp* regime, whereas it was highest, as expected, in the *Spatial*. This match to the neutral SNP pattern is unexpected because female body mass is almost certainly not neutral and we observed phenotypic differentiation populations from the two constant environments (Fig. 2, middle panels). Although we have no satisfying explanation for this result and it could simply be a result of measurement error as the relative ordering of V_{sire} for *Cad*, *Salt*, and *Temp* is far from certain (Fig. 3D). One alternative possibility is that, temporal heterogeneity (as opposed to spatial variation) might favor a buffering genotype (Kassen 2002; Condon et al. 2014) that reduces the average phenotypic effect size of variants (a). This reduction in phenotypic effect size would lead to the unexpectedly low sire variance in *Temp* populations even if polymorphism (pq) at selected sites is high because $V_A = 2pqa^2$. However, it seems unlikely a "buffering genotype" would evolve in approximately 45 generations because canalization evolves via indirect selection from how it affects the expression of other loci (Wagner et al. 1997; Ketola et al. 2014).

INSIGHTS INTO GENE ACTION OF DIFFERENTIALLY SELECTED LOCI

Both the sequence data and the QG data are suggestive that both EAS and CN forms of gene action occur. In the sequence data, the pattern of diversity for the sites under differential selection followed the pattern expected under EAS: *Cad*, *Salt* < *Temp* < *Spatial*. A similar pattern was observed for QG variation in male body mass. Although the major patterns of sequence diversity indicated EAS, some of the more subtle patterns relating sequence diversity to initial polymorphism levels suggested some loci were CN. For QG variances of survival and female mass assayed in cadmium, the rank order was *Cad* (*Adapated*) < *Temp* < *Spatial* < *Salt* (*Nonadapted*). This pattern of variation is most consistent with the prediction from the CN model with cadmium-selected alleles being neutral in salt. The pattern is also compatible with the observation that there are approximately 13% more sites that are putatively selected in cadmium, but neutral in salt than the reverse type of condition neutrality (7498 vs. 6620 by ad hoc categorization, Huang et al. 2014). Another potential piece of evidence is that the mean survival of genotypes from heterogeneous populations in cadmium is almost as high as those from the adapted *Cad* populations (Fig. 2, top left), suggesting the alleles adapted to cadmium are able to spread in heterogeneous populations, without being selected against in salt (neutral in salt). Of course, none of the loci involved may be strictly described by either the EAS or CN models of gene action as there are a variety of intermediate models that share features of both EAS and CN models that could explain these observations. Moreover, all arguments based on the rank ordering of treatments should be regarded as highly tentative because measurement error of QG variances makes our certainty in the ordering low.

COMPARISON IN PATTERNS OF VARIATION AMONG TRAITS

Qualitatively, some QG patterns are similar across all three traits. For all three traits in both environments V_{sire} in the *Cad* treatment is low; V_{sire} in the *Spatial* treatment tends to be high. Patterns for survival and female body size are nearly identical within each environment and differ in the same way between environments. Male body size differs more from the other two traits. In particular, the *Temp* treatment has reasonably high V_{sire} for male body size but has low V_{sire} for survival and female body size. Conversely, for the *Salt* treatment measured in cadmium V_{sire} is high for survival and female body size, but relatively low for male body size. The aberrant patterns with respect to male body size may be due to measurement error or be because this trait has a less direct connection to fitness than the others. As mentioned above, male body size is the trait least likely to be under strong directional selection, possibly because of a trade-off due to correlated selection on development time. (Because male flies develop more slowly than females, males are more likely to face the risk of failing to complete development before day 12 in the maintenance schedule when adults are chosen as parents for the next generation.) We see reasonably strong divergence between homogeneous populations in mean male body size, as expected with directional selection, but less so than for the other two traits (Fig. 2). Although this implies a history of directional selection during the course of adaptation, variation in body size within adapted populations may now be more heavily influenced by genes with antagonistic effects on development time. Even if directional selection is weaker (or stabilizing rather than directional), it is unclear why this would cause the particular set of observed differences.

Conclusion

For this set of population, the sequence data provide clues to interpreting the unexpected patterns observed in the QG data, whereas the QG variances offer more meaningful insight into the evolvability of traits in specific environments. Although the patterns in QG data were more ambiguous, both sequence and QG approaches were consistent with the basic predictions that environmental heterogeneity helps to maintain genetic variance in fitness and that spatial heterogeneity does this more effectively than temporal heterogeneity. However, some of the patterns vary between site types as well as among traits or environments. The measurement error in the QG data limits our confidence in these patterns, but, assuming the patterns are real, some of this variation may reflect differing forms of selection (EAS vs. CN) and/or differences in the relative importance of selection and drift. Although this experiment is only a single snapshot of genetic variance in populations evolving in artificial environments, it reveals a diversity of outcomes even under well-controlled conditions.

Although experimental evolution provides a means to test key principles under simpler conditions (Kawecki et al. 2012), the ultimate goal is to understand the maintenance of genetic variation within populations in complex natural environments (e.g., Schmidt and Conde 2006). This will likely require a careful combination of quantitative and population genomic approaches (Stinchcombe and Hoekstra 2007; Savolainen et al. 2013), but presently most such work has been limited to cases of spatial heterogeneity with relatively limited migration (e.g., Fournier-Level et al. 2011; Jones et al. 2012; Langley et al. 2012; Siepielski et al. 2013; but see one temporal heterogeneity case: Bergland et al. 2014). Hopefully, as we gain insights from artificial experiments and the power of genomic inference increases, we will be better prepared to tackle these problems in a broader set of natural systems.

ACKNOWLEDGMENTS

The authors thank H. Rundle and N. Sharp for helpful discussion on statistics. They are grateful to the editors and two anonymous reviewers

for helpful feedback on earlier versions of the manuscript. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

DATA ARCHIVING

The doi for our data is doi:10.5061/dryad.sr3g0.

LITERATURE CITED

Baldwin-Brown, J. G., A. D. Long, and K. R. Thornton. 2014. The power to detect quantitative trait loci using resequenced, experimentally evolved populations of diploid, sexual organisms. Mol. Biol. Evol. 31:1040–1055.

Bangham, J. 2002. Effects of body size, accessory gland and testis size on pre- and postcopulatory success in *Drosophila melanogaster*. Anim. Behav. 64:915–921.

Barton, N. H. 2000. Genetic hitchhiking. Philos. Trans. R. Soc. B Biol. Sci. 355:1553–1562.

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

Bergland, A. O., L. B. Emily, K. R. O'Brien, P. S. Schmidt, and D. A. Petrov. 2014. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scale in *Drosophila*. PLoS Genet 10:e1004775.

Bonduriansky, R., M. A. Mallet, D. Arbuthnott, V. Pawlowsky-Glahn, J. J. Egozcue, and H. D. Rundle. 2015. Differential effects of genetic vs. environmental quality in *Drosophila melanogaster* suggest multiple forms of condition dependence. Evol. Lett. 18:317–326.

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

Byrne, P. G., and W. R. Rice. 2006. Evidence for adaptive male mate choice in the fruit fly *Drosophila melanogaster*. Proc. R. Soc. B 273:917–922.

Charlesworth, B., and K. A. Hughes. 2000. The maintenance of genetic variation in life history traits. Pp. 369–391 *in* R. S. Singh, and C. B. Krimbas, eds. Evolutionary genetics from molecules to morphology. Cambridge Univ. Press, Cambridge, U.K.

Charlesworth, D. 2006. Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet 2:e64.

Condon, C., B. S. Cooper, S. Yeaman, and M. J. Angilletta, Jr. 2014. Temporal variation favors the evolution of generalists in experimental populations of *Drosophila melanogaster*. Evolution 68:720–728.

Dempster, E. R. 1955. Maintenance of genetic heterogeneity. Cold Spring Harb. Symp. Quant. Biol. 20:25–32.

Falconer, D. S., and T. F. C. Mackay. 1996. 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.

Fournier-Level, A., A. Korte, M. D. Cooper, M. Nordborg, J. Schmitt, and A. M. Wilczek. 2011. A map of local adaptation in *Arabidopsis thaliana*. Science 334:86–89.

Friberg, U., and G. Arnqvist. 2003. Fitness effects of female mate choice-preferred males are detrimental for *Drosophila melanogaster* females. J. Evol. Biol. 16:797–811.

Fry, J. D. 1996. The evolution of host specialization: are trade-offs overrated? Am. Nat. 148:S84–S107.

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

Garcia-Gonzalez, F., L. W. Simmons, J. L. Tomkins, J. S. Kotiaho, and J. P. Evans. 2012. Comparing evolvabilities: common errors surrounding the calculation and use of coefficients of additive genetic variation. Evolution 66:2341–2349.

Gillespie, J. H. 1997. Junk ain't what junk does: neutral alleles in a selected context. Gene 205:291–299.

Hadfield, J. D. 2010. MCMC methods for multi-response generalised linear mixed models: the MCMCglmm R Package. J. Stat. Softw. 33:1–22.

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.

Hansen, T. F., C. Pélabon, and D. Houle. 2011. Heritability is not evolvability. Evol. Biol. 38:258–277.

Hoffmann, A. A., and J. Merila. 1999. Heritable variation and evolution under favourable and unfavourable conditions. Trends Ecol. Evol. 14: 96–101.

Houle, D. 1992. Comparing evolvability and variability of quantitative traits. Genetics 130:195–204.

Houle, D., C. Pélabon, G. P. Wagner, and T. F. Hansen. 2011. Measurement and meaning in biology. Quart. Rev. Biol. 86:3–34.

Huang, Y., S. I. Wright, and A. F. Agrawal. 2014. Genome-wide patterns of genetic variation within and among alternative selective regimes. PLoS Genet 10:e1004527.

Hudson, R. R., and N. L. Kaplan. 1988. The coalescent process in models with selection and recombination. Genetics 120:831–840.

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

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.

Kassen, R. 2002. The experimental evolution of specialists, generalists, and the maintenance of diversity. J. Evol. Biol. 15:173–190.

Kawecki, T. J. 1994. Accumulation of deleterious mutations and the evolutionary cost of being a generalist. Am. Nat. 144:833–838.

Kawecki, T. J., R. E. Lenski, D. Ebert, B. Hollis, I. Olivieri, and M. C. Whitlock. 2012. Experimental evolution. Trends Ecol. Evol. 27:547–560.

Ketola, T., V. M. Kellermann, Volker Loeschcke, A. López-Sepulcre, and T. N. Kristensen. 2014. Does environmental robustness play a role in fluctuating environments? Evolution 68:587–594.

Kofler, R., and C. Schlötterer. 2014. A guide for the design of evolve and resequencing studies. Mol. Biol. Evol. 31:474–483.

Langley, C. H., K. Stevens, C. Cardeno, Y. C. G. Lee, D. R. Schrider, J. E. Pool, S. A. Langley, C. Suarez, R. B. Corbett-Detig, B. Kolaczkowski, et al. 2012. Genomic variation in natural populations of *Drosophila melanogaster*. Genetics 192:533–598.

Leffler, E. M., K. Bullaughey, D. R. Matute, W. K. Meyer, L. Ségurel, A. Venkat, P. Andolfatto, and M. Przeworski. 2012. Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol 10:e1001388.

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

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

Long, T. A. F., L. Rowe, and A. F. Agrawal. 2013. The effects of selective history and environmental heterogeneity on inbreeding depression in experimental populations of *Drosophila melanogaster*. Am. Nat. 181:532–544.

Lynch, M., and B. Walsh. 1997. Genetics and analysis of quantitative traits. Sinauer Associations, Sunderland, MA.

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

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

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.

Pitnick, S., and F. Garcia-Gonzalez. 2002. Harm to females increases with male body size in *Drosophila melanogaster*. Proc. R. Soc. B Biological Sciences 269:1821–1828.

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.

Prasad, N. G., S. Bedhomme, T. Day, and A. K. Chippindale. 2007. An evolutionary cost of separate genders revealed by male-limited evolution. Am. Nat. 169:29–37.

Quinn, G. P., and M. J. Keough. 2002. Experimental design and data analysis for biologist. Pp. 196–198. Cambridge Univ. Press, Cambridge, U.K.

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.

Robertson, A. 1959. Experimental design in the evaluation of genetic parameters. Biometrics 15:219–226.

Savolainen, O., M. Lascoux, and J. Merilä. 2013. Ecological genomics of local adaptation. Nat. Rev. Genet. 14:807–820.

Schmidt, P. S., and D. R. Conde. 2006. Environmental heterogeneity and the maintenance of genetic variation for reproductive diapause in *Drosophila melanogaster*. Evolution 60:1602–1611.

Sharp, N. P., and A. F. Agrawal. 2012. Evidence for elevated mutation rates in low-quality genotypes. Proc. Natl. Acad. Sci. 109:6142–6146.

Siepielski, A. M., K. M. Gotanda, M. B. Morrissey, S. E. Diamond, J. D. DiBattista, and S. M. Carlson. 2013. The spatial patterns of directional phenotypic selection. Ecol. Lett. 16:1382–1392.

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.

Stinchcombe, J. R., and H. E. Hoekstra. 2007. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity 100:158–170.

Taylor, J. E. 2013. Theoretical population biology. Theor. Popul. Biol. 87:34–50.

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.

Venail, P. A., O. Kaltz, I. Olivieri, T. Pommier, and N. Mouquet. 2011. Diversification in temporally heterogeneous environments: effect of the grain in experimental bacterial populations. J. Evol. Biol. 24:2485–2495.

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

Wagner, G. P., G. Booth, and H. Bagheri-Chaichian. 1997. A population genetic theory of canalization. Evolution 51:329–347.

Wallace, B. 1975. Hard and soft selection revisited. Evolution 29:465–473.

Whitlock, M. C. 1996. The red queen beats the jack-of-all-trades: the limitations on the evolution of phenotypic plasticity and niche breadth. Am. Nat. 148:S65–S77.

Yeaman, S., Y. Chen, and M. C. Whitlock. 2010. No effect of environmental heterogeneity on the maintenance of genetic variation in wing shape in *Drosophila melanogaster*. Evolution 64:3398–3408.

Associate Editor: I. Dworkin
Handling Editor: T. Lenormand

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1. Sire variance standardized by total variance for survival (A and B), female body mass (C and D), and male body mass (E and F) in cadmium (left) and salt (right) assay environment for each replicate of different selective histories.

Figure S2. Isire for survival (top, A and B), female body mass (middle, C and D), and male body mass (bottom, E and F) in cadmium (left) and salt (right) assay environment for each replicate of different selective regimes.

Table S1. The raw P-value (top line) and adjusted P-value using sequential Bonferroni correction across all 24 tests (bottom line) for the hypothesis testing of three traits: viability (V), female (F), and male (M) body mass.