doi:10.1111/evo.12309

SENSITIVITY OF THE DISTRIBUTION OF MUTATIONAL FITNESS EFFECTS TO ENVIRONMENT, GENETIC BACKGROUND, AND ADAPTEDNESS: A CASE STUDY WITH *DROSOPHILA*

Alethea D. Wang,[1] Nathaniel P. Sharp,[1] and Aneil F. Agrawal[1,2]

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

[2]*E-mail: a.agrawal@utoronto.ca*

Received May 3, 2013
Accepted October 22, 2013

Heterogeneity in the fitness effects of individual mutations has been found across different environmental and genetic contexts. Going beyond effects on individual mutations, how is the distribution of selective effects, *f*(*s*), altered by changes in genetic and environmental context? In this study, we examined changes in the major features of *f*(*s*) by estimating viability selection on 36 individual mutations in *Drosophila melanogaster* across two different environments in two different genetic backgrounds that were either adapted or nonadapted to the two test environments. Both environment and genetic background affected selection on individual mutations. However, the overall distribution *f*(*s*) appeared robust to changes in genetic background but both the mean, *E*(*s*), and the variance, *V*(*s*) were dependent on the environment. Between these two properties, *V*(*s*) was more sensitive to environmental change. Contrary to predictions of fitness landscape theory, the match between genetic background and assay environment (i.e., adaptedness) had little effect on *f*(*s*).

KEY WORDS: **Distribution of fitness effects, epistasis, mutation, selection, stress.**

A major property determining the evolutionary fate of a new mutation is its fitness effect. A mutation's fitness effect may depend on the exact context in which it arises. Many studies have shown that a mutation's effect on fitness depends on the environment in which it is assayed (e.g., Kishony and Leibler 2003; Jasnos et al. 2008; Hillenmeyer et al. 2008; Wang et al. 2009). Differences in the fitness effects of individual mutations have also been detected among genetic backgrounds (Remold and Lenski 2004; Wang et al. 2012). This suggests that the influence of environmental and genetic conditions on mutational fitness effects is pervasive. Yet, we lack a general understanding of the magnitude and consistency of changes caused by environmental and genetic conditions, and whether these condition effects can alter the overall distribution of mutational fitness effects.

Knowledge of the frequency distribution of fitness effects for new mutations, $f(s)$, is pertinent to a number of major questions in evolutionary genetics. These include predictions for levels of standing variation (Charlesworth and Hughes 2000) and inbreeding depression (Charlesworth and Charlesworth 1998), the evolution of sex and recombination (Keightley and Otto 2006; Hartfield et al. 2010), the nature of adaptation (Orr 1998), and extinction risks (Lynch et al. 1995a,1995b). Understanding how $f(s)$ changes across contexts thus has important theoretical and practical implications. Furthermore, selection can occur over multiple contexts within a single species and this too has important consequences. For example, autocorrelated temporal fluctuations in the environment can accelerate the rate of fitness loss through Muller's Ratchet (Wardlaw and Agrawal 2012). Spatial variation in the

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

intensity of selection alters the equilibrium frequency of deleterious alleles and the resulting mutation load, but only if these changes are positively correlated across loci (Roze 2012).

Empirical studies of changes in $f(s)$ typically involve estimation of its major properties—the mean, $E(s)$, and variance, $V(s)$—as we generally have limited power to obtain precise estimates for individual alleles with small fitness effects (Eyre-Walker and Keightley 2007). Consequently, we focus on how environmental and genetic factors affect these aspects of $f(s)$. Several hypotheses have been proposed to explain heterogeneity in these moments of $f(s)$. A persistent idea is that stressful contexts are more selective than benign ones, where "stress" is regarded as the reduction in absolute fitness relative to maximal fitness in a benign environment. A literature survey found the effects of stress on mean selection were inconsistent (Agrawal and Whitlock 2010). Some studies have found an increase in $E(s)$ (e.g., Remold and Lenski 2001; Szafraniec et al. 2001), but others have found no effect (e.g., MacLellan et al. 2012), or that stress ameliorated the average deleterious fitness effect of mutations (e.g., Kishony and Leibler 2003; Jasnos et al. 2008).

Using Fisher's fitness landscape model (1930), Martin and Lenormand (2006a) examined how $f(s)$ changes as a function of the level of adaptation of the genetic background to the environment. This conceptual framework emphasizes that the context for $f(s)$ is neither the genetic background nor the environment alone, but the two together, and the match between them. This model has been used to interpret empirical studies measuring changes in $E(s)$ or $V(s)$ across environments (e.g., Martin and Lenormand 2006b; Lalić et al. 2011; Vale et al. 2012). Environments where the baseline fitness is high are taken to represent situations of strong adaptation, whereas environments where the baseline fitness is low are thought to represent poor adaptation. A difficulty with most such comparisons is that any observed change in $E(s)$ or $V(s)$ could be attributed to the difference in adaptation between contexts or could be attributed to the difference in environments themselves. The theory is based on the level of adaptation within an environment but does not make predictions for how the fitness landscape and, consequently, $f(s)$ changes across environments.

Conceptually, $f(s)$ could be affected by the genetic background (G), the environment (E), and/or the match between them ("adaptedness," a type of G $\times$ E interaction). The most direct approach to examining the importance of these factors is to measure the fitness effect of mutations in several environments and in genetic backgrounds that are adapted to some of these test environments but not others. Remold and Lenski (2004) performed such a study in *Escherichia coli* and found that each of these three factors affected individual mutations to varying degrees. However, their data were not well suited to examining how these factors influenced the overall properties of $f(s)$. Here, we measured the fitness effects of 36 individual mutations in two environments in each of two *Drosophila melanogaster* populations, one adapted to each of the two test environments.

This experiment has several limitations that are worth noting at the outset. First, as in most experimental studies examining individual mutation effects, we do not use a random sample of spontaneous mutations. Rather, we measure the fitness effects of a random set of gene disruption mutations, which we would expect to have larger effects than spontaneous mutations, on average. However, our interest is not in the estimate of $E(s)$ per se but rather how it, and other aspects of $f(s)$, are affected by the context in which selection is measured. Second, we measure selection in only two environments and two genetic backgrounds. Any effect (or lack thereof) of environment, background, or adaptedness is, in a strict sense, limited to this specific combination of environments and backgrounds. Third, the two environments used here are not related to natural habitat variation in this species. Nonetheless, these environments are sufficient to examine the effects of adaptation, environment, and genetic background in terms of evaluating conceptual principles. Fourth, our measure of selection only measures egg-to-adult survivorship. Effects through adult fitness (e.g., siring success and female fecundity) could be large and might be affected by context differently than survivorship, though we have no reason to believe this is so. Moreover, juvenile and adult fitness effects of new mutations tend to be positively correlated (Keightley and Ohnishi 1998; Mallet et al. 2012). Alternatively, and in the spirit of measuring selection in arbitrarily defined environments, we can imagine that these populations are artificially maintained with no variance in adult reproductive success (i.e., all surviving parents contribute equally to the next generation) so that survivorship effects represent a complete measure of selection. Bearing these issues in mind, we view the study reported below as an imperfect but useful case study of how $f(s)$ is influenced by genetic background, environment, and the match between the two.

Methods

MUTATIONS

We examined the fitness effects of 36 individual gene disruption mutations (Table S1). They were selected from a larger collection of inserts, originally generated by Exelixis Corporation on a fully isogenic background bearing w^{1118} by transposon insertion mutagenesis (Thibault et al. 2004), and distributed by the Bloomington *Drosophila* Stock Center. These mutations occur in nonoverlapping genes randomly distributed across the X-chromosome, and all are marked by the mini-white gene (w^{+mC}) originating from the transposon construct used in their generation.

The X chromosome from the original isogenic progenitor strain, marked with w^{1118}, was used as the "wild-type" control in our fitness assays. w^{1118} is a recessive loss-of-function mutation

in the *white* gene, which decreases normal eye-color pigmentation levels by >99%, resulting in white eyes. As w^{+mC} is a cDNA sequence of the wild-type *white* gene, it restores eye pigmentation on the w^{1118} background. All 36 mutants, hereafter referred to as Ex(w^{+mC}), carry both w^{1118} and w^{+mC} genes having eyes that range from light orange to brown and are distinguishable from the w^{1118} white-eyed control. They have no other obvious phenotypic effects. We attempted to estimate the fitness effect of having a functional *white* gene by assaying the fitness effect of the actual wild-type *white* allele (w^+). We introgressed all 36 mutations and the w^+ allele into the w^{1118} isogenic background through 10 generations of serial backcrossing to homogenize the X chromosome among lines (i.e., to remove any additional mutational differences that may have accumulated on the X chromosome in individual mutant lines). Each generation of backcrossing involved ∼100 randomly sampled individuals from the w^{1118} population.

GENETIC BACKGROUNDS

Two *D. melanogaster* populations with different adaptive histories were used to create the genetic backgrounds on which mutational fitness effects were examined. These two populations were both derived from an outbred lab stock (SIM) originally collected from the Similkameen Valley, British Columbia, Canada, in September 2005 by S. Yeaman. The SIM stock was subsequently maintained in large population cages, first at the University of British Columbia and then in our lab, under constant conditions (25°C, 70% relative humidity, 12L:12D photoperiod) on standard cornmeal–sugar–yeast media. The two test populations used here, "Cad-Adapted" (*CA*) and "Salt-Adapted" (*SA*), were established separately in rearing environments in which the standard food media was supplemented with $CdCl_2$ or NaCl in July 2007 and October 2008, respectively. During the course of experimental adaptation, populations were composed of at least 1000 adults per generation. The amount of $CdCl_2$ or NaCl added to the food medium was progressively increased each generation, during which time the populations experienced fitness gains in their selected environment. Concentrations of the supplements reached 75 μg L^{-1} $CdCl_2$ and 8% NaCl, after which further increases were ceased, and the populations were maintained at these concentrations.

To create the two genetic background lineages in which mutational fitness effects were measured, 10 generations of backcrossing were conducted simultaneously between the w^{1118} control line and randomly sampled genomes (∼200) isolated from either the *CA* or *SA* population beginning in November 2010. The two resulting lineages, w^{1118}*; CA* and w^{1118}*; SA*, share the isogenic X chromosome from the w^{1118} control line and randomized autosomes derived from either the *CA* or *SA* populations, respectively. To avoid the possible accumulation of new divergent adaptive mutations on the X chromosome of these two lineages, they were created and maintained on standard cornmeal food medium. After establishment, the two genetic background lineages were expanded to population sizes of >1000 individuals each. To ensure that the autosomes of the two constructed populations retained their environment-specific adaptations (without losing the X from the w^{1118} progenitor stock), we used a specialized migration procedure each generation (Fig. S1).

In sum, the flies used in the fitness assays described below carry an X chromosome with or without an Ex(w^{+mC}) mutation that is otherwise isogenic. The autosomal background is genetically variable, reflecting the segregating genetic variation within the *CA* or *SA* populations.

ENVIRONMENTS

The two environments, cadmium and salt, used in our experiment involved the same chemical food supplements ($CdCl_2$ or NaCl) as the selective environments in which the *CA* and *SA* populations were maintained, but at lower concentrations (60 μg L^{-1} $CdCl_2$ and 4.7% NaCl). This was necessary as the higher concentrations were nearly lethal for the corresponding nonadapted population.

FITNESS ASSAYS

Fitness assays of mutational effects were conducted over 21 blocks beginning in June 2011 and ending in July 2012. In each block, one to three Ex(w^{+mC}) mutations were assayed (Table S1). Introgressions of individual mutations into the w^{1118}*; CA* and the w^{1118}*; SA* genetic backgrounds were conducted in parallel, via serial backcrossing (∼150 individuals per backcross) on standard cornmeal food media, in the three generations immediately prior to the start of the fitness assay. This created two lineages: Ex(w^{+mC})*; CA* and Ex(w^{+mC})*; SA*.

We measured the fitness effect of each mutation in four contexts (two genetic backgrounds in two environments): (1) *CA* in cadmium, (2) *CA* in salt, (3) *SA* in cadmium, and (4) *SA* in salt. To assay fitness (measured as egg-to-adult viability), heterozygous mutant Ex(w^{+mC})*; CA* or Ex(w^{+mC})*; SA* females were mated to hemizygous w^{1118}*; CA* or w^{1118}*; SA* males, respectively (Fig. 1). To create these parental individuals, we crossed hemizygous Ex(w^{+mC})*; CA* or Ex(w^{+mC})*; SA* males with females from the w^{1118}*; CA* or w^{1118}*; SA* lineages (∼600 individuals for each cross, in groups of three males × six females). All parents used for the fitness assays were raised on standard cornmeal–sugar–yeast media in 37 mL vials at moderate density. Four days following their emergence as adults, these parental flies were released into cages to lay eggs (three cages per genetic background lineage; ∼2000 flies per cage). Eggs were laid overnight on grape juice agar lay plates. From each of the two genetic background lineages, groups of 600 ± 100 eggs were transferred via egg washing (with phosphate buffered saline solution) into bottles of either the

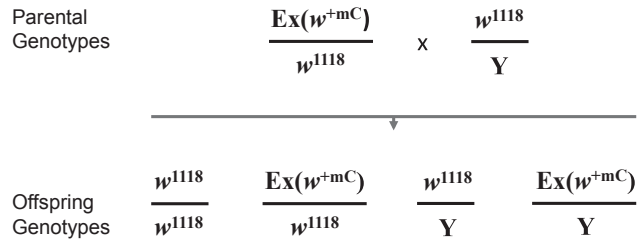

Figure 1. Fitness assay design used to measure the fitness effect of individual Ex(w^{+mC}) mutations in each of the four genetic background by environment contexts. Only the genotype with respect to the X and Y chromosomes is shown. Heterozygous mutant Ex(w^{+mC}); *CA* or Ex(w^{+mC}); *SA* females were mated to hemizygous w^{1118}; *CA* or w^{1118}; *SA* males ("parental genotypes"), and their offspring were collected to be raised in each of the two environments (cadmium or salt). Based on neutral fitness effects of the Ex(w_{+mC}) mutation, this cross is expected to produce four offspring genotypes in equal frequency (one-fourth each): homozygous w^{1118} females, heterozygous Ex(w^{+mC}) females, hemizygous w^{1118} males, and hemizygous Ex(w_{+mC}) males ("offspring genotypes"). Mutants carrying Ex(w^{+mC}) will have reddish eyes whereas flies homozygous (females) or hemizygous (males) for w^{1118} have white eyes.

cadmium or salt food environment (day 0). A total of 40–50 replicate bottles were set up for each of the four contexts over 3 days.

Replicate bottles from all four treatments were mixed together in trays and offspring were allowed to develop under standard culture conditions (25°C, 70% relative humidity, on a 12L:12D photoperiod). After 14 days, the number of surviving offspring of each genotype was determined by phenotypic scoring. The offspring of each cross are of four expected genotypes: homozygous w^{1118} females, heterozygous Ex(w^{+mC}) females, hemizygous w^{1118} males, and hemizygous Ex(w^{+mC}) males (Fig. 1).

Heterozygous Ex(w^{+mC}) females were more difficult to identify than hemizygous Ex(w^{+mC}) males. Specifically, heterozygous Ex(w^{+mC}) females for some loci have very faintly colored eyes and some individuals were likely mis-scored as w^{1118} individuals. Such a bias in scoring would result in an overestimate of selection against mutations in females. The difficulty in scoring heterozygous Ex(w^{+mC}) females varied among the 36 Ex(w^{+mC}) mutations, because of differences in their eye-color phenotypes, but did not seem to vary among genetic background and environment contexts. For this reason, we suspect that the degree of bias in estimating female selection was variable among loci, but reasonably constant within a locus among contexts.

Although we estimated selection in both males and females, these data are not suitable for formal comparisons of selection between the sexes. First, selection against mutants is measured in the hemizygous state in males but in the heterozygous state in females. Second, there is likely substantial upward bias in estimating selection in females but not in males.

FITNESS OF ALTERNATIVE GENETIC BACKGROUNDS

Prior to our fitness assays to measure mutational effects, we measured the fitness of the two genetic background lineages, w^{1118}; *CA* and w^{1118}; *SA*, in each of the two food environments. The design of this assay was identical to that of the fitness assays described earlier, but because we were also interested in measuring the fitness effect of the true wild-type *white* gene (w^+), the control assay involved mating heterozygous w^+; *CA* or w^+; *SA* females with hemizygous w^{1118}; *CA* or w^{1118}; *SA*, respectively. In addition, a larger total number of replicate bottles (123–130) were set up for each of the four contexts over the course of 4 days.

Because all of the Exelixis gene disruption mutations carry the *mini-white* construct, it is possible that differences in mean selection between contexts could be due to context dependent fitness effects of a functional *white* protein rather than selection on the disrupted target genes. To assess selection on having a functional *white* protein, we estimated the fitness effect of w^+ relative to the w^{1118} reference in each of the four contexts (see Fig. S2 for details). Estimates of selection on w^+ were significantly different across contexts. However, the patterns were different from the results for average selection for the 36 Ex(w^{+mC}) mutations. It is thus unlikely that differences in selection on the shared *mini-white* element underlie the context-dependent changes in $E(s)$ that we observed in our experiment.

STATISTICAL ANALYSIS

Our objective was to examine the fitness effects of the Ex(w^{+mC}) mutations across the four genetic background by environment contexts. As all 36 Ex(w^{+mC}) mutations are located on the X chromosome, fitness effects were calculated separately as hemizygous selection in males, s_M, and heterozygous selection in females, s_F. Within each sex, the expected frequency of Ex(w^{+mC}) individuals among the surviving offspring is $f_{Ex}' = f_{Ex}(1 - s)/(1 - f_{Ex}s)$, where $f_{Ex} = 1/2$ is the initial frequency of Ex(w^{+mC}) genotypes. Estimates of the selection coefficient were obtained by rearranging this equation as $\hat{s} = (1 - 2f_{Ex}')/(1 - f_{Ex}')$ and using the observed frequency of Ex(w^{+mC}) genotypes after summing individuals across replicates for f_{Ex}'. We used resampling of the replicate bottles to produce a set of 10^5 bootstrap estimates of $\widehat{s_M}$ and $\widehat{s_F}$ for each of the four contexts. From these, we obtained bootstrap mean estimates of $\widehat{s_M}$ and $\widehat{s_F}$ as well as 95% confidence intervals (CIs) for each gene in each context that were used in further analyses. Replicate bottles with very low survivorship (fewer then 10 total offspring) were omitted from the estimation procedure. The variance among bootstrap estimates was used as an approximation of the error variance in the likelihood analysis (below).

Using the estimates of $\widehat{s_M}$ and $\widehat{s_F}$ for all 36 mutations, we examined context-dependent changes in the distribution of fitness effects $f(s)$ for each sex separately. We used two approaches; the first was based on simple standard statistical tests (e.g., ANOVA, Levene's tests) and the second was based on maximum likelihood analyses of the full distributions. First, we used a two-way analysis of variance (ANOVA—"aov" function in *R*) to test for difference in the average mutational fitness effect, $E(s)$, across contexts, using the genetic background, the environment, and their interaction as fixed effects and gene as a random effect. Next, we tested for differences in the variance of selection $V(s)$ across contexts using Levene's tests. For testing specific effects on the variance, we contrasted pairs of contexts using Levene's tests and then applied the Z-transform test (Whitlock 2005) to test the significance of combined probabilities across separate but parallel contrasts. For example, to test whether the environment affected $V(s)$, we performed two separate Levene's tests, one contrasting the variance in cadmium to the variance in salt in the *CA* genetic background and a second test contrasting variances across environments in the *SA* genetic background. In addition to reporting the outcome of these individual tests, we also report the combined probability based on the Z-transform test.

To further examine changes in the overall distribution of fitness effects across contexts, we also performed a maximum likelihood analysis. Compared to the classical tests described above, this analysis imposed fewer distributional assumptions and allowed us to explicitly incorporate differences in measurement error between assay environments. For each sex separately, bootstrap estimates of $\hat{s}$ were fit to a likelihood model comprised of two parts, representing (i) the probability that a mutation in context C will have selection coefficient s, and (ii) the probability that the estimate of selection on gene g will differ from the true value of selection on this gene by the amount $\varepsilon_{C,g} = \hat{s}_{C,g} - s_{C,g}$ because of measurement error. The log-likelihood of the data is given by

$$\log(L[parameters|\{\hat{s}_{1,1}, \hat{s}_{2,1}, \ldots, \hat{s}_{4,36}\}) = \sum_{C=1}^{4}\sum_{g=1}^{36} \log\left[\int f_C(s|\bar{s}_C, \sigma_C^2, \alpha_C) N(\varepsilon_{C,g}|0, V_{Err,C,g}) ds\right],$$

where the outer most summation is over the four genetic background × environment contexts, the next summation is over the 36 genes, and the integration is over all possible values of s within context C.

For context C, the distribution of fitness effects, $f_C(s)$, is assumed to follow a displaced γ distribution that has mean $\bar{s}_C$, variance $\sigma_{s_C}^2$, and a "shape" α_C. Typically, a γ distribution is parameterized by two parameters, a and b, resulting in a mean of $m = ab$ and a variance of $v = ab^2$. In our parameterization of the displaced γ, we used a typical γ with $a = \alpha^2/{\sigma_s}^2$ and $b = {\sigma_s}^2/\alpha$ but then shift the distribution to have a mean of $\bar{s}$. In other words, let $s = z - ab + \bar{s}$, where $z \sim \Gamma(a, b)$. For computational simplicity, we used a discrete approximation for the shifted γ distribution that was characterized by 501 equally weighted values based on quantiles of the continuous distribution. $N(\varepsilon_{C,g})$ is the probability density function of the measurement error, $\varepsilon_{C,g}$. We assumed measurement error was normally distributed with mean zero and variance $V_{Err,C,g}$; the bootstrap sampling variance of $\hat{s}_{C,g}$ for each mutation in each context was used as the measurement error variance.

Because there are four contexts, the "full model" involves 12 separate parameters (3 parameters per distribution × 4 contexts). In contrast, our "base model" assumes the distribution is the same across all four contexts, requiring only three parameters. A series of models intermediate between these two extremes were considered. These additional models were organized into three sets based on our interest in examining the importance of the three main factors of interest: genetic background, environment, and adaptedness. Within each set, we conducted paired contrasts, where we assumed that the distributions could differ *between* the contexts of interest with respect to one, two, or all three of the parameters ($\bar{s}$, σ_s^2, and α). However, we constrained the distributions to be the same *within* the context of interest. For example, when comparing between genetic backgrounds, we allowed the distributions to differ between *CA* and *SA* backgrounds, but assumed that distribution of s was the same across environments within each genetic background (note the assumption is that the distribution is the same across environments, not that individual mutations have the same effect). Seven models (Table 1) were examined for each of the three paired contrast sets. Including the "full" and "base" models, this is a total of 23 models. These models were compared using log-likelihood and AIC scores.

A multistep approach was used to ensure we found the global maximum likelihood values rather than local maxima. This was done with the optimization function *optim* in *R* (R Development Core Team 2012), using the Nelder–Mead algorithm, followed by more refined optimization using BFGS. For each model, we ran this optimization procedure 15 times, each time starting with different random initial parameter values.

To quantify the frequency of environment, genetic background, and G × E interaction effects at the level of individual genes, we examined allele-specific fitness effects across contexts for each of the 36 Ex($w^{+\mathrm{mC}}$) mutations (Fig. S4). Using a generalized linear model (GLM—"*glm*" function in *R*), the frequency of mutant offspring (egg-to-adult survival) was fit as a function of genetic background, environment, and their interaction. The response variable in the GLM was a matched pair of counts (mutant offspring:nonmutant offspring) that was analyzed as proportion data using a quasi-binomial error structure. As phenotypic

Table 1. Models used in maximum likelihood analysis. In each context *C*, the distribution of fitness effects is modeled as a displaced γ distribution with mean ($\bar{s}_C$), variance (σ_C^2), and shape (α_C). In the Base model, the distribution is constrained to be the same across all four contexts, such that there is only one value each of $\bar{s}$, σ^2, and α being estimated (three parameters total). In contrast, in the Full model, each of the four distributions is free to take different values of $\bar{s}$, σ^2, and α, such that there are 12 parameters in total. We also considered a series of models (Models 1–7) in which the four contexts were divided into two pairs. The two contexts within each pair were constrained to have the same distributions but distributions could vary between the two pairs. Models 1–7 differ in the type and degree of constraint assumed between pairs. The division of contexts into pairs was done in three different ways so that the two pairs differed by environment, by genetic background, or by adaptedness (i.e., whether there was a match between environment and genetic background). Models 1–7 were repeated for each of these alternative pairings.

		Shared constraints			
Model		$\bar{s}_C$	σ_C^2	α_C	Parameters
Base model		Yes	Yes	Yes	3
Paired contrast models	Model 1	No	Yes	Yes	4
	Model 2	Yes	No	Yes	4
	Model 3	Yes	Yes	No	4
	Model 4	No	No	Yes	5
	Model 5	No	Yes	No	5
	Model 6	Yes	No	No	5
	Model 7	No	No	No	6
Full model		No	No	No	12

scoring of the offspring emerging from our fitness assays involved multiple experimenters ("Scorers"), a second GLM was run for each of the mutations that included main and interaction effects of Scorer. If none of the Scorer terms in the second GLM was significant at the $P < 0.05$ level then we reported the results from the first GLM. In almost all cases, even if one or more of the Scorer terms were significant, results from the second GLM were qualitatively similar to the conclusions of the first model. Significance of the main and interaction effects were evaluated using analysis of deviance ("anova" function in *R*) of the GLM parameter fits. Summaries of the number of individual mutations that show dependence on genetic background, environment, and their interaction were compiled.

Results

We confirmed that within each environment, the adapted population had higher egg-to-adult survival than the nonadapted population (Fig. 2). The number of surviving *CA* flies was 29% higher than the number of *SA* flies in the cadmium environment ($t = 10.06$, df $= 253$, $P = 10^{-3}$). Similarly, the number of surviving *SA* flies was 14% higher than the number of *CA* flies in the salt environment ($t = 4.33$, df $= 251$, $P < 10^{-3}$). These results indicate that the lineage with a selective history in a given environment is closer to the fitness optimum of that environment than the other lineage.

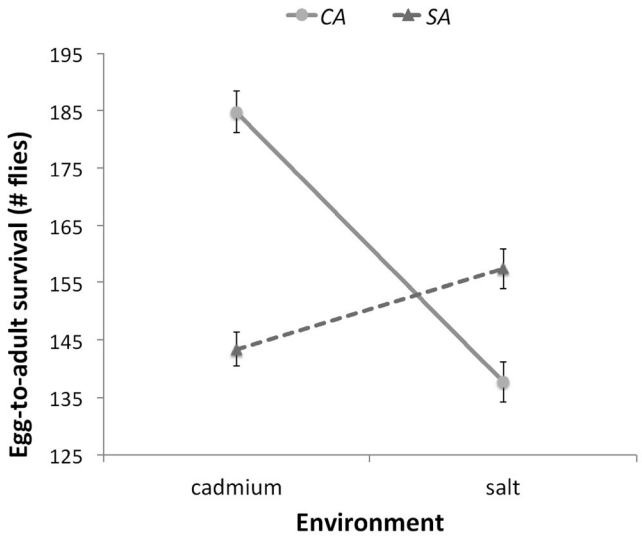

Figure 2. Egg-to-adult survival ± SE (measured based on the total number of flies emerging) in each of the two environments, cadmium and salt, for each of the two genetic background lineages, *CA* and *SA*, in the absence of Ex(w^{+mC}) mutations.

CONTEXT-DEPENDENT CHANGES IN THE PROPERTIES OF *f(s)*

Fitness effects were estimated separately as hemizygous selection in males, $\widehat{s_M}$, and heterozygous selection in females, $\widehat{s_F}$; the distributions are shown in Figure 3. Most mutations are deleterious and of small effect, and all distributions are right-skewed. Several mutations were estimated to have negative values of *s*, suggesting they may have beneficial fitness effects but all had 95% CIs that overlap 0 (Table S2). The confidence intervals on individual estimates are large so the lack of statistically significant beneficial effects should not be regarded as strong evidence against beneficial effects. In addition, likelihood models in which the distributions were constrained to have no beneficial effects were not significantly worse (γ vs. displaced γ; not shown) but this likely reflects a lack of power. We did not formally compare mean selection between the sexes (i.e., $E(\widehat{s_F})$ vs. $E(\widehat{s_M})$) because (i) selection in females is on heterozygous effects whereas selection in males is on hemizygous effects and (ii) mis-scoring of heterozygous mutant females as wild types (see Methods) likely caused an upward bias in the estimation of $\widehat{s_F}$ but there was no such bias in $\widehat{s_M}$.

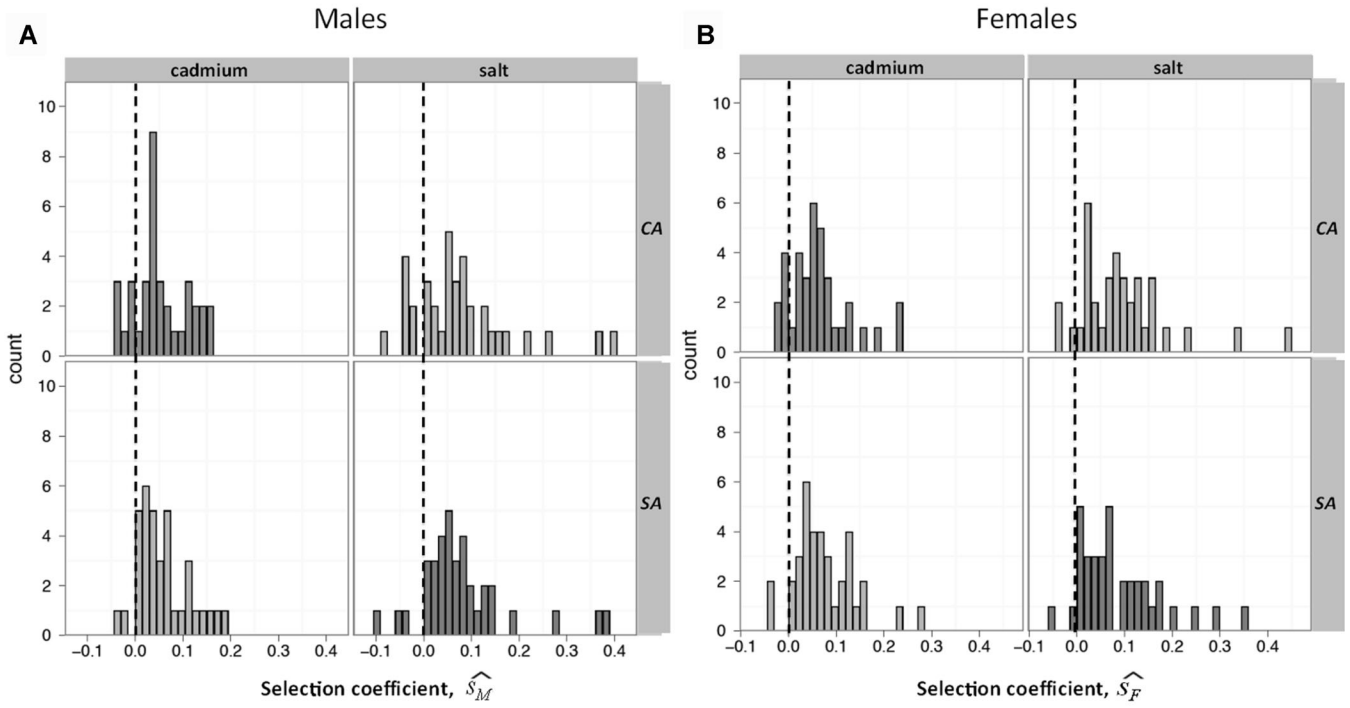

Figure 3. **Histograms of the distribution of fitness effects in (A) males and (B) females in each of four genetic background by environment contexts. The dashed black line in each panel marks where selection is 0 (i.e., mutant fitness effects are neutral). Positive values of the selection coefficient indicate deleterious fitness effects. Darkly (lightly) shaded histograms indicate contexts where there is a match (mismatch) between environment and genetic background.**

Table 2. **Pearson correlation coefficients comparing mutational fitness effects: (1) between male and female fitness effects within each of the four genetic background by environment contexts (diagonal cells), (2) *between* contexts for male effects (above the diagonal), and (3) *between* contexts for female effects (below the diagonal). An asterisk (*) indicates that the correlation is significant at $P < 0.05$.**

Genetic background/ environment	Genetic background/environment			
	CA/ cadmium	*SA*/ cadmium	*CA*/ salt	*SA*/ salt
CA/cadmium	0.363*	0.144 ♂	0.319 ♂	0.198 ♂
SA/cadmium	0.666* ♀	0.411*	0.139 ♂	0.262 ♂
CA/salt	0.686* ♀	0.637* ♀	0.330*	0.787* ♂
SA/salt	0.694* ♀	0.663* ♀	0.679* ♀	0.311

Averaging estimates across all four contexts, selection estimates in males and females ($\widehat{s_M}$ and $\widehat{s_F}$) were positively correlated ($r = 0.35$, $P = 0.04$, df = 34). Positive correlations between hemizygous and heterozygous fitness effects were also observed *within* each of the four contexts (significant for 3 of 4, diagonal elements of Table 2). This suggests that mutational fitness effects are generally concordant between the sexes. In addition, correlations *between* each of the contexts were calculated separately within each sex (off-diagonal elements of Table 2, also see Fig. S3). These values ranged from $r = 0.14$ to 0.79, and were significant in 7 of 12 cases. The cross-context correlations in females tended to be higher than in males. We suspect this is an artifact of the scoring bias that occurs in females (but not males); the scoring bias creates an artificial correlation because it is worse for some mutations than others (i.e., upwardly biasing $\widehat{s_F}$ more for some mutations than others) in a context-independent manner.

Analysis of variance was used to examine changes in the average fitness effect of mutations across contexts (Fig. 4). Mutations were deleterious on average in all four contexts for both sexes, but in neither sex was the average strength of selection significantly affected by genetic background (males $F = 0.20$, $P = 0.66$; females $F = 0.07$, $P = 0.80$) or G × E (males $F < 10^{-3}$, $P = 0.98$; females $F = 0.49$, $P = 0.49$). However, there was a significant effect of environment on the average mutational fitness effect in both sexes (males $F = 4.65$, $P = 0.03$; females $F = 6.78$, $P = 0.01$). $E(s)$ was 36% higher in the salt environment (on average across genetic backgrounds and sexes).

Changes in the variance of fitness effects, $V(s)$, were compared using Levene's homogeneity of variance tests. There is a clear effect of environment, which is significant for males, with $V(s_M)$ being ~2.5-fold higher in salt than cadmium, averaging

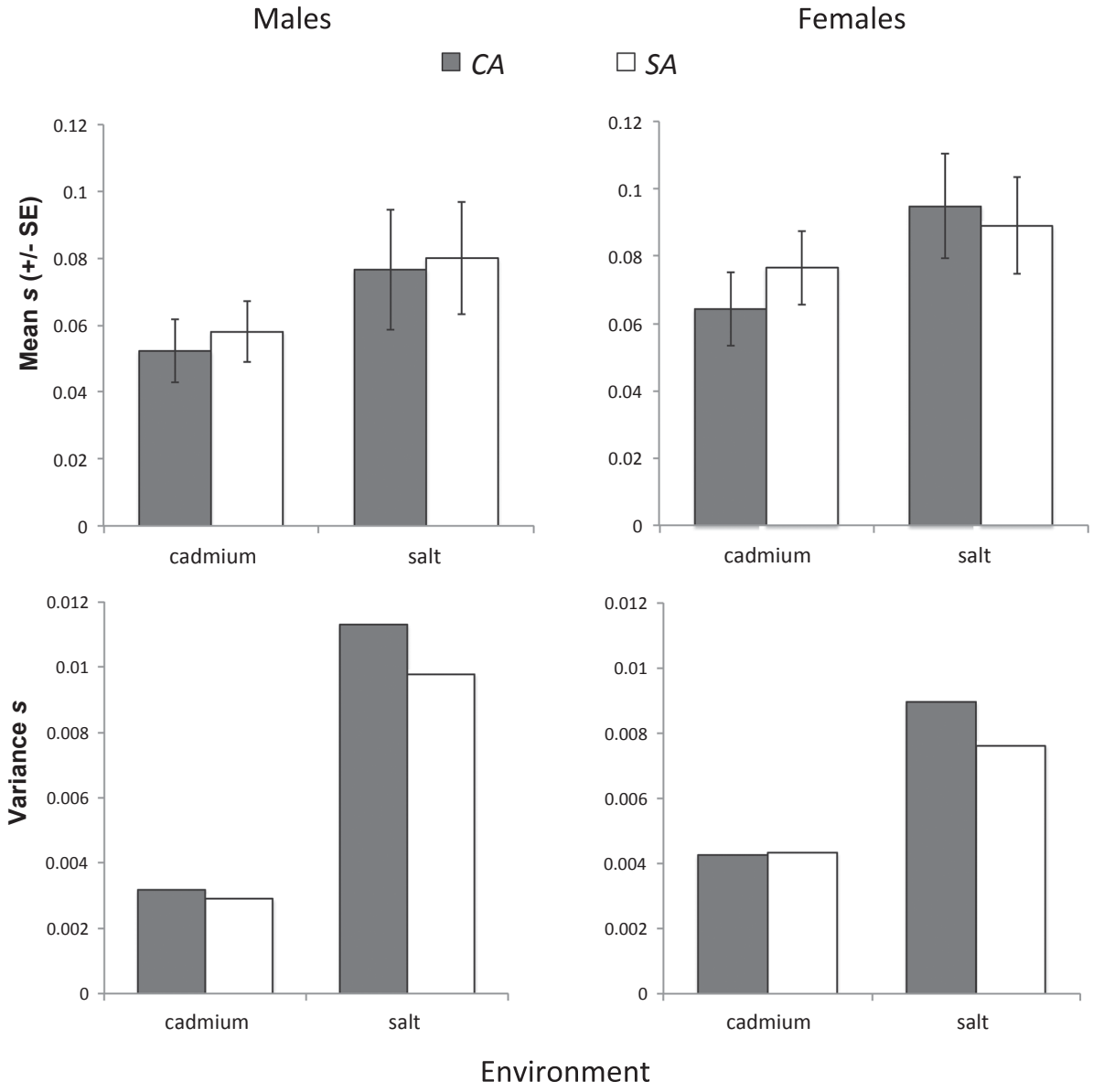

Figure 4. **Mean ± SE (top) and variance (bottom) of $\hat{s}$ for males (left) and females (right) in each of the four genetic background by environment contexts. Means ± SE: males (from left to right as shown) 0.051 ± 0.009, 0.057 ± 0.009, 0.075 ± 0.018, 0.081± 0.017; females 0.063 ± 0.017, 0.075 ± 0.011, 0.092 ± 0.016, 0.087 ± 0.015. Variances: males 0.0031, 0.0029, 0.0112, 0.0099; females 0.0041, 0.0042, 0.0087, 0.0077.**

across genetic backgrounds (Table 3A and Fig. 4). For females, $V(s_F)$ was about twice as high in salt than cadmium, on average. Considering the tests from both sexes together, there is strong evidence that the variance in selection among genes is significantly different across environments ($Z = -2.95$, $P = 0.003$). Analogous sets of comparisons were performed to evaluate whether genetic background or adaptedness affected the variance in selection. Neither of these factors appeared to have a strong effect (Table 3B, C and Fig. 4).

As a fraction of the variance in selection estimates arises from the measurement error, the significant difference observed in $V(s)$ across environments could be influenced by disparities in measurement error. Comparison of the bootstrap sampling variances shows that there was indeed a significant difference between the cadmium and salt environments (males $t = 4.16$, df $= 34$, $P < 10^{-3}$, females $t = 5.81$, df $= 34$, $P < 10^{-5}$). However, the magnitude of the difference observed in sampling variance across environments (males: 0.0005, females: 0.0006) was an order of magnitude smaller than the observed overall differences in $V(s_M)$ and $V(s_F)$ (males 0.0075, females 0.0040). This suggests that measurement error alone cannot account for the differences in $V(s)$ we observed.

Table 3. **Summary of the results from the Levene's tests conducted to assess the effect of (A) environment, (B) genetic background, and (C) adaptedness (i.e., the match between background and environment) on *V*(*s*) for each sex. For (A), each contrast is between environments for a single genetic background. For (B), the contrast is between genetic backgrounds. In the first contrast, each background is assessed in the environment to which it is adapted; in the second contrast, each background is assessed in its nonadapted environment. For (C), the contrast is between the adapted and nonadapted background, assessed in a single environment. The *Z*-transform is a combined probability test used to pool results across the parallel contrasts within each of the main contexts of interest.**

	Males		Females	
Between context contrasts	$F_{1,70}$	P	$F_{1,70}$	P
(A) Environment				
CA in cadmium vs. salt	3.98	0.05	1.48	0.23
SA in cadmium vs. salt	2.69	0.11	1.28	0.26
	Combined: $Z = -2.53$; $P = 0.011$		Combined: $Z = -1.65$; $P = 0.10$	
(B) Genetic background				
CA in cadmium vs. *SA* in salt	2.26	0.14	1.47	0.23
SA in cadmium vs. *CA* in salt	4.56	0.04	1.30	0.26
	Combined: $Z = 0.43$; $P = 0.67$		Combined: $Z = -0.05$; $P = 0.96$	
(C) Adaptedness				
CA vs. *SA* in cadmium	0.04	0.84	0.01	0.93
SA vs. *CA* in salt	0.18	0.68	< 0.01	0.95
	Combined: $Z = -0.44$; $P = 0.66$		Combined: $Z = -0.02$; $P = 0.98$	

Results from our maximum likelihood analysis corroborated those of the simpler statistical tests described earlier. This analysis was performed separately for males and females, as well as considering estimates of selection in both sexes. Details of the maximum likelihood results are given in Table 4. The suite of models tested (Table 1) was designed to allow inference about changes in $f(s)$ across the four contexts in our study. Our most constrained model ("Base" model) assumes there is a single distribution of mutational fitness effects that applies across all contexts. At the other extreme, the "Full" model assumes that there is a separate distribution for each context. Each of the remaining models fell into one of three categories based on whether it was structured to evaluate differences in $f(s)$ due to genetic background, environment, or the match between genetic background and environment (adaptedness). In Table 4, we present results for only the best of the seven models fit within each category (see Table S3 for a more detailed list of model results), as well as the Base and Full models.

For both sexes, allowing for different distributions among contexts improved the Akaike Information Criterion (AIC) score relative to the Base model. For males, the best model was that which assumed the mean and variance of the distribution differed across environments ("By environment—Model 4"). In females, however, estimates of fitness effects were best fit by the model allowing the shape of the distribution to vary depending on the match between genetic background and environment ("By adaptedness—Model 3"). When considering fitness effects in both males and females, the most strongly supported model was once again that which allows the mean and variance in the distribution of fitness effects to change across environments ("By environment —Model 4").

Although the environment seemed to be more important than other factors when examining changes in the mean and variance of the entire distribution, this was not the case when we considered effects on individual genes. We found similar numbers of cases where the fitness effects of individual genes varied with differences in genetic background, environment, or the G × E interaction (Fig. S4). Summaries of the number of mutations having fitness effects dependent on each of these three factors is given in Table 5. Nineteen out of the 36 Ex(w^{+mC}) mutations tested had male fitness effects that varied significantly with at least one of the three factors. Similarly, 18 of the 36 had female fitness effects that varied significantly with at least one of the three factors. In four of the mutations in males, and six of them in females, fitness effects were significantly changed by more than one factor.

Discussion

Here we measured the distribution of selective effects, $f(s)$, in two differentially adapted genetic backgrounds, assayed in reciprocal environments. Our goal was to assess if features of $f(s)$ were strongly dependent on either the environment or genetic background, or if $f(s)$ was more heavily influenced by the degree of adaptation. Our results indicate that the environment had the largest effect on $f(s)$; mean selection was stronger and the variance in selection was greater in the salt environment, regardless of genetic background.

Table 4. Results of maximum likelihood analyses for (A) fitness effects in males, (B) fitness effects in females, and (C) both sexes. The Base and Full model represent the fully constrained and the unconstrained models, respectively. A series of models were run contrasting distributions by (i) environment, (ii) genetic background, or (iii) adaptedness. Within each of these three categories, seven models (see models 1–7 in Table 1) were examined and the one with the best AIC score is shown here. The columns denoted as UP list the unconstrained parameters (one or more of $\bar{s}$, σ^2, α) that were allowed to vary across the context of interest (e.g., environment) and the associated model number. The columns denoted as NP give the number of parameters in the model. The best of the 23 models (based on AIC) within each column is highlighted in bold. Maximum likelihood analyses were performed separately for males and females. The results reported in (C) were obtained by summing log-likelihoods of separate analyses performed on male and female data for models with the same structure (i.e., both sexes have the same set of free parameters but the parameters take different values for the two sexes).

	(A) Males				(B) Females				(C) Both			
Model type	UP	NP	Log-likelihood	AIC	UP	# Par.	Log-likelihood	AIC	UP	NP	Log-likelihood	AIC
Base model	None	3	183.2	−360.3	None	3	186.9	−367.8	None	6	370.0	−728.1
By environment	**$\bar{s}$, σ^2 Model 4**	**5**	**187. 1**	**−364.2**	$\bar{s}$, σ^2 Model 4	5	188.8	−367.7	**$\bar{s}$, σ^2 Model 4**	**10**	**375.9**	**−731.9**
By Background	$\bar{s}$ Model 1	4	183.8	−359.6	α Model 3	4	187.7	−367.4	α Model 3	8	371.4	−726.7
By adaptedness	σ^2 Model 2	4	183.2	−358.4	**α Model 3**	**4**	**188.6**	**−369.2**	α Model 3	8	371.8	−727.5
Full model	All	12	188.1	−352.2	All	12	191.0	−358.0	All	24	379.1	−710.3

We would like to understand how and why $f(s)$ changes across contexts. The most well developed theoretical studies of $f(s)$ are based on fitness landscape models (i.e., Fisher's geometric model; Martin and Lenormand 2006a; Chevin et al. 2010). These studies typically assume that fitness is a Gaussian function of an unmeasured multivariate phenotype. Martin and Lenormand (2006a) found that, in the Gaussian model, $f(s)$ can be approximated by a displaced γ distribution. The mean of $f(s)$ depends on the strength of stabilizing selection in each phenotypic dimension (**S**, sensu Martin and Lenormand 2006a), the mutational input along each dimension (**M**), and the correspondence between the two. Curiously, the mean of $f(s)$ is predicted to be unaffected by a displacement of the population from the optimum. This prediction arises because the curvature of the Gaussian function is constant on the log fitness scale. Unlike the mean, the variance in s is predicted to increase with the distance from the optimum. This latter result applies beyond the Gaussian to any concave fitness function and is intuitive with respect to the sign of selection; all mutations are deleterious for a population at the optimum, whereas an increasing fraction of mutations will be beneficial as a population is displaced further from this point. Although it has often been viewed as merely a heuristic model, the Gaussian landscape model has had a number of quantitative successes with empirical data. For example, it has been used to predict the distribution of epistatic effects based on the distribution of single mutation effects (Martin et al. 2007). Trindade et al. (2012) used the Gaussian landscape model to accurately predict the maximum attainable fitness for *E. coli* in two of three environments based on measures of $f(s)$ in nonoptimal genetic backgrounds in each environment.

Although the landscape model of Martin and Lenormand (2006a) is best suited to describing $f(s)$ within a single Gaussian environment, it has often been used to predict changes in $f(s)$ across environments. This application of the model should be done with caution for two reasons. First, there is no reason to believe that all landscapes are Gaussian; even if the Gaussian landscape works as a good approximation for one environment, it may not for another. Although many landscapes may be Gaussian-like close to their optimums, we are often interested in the case when a population experiences a new environment where it will not be close to the optimum and, thus, the Gaussian approximation may fail. Second, even if Gaussian approximations are reasonable in multiple environments, each environment may be described by a different Gaussian function (i.e., strength of selection around each environment's optimum may differ in one or more dimensions, $\mathbf{S}_A \neq \mathbf{S}_B$). If phenotypic selection **S** changes between environments, so will the moments of $f(s)$, including the mean, even if the fitness surface remains Gaussian.

The typical application of the predictions of the model of Martin and Lenormand (2006a) to changes across environments implicitly assumes that a single Gaussian function describes

Table 5. Gene-specific fitness effects in (A) males and (B) females. Each row summarizes the number of mutations having fitness effects that varied significantly with a given factor. The last row summarizes the number of mutations that had fitness effects in which more than one of the three terms was significant. Significance levels are: (ns) $P > 0.1$; (.) $0.05 < P < 0.1$; (*) $0.01 < P < 0.05$; (**) $0.001 < P < 0.01$; (***) $P < 0.001$. Analyses for individual genes are shown in Figure S4.

	ns	.	*	**	***	$P < 0.05$	$P < 0.1$
(A) Males							
Environment	24	6	0	3	3	6	12
Background	25	1	3	3	4	10	11
G × E	30	3	2	1	0	3	6
Multiple						4	
(B) Females							
Environment	24	4	3	1	4	8	12
Background	27	5	3	1	0	4	9
G × E	27	3	5	0	1	6	9
Multiple						6	

selection in different environments with only the phenotypic optimum changing (i.e., no change in the shape of the surface). In this case, neither environment nor genetic background is intrinsically a strong determinant of $f(s)$; only the position of the genetic background relative to the phenotypic optimum is relevant. Specifically, $f(s)$ is determined by the fitness difference between the genetic background and the optimum, s_o. In this model, it does not matter if the fitness difference s_o is created by changing the genetic background (i.e., choosing a genetic background further from the optimum) or changing the environment (i.e., moving the optimum further from the genetic background). From this set of assumptions, one can interpret changes in $f(s)$ across environments as equivalent to changes in $f(s)$ across genetic backgrounds.

Martin and Lenormand (2006b) surveyed data from nine mutation accumulation experiments for which fitness assays had been performed in multiple environments, with at least one "benign" and one "stressful" environment in each experiment. They found changes in the mean tended to be relatively small and bidirectional with respect to stress. Changes in the variance tended to be comparatively larger and more consistent in direction, with higher variance occurring in more stressful environments. They concluded these patterns were consistent with a simple model of Gaussian fitness surfaces of constant shape (but differing optimums) across environments. From this perspective, a stressful environment is interpreted as one in which the reference genetic background is further from the optimum than it is in the benign environment but the fitness surface is otherwise identical in shape.

Our results provide an opportunity to examine some predictions of the landscape model. In its most extreme form (i.e., Gaussian functions of constant shape), the landscape model predicts that mean selection should be constant across environments and across genetic backgrounds. In contrast, we found that mean selection differed between environments. Similarly, a number of other studies of single mutant effects have also reported significant changes in mean selection across environments (e.g., Remold and Lenski 2001; Kishony and Leibler 2003; Jasnos et al. 2008; Young et al. 2009). Such results do not disprove the Gaussian model but they indicate that if landscapes are Gaussian then they must differ in shape between environments.

Within an environment, mean selection should be constant across genetic backgrounds but the variance in selection should be lower in the better-adapted background if landscapes are Gaussian. Our results are consistent with a constant mean across backgrounds within environments but, however, we did not find evidence of the predicted changes in the variance between adapted and nonadapted backgrounds. These negative findings (for both the mean and variance) could be due to a lack of statistical power. However, we note that we were able to detect differences between environments (Fig. 4) and that our genetic backgrounds did show substantial differences in fitness (Fig. 2).

An alternate approach to evaluating the Gaussian landscape model is to explicitly consider the underlying parameters that lead to the predictions previously discussed. Martin and Lenormand (2006a, eqs. 5 and 6) showed that, under the Gaussian landscape model, $f(s)$ will be a displaced γ distribution that is a function of three landscape parameters: the fitness distance to the optimum s_o, the effective number of phenotypic dimensions n_e, and a scaling parameter λ_e related to the strength of landscape curvature and the phenotypic magnitude of mutations. We can reanalyze our data after parameterizing the likelihood model in terms of s_o, n_e, and λ_e (described in Table S4). For a given assay environment, the distance to the optimum is unknown but we do know that one population is further than the other. More explicitly, we can describe the distance to the optimum of the nonadapted population as $s_{o,\text{non-adapted}} = 1 - (1 - s')(1 - s_{o,\text{adapted}})$, where s' is the reduction in relative fitness of the nonadapted population relative to the adapted population (estimated from the data in Fig. 2). When two populations (*CA* and *SA*) are assayed in the same

environment, they are on the same landscape and, thus, the same values of n_e and λ_e should apply to both populations according to the Gaussian model. Contrary to this prediction, we obtain significantly higher likelihoods in alternative models in which one or more of the constraints predicted by the Gaussian model are relaxed (males in Cd: $P = 0.04$; females: $P = 0.009$; see Table S4 for details).

Taking these results at face value, the Gaussian landscape appears to be an inadequate model for these data. However, these analyses should be regarded with caution for three reasons. First, this analysis may be affected by the bias in estimating female selection, although this concern does not pertain to the male results. Second, this test is strongly dependent on our estimates of s', which come from measures of viability of each genetic background measured separately, rather than in direct competition. If our estimates of s' are invalid, the subsequent analysis cannot be trusted. In experiments with other fly genotypes, fitness differences tend to be larger in competition (e.g., Ho and Agrawal 2012), so we speculate that our estimates of s' may be too low. This would make our test conservative as the likelihood of the Gaussian model declines if larger values of s' are used. Third, the model of Martin and Lenormand (2006a) assumes that mutations are measured on an isogenic background (as is typical in microbe studies) but we measured selective effects in a genetically variable background. Simulations (G. Martin, pers. comm.) examining the consequences of using a genetically variable background suggest that the main prediction are similar: $E(s)$ can change between adapted and nonadapted populations but not by much and $V(s)$ is expected to be greater in nonadapted populations than adapted ones. Although our data do not appear consistent with the Gaussian model, the strength of this conclusion should be tempered by the limitations of the data described here and in the Introduction.

Although the landscape theory is well developed for examining $f(s)$ in a single environment across multiple genetic backgrounds that differ in their proximity to the optima, this has rarely been studied. Burch and Chao (2004) examined the consequences of mutation accumulation in five lines that differed in their initial fitness. In contrast to the prediction of a constant $E(s)$, they inferred that mean selection declined with distance from the optimum. Martin and Lenormand (2006b) commented that these results imply a linear log-fitness function rather than a quadratic one (i.e., Gaussian). Neither our results nor those of Burch and Chao (2004) support the Gaussian model but more data examining $f(s)$ across genotypes of varying degrees of adaptedness are needed. The lack of a strong effect of genetic background on $f(s)$ is also relevant to discussions of epistasis. If negative (i.e., synergistic) epistasis was prevalent, then we would expect that $E(s)$ would be larger in maladapted backgrounds. Because we did not observe such an effect, our data is consistent with the emerging consensus that there is no predominant direction of epistasis (de Visser and Elena 2007; Jasnos and Korona 2007; Martin et al. 2007).

Based on Fig. 2, the difference in fitness between environments (for a given background) is of similar magnitude to the difference between genetic backgrounds (for a given environment). Yet, we see much more pronounced differences in $f(s)$ between environments than between backgrounds. However, when we examine effects on individual genes, we find that environment and genetic background each affect selection on a similar number of genes (Table 5 and Fig. S4). Discrepancies between how context affects individual genes versus distributional properties could arise in any theoretical model, including the landscape model, in which not all genes are affected equally by changes in context. In our case, genetic background effects tend to cancel each other out (e.g., the *CA* background increases selection on some genes but makes equivalent decreases on others) whereas environmental effects do not. The question of why this occurs remains open.

Our discussion has centered on adaptedness as a major determinant of $f(s)$ because this is the focus of the best developed theory (Martin and Lenormand 2006a). However, in our experiment "environment" was the most obvious factor affecting $f(s)$. Unfortunately, we have no robust theory for predicting which environments will cause the types of increases in the mean and variance of $f(s)$ that we observed. Mean selection was stronger and the variance in s was higher in salt than in cadmium but we do not know why. What are the key features of environments that make some more selective than others? "Stress" does not appear to be a good explanation for our results because the salt environment was considerably more stressful for the cadmium-adapted genetic background than the salt-adapted background, yet there was no evidence that genetic background affected $f(s)$. It has been suggested that density-dependence may be a key feature (Agrawal and Whitlock 2010). In environments where fitness is strongly density dependent, the best individuals tend to worsen the environment experienced by others, which will tend to exaggerate the selective differences between genotypes. Under this hypothesis, we would predict that fitness in the salt environment is more strongly density dependent than the cadmium environment. However, the relationship between density dependence and $f(s)$ has not been formally tested here or elsewhere.

Further studies on how $f(s)$ is influenced by adaptedness, density dependence, and other factors are needed. Such data are most likely to come from microorganisms where the experiments can be performed with greater statistical power and precision. Although such data are of great value, we should not rely exclusively on these taxa. There are well-known differences in genomic architecture (e.g., gene number and density, regulatory DNA per gene, intron abundance) between microorganisms and multicellular eukaryotes (Lynch 2007). Moreover, surveys of mutational effects studies from different taxa have been used to infer differences

in the effective dimensionality (Martin and Lenormand 2006a) and patterns of epistasis (Sanjuán and Elena 2006) between such groups. Although studies of $f(s)$ in multicellular organisms will often involve fewer genes or noisier estimates of selection, they provide a necessary point of comparison and validation before extrapolating too far from microorganisms.

ACKNOWLEDGMENTS

The authors are grateful to a number of laboratory assistants, especially L. Sritharan, J. Won, and G. Shang, and to C.C. Spencer for establishing the original cadmium-adapted population. The authors thank G. Martin and an anonymous reviewer for helpful comments; further, G. Martin performed simulations to examine how predictions on $f(s)$ from a Gaussian model are affected by whether mutational effects are measured in an isogenic versus genetically variable background. This work was supported by the Natural Sciences and Engineering Research Council of Canada (ADW, NPS, AFA).

DATA ARCHIVING

The doi for our data is doi:10.5061/dryad.9gg30.

LITERATURE CITED

Agrawal, A. F., and M. C. Whitlock. 2010. Environmental duress and epistasis: how does stress affect the strength of selection on new mutations? Trends Ecol. Evol. 25:450–458.

Burch, C. L., and L. Chao. 2004. Epistasis and its relationship to canalization in the RNA virus ϕ6. Genetics 167:559–567.

Charlesworth, B., and D. Charlesworth. 1998. Some evolutionary consequences of deleterious mutations. Genetica 102–3:3–19.

Charlesworth, B., and K. A. Hughes. 2000. The maintenance of genetic variation in life history traits. Pp. 369–392 in R. S. Singh and C. B. Krimbas, eds. Evolutionary genetics from molecules to morphology. Cambridge Univ. Press, Cambridge, U.K.

Chevin, L.-M., G. Martin, and T. Lenormand. 2010. Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. Evolution 64:3213–3231.

de Visser, J. A. G. M., and S. F. Elena. 2007. The evolution of sex: empirical insights into the roles of epistasis and drift. Nat. Rev. Genet. 8:139–149.

Eyre-Walker, A., and P. D. Keightley. 2007. The distribution of fitness effects of new mutations. Nat. Rev. Genet. 8:610–618.

Fisher, R. A. 1930. The genetical theory of natural selection. 1st ed. Oxford Univ. Press, Oxford, U.K.

Hartfield, M., S. P. Otto, and P. D. Keightley. 2010. The role of advantageous mutations in enhancing the evolution of a recombination modifier. Genetics 184:1153–1164.

Hillenmeyer, M. E., E. Fung, J. Wildenhain, S. E. Pierce, S. Hoon, W. Lee, M. Proctor, R. P. St Onge, M. Tyers, D. Koller, et al. 2008. The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science 320:362–365.

Ho, E., and A. F. Agrawal. 2012. The effects of competition on the strength and softness of selection. J. Evol. Biol. 25:2537–2546.

Jasnos, L., and R. Korona. 2007. Epistatic buffering of fitness loss in yeast double deletion strains. Nat. Genet. 39:550–554.

Jasnos, L., K. Tomala, D. Paczesniak, and R. Korona. 2008. Interactions between stressful environment and gene deletions alleviate the expected average loss of fitness in yeast. Genetics 178:2105–2111.

Keightley, P. D., and O. Ohnishi. 1998. EMS-induced polygenic mutation rates for nine quantitative characters in *Drosophila melanogaster*. Genetics 148:753–766.

Keightley, P. D., and S. P. Otto. 2006. Interference among deleterious mutations favours sex and recombination in finite populations. Nature 443:89–92.

Kishony, R., and S. Leibler. 2003. Environmental stresses can alleviate the average deleterious effect of mutations. J. Biol. 2:14.

Lalić, J., J. M. Cuevas, and S. F. Elena. 2011. Effect of host species on the distribution of mutational fitness effects for an RNA virus. PLoS Genet. 7:e1002378.

Lynch, M. 2007. The origins of genome architecture. Sinauer Associates Inc., Sunderland, MA.

Lynch, M., J. Conery, and R. Bürger. 1995a. Mutation accumulation and the extinction of small populations. Am. Nat. 146:489–518.

———. 1995b. Mutational meltdowns in sexual populations. Evolution 49:1067–1080.

MacLellan, K., L. Kwan, M. C. Whitlock, and H. D. Rundle. 2012. Dietary stress does not strengthen selection against single deleterious mutations in *Drosophila melanogaster*. Heredity 108:203–210.

Mallet, M. A., C. M. Kimber, and A. K. Chippindale. 2012. Susceptibility of the male fitness phenotype to spontaneous mutation. Biol. Lett. 8:426–429.

Martin, G., and T. Lenormand. 2006a. A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. Evolution 60:893–907.

———. 2006b. The fitness effect of mutations across environments: a survey in light of fitness landscape models. Evolution 60:2413–2427.

Martin, G., S. F. Elena, and T. Lenormand. 2007. Distributions of epistasis in microbes fit predictions from a fitness landscape model. Nat. Genet. 39:555–560.

Orr, H. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution 52:935–949.

R Development Core Team, 2012 R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Remold, S. K., and R. E. Lenski. 2001. Contribution of individual random mutations to genotype-by-environment interactions in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 98:11388–11393.

———. 2004. Pervasive joint influence of epistasis and plasticity on mutational effects in *Escherichia coli*. Nat. Genet. 36:423–426.

Roze, D. 2012. Spatial heterogeneity in the strength of selection against deleterious alleles and the mutation load. Heredity 109:137–145.

Sanjuán, R., and S. F. Elena. 2006. Epistasis correlates to genomic complexity. Proc. Natl. Acad. Sci. USA 103:14402–14405.

Szafraniec, K., R. Borts, and R. Korona. 2001. Environmental stress and mutational load in diploid strains of the yeast *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA 98:1107–1112.

Thibault, S. T., M. A. Singer, W. Y. Miyazaki, B. Milash, N. A. Dompe, C. M. Singh, R. Buchholz, M. Demsky, R. Fawcett, H. L. Francis-Lang, et al. 2004. A complementary transposon tool kit for *Drosophila melanogaster* using *P* and *piggyBac*. Nat. Genet. 36:283–287.

Trindade, S., A. Sousa, and I. Gordo. 2012. Antibiotic resistance and stress in the light of Fisher's model. Evolution 66:3815–3824.

Vale, P. F., M. Choisy, R. Froissart, R. Sanjuán, and S. Gandon. 2012. The distribution of mutational fitness effects of phage φX174 on different hosts. Evolution 66:3495–3507.

Wang, A. D., N. P. Sharp, C. C. Spencer, K. Tedman-Aucoin, and A. F. Agrawal. 2009. Selection, epistasis, and parent-of-origin effects on

deleterious mutations across environments in *Drosophila melanogaster*. Am. Nat. 174:863–874.

Wang, Y., C. Díaz Arenas, D. M. Stoebel, and T. F. Cooper. 2012. Genetic background affects epistatic interactions between two beneficial mutations. Biol. Lett. 9:20120328.

Wardlaw, A. M., and A. F. Agrawal. 2012. Temporal variation in selection accelerates mutational decay by Muller's ratchet. Genetics 191:907–916.

Whitlock, M. C. 2005. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. J. Evol. Biol. 18:1368–1373.

Young, J. A., C. P. Yourth, and A. F. Agrawal. 2009. The effect of pathogens on selection against deleterious mutations in *Drosophila melanogaster*. J. Evol. Biol. 22:2125–2129.

Associate Editor: A. De Visser

Supporting Information

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

Table S1. List of mutations tested.
Table S2. Context-dependent fitness effects of individual Exelixis mutations.
Table S3. Maximum likelihood estimates for the distribution of fitness effects for males (top table) and females (bottom table).
Table S4. Maximum-likelihood models using Gaussian landscape parameterization.
Figure S1. Migration procedure to ensure experimental populations retained their environment-specific adaptations.
Figure S2. Comparison of the fitness effects of w+ and the average fitness effects of the Exelixis mutations (all measured against the w1118 reference).
Figure S3 (A) Correlation between selection coefficients for individual Exelixis mutations across the cadmium (x-axis) and salt (y-axis) environments within the CA genetic background in males (left) and females (right). (B) Correlation between selection coefficients for individual Exelixis mutations across the cadmium (x-axis) and salt (y-axis) environments within the SA genetic background in males and females.
Figure S4. Gene-specific analyses of environment and genetic background effects.