Testing for Environmentally Induced Bias in Phenotypic Estimates of Natural Selection: Theory and Practice

John R. Stinchcombe,[1,*] Matthew T. Rutter,[1,†] Donald S. Burdick,[2] Peter Tiffin,[3] Mark D. Rausher,[1] and Rodney Mauricio[4]

1. Department of Biology, Duke University, Durham, North Carolina 27708-0338;
2. Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina 27708-0251;
3. Department of Plant Biology, University of Minnesota, St. Paul, Minnesota 55108-1095;
4. Department of Genetics, University of Georgia, Athens, Georgia 30602-7223

Submitted June 14, 2001; Accepted April 11, 2002

abstract: Measuring natural selection has been a fundamental goal of evolutionary biology for more than a century, and techniques developed in the last 20 yr have provided relatively simple means for biologists to do so. Many of these techniques, however, share a common limitation: when applied to phenotypic data, environmentally induced covariances between traits and fitness can lead to biased estimates of selection and misleading predictions about evolutionary change. Utilizing estimates of breeding values instead of phenotypic data with these methods can eliminate environmentally induced bias, although this approach is more difficult to implement. Despite this potential limitation to phenotypic methods and the availability of a potential solution, little empirical evidence exists on the extent of environmentally induced bias in phenotypic estimates of selection. In this article, we present a method for detecting bias in phenotypic estimates of selection and demonstrate its use with three independent data sets. Nearly 25% of the phenotypic selection gradients estimated from our data are biased by environmental covariances. We find that bias caused by environmental covariances appears mainly to affect quantitative estimates of the strength of selection based on phenotypic data and that the magnitude of these biases is large. As our estimates of selection are based on data from spatially replicated field experiments, we suggest that our findings on the prevalence of bias caused by environmental covariances are likely to be conservative.

* Corresponding author. Present address: Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912; e-mail: john_stinchcombe@brown.edu.

† Present address: Department of Biology, University of Maryland, College Park, Maryland 20742.

Am. Nat. 2002. Vol. 160, pp. 511–523.

Keywords: environmental covariance, natural selection, phenotypic selection, quantitative genetics, selection gradients, plant size, herbivory.

Natural selection is a major force of evolutionary change, and for this reason evolutionary biologists have long been interested in describing and quantifying the action of selection in natural populations (Endler 1986; Kingsolver et al. 2001). Although attempts to measure natural selection date to the turn of the twentieth century (Bumpus 1899; Weldon 1901), the past 20 yr have seen an explosion in the quantification of selection in nature. An important stimulus for much of this work has been the development of methods for estimating the pattern of selection acting on multiple quantitative traits (Lande 1979; Lande and Arnold 1983; Arnold and Wade 1984*a*, 1984*b*; Schluter 1988; Phillips and Arnold 1989; Rausher 1992; Schluter and Nychka 1994; Janzen and Stern 1998; Scheiner et al. 2000). The approach developed by Lande and Arnold (1983), which provides a simple means by which field biologists can estimate selection acting on suites of potentially correlated traits, has been particularly influential (Brodie et al. 1995).

The Lande-Arnold approach is particularly suited for estimating the magnitude of selection in natural populations because it requires only measurements on an individual's phenotype and does not require information about relationships among individuals. These phenotypic measurements are then used to perform a multiple linear regression of relative fitness on the trait values. Each of the resulting partial regression coefficients, termed a "directional selection gradient," is an estimate of the magnitude of directional selection acting on one of the characters, with the effects of selection on all of the other measured characters removed. Similarly, each of the second-order coefficients from a quadratic multiple regression of phenotypic values of relative fitness on the character values provides an estimate of the magnitude of stabilizing, disruptive, or correlational selection acting on

each character, with the effects of selection on all of the other measured traits removed.

Although this regression method has proven extremely useful, several authors, including Lande and Arnold (1983) themselves, have pointed out potentially important statistical and biological limitations in its use (Mitchell-Olds and Shaw 1987; Wade and Kalisz 1990; Willis 1996). One of these limitations has been particularly well explored theoretically: selection gradients can be biased if covariances between traits and fitness are caused by each being influenced by the same or correlated environmental factors (van Noordwijk et al. 1988; Schluter et al. 1991; Rausher 1992; van Tienderen and de Jong 1994). Such environmental covariances are likely to be of particular concern to field biologists.

Environmentally induced covariances between traits and fitness can arise whenever environmental factors that influence trait values also affect fitness either directly or through an unmeasured trait (reviewed in Mauricio and Mojonnier 1997). Rausher (1992) provided the following example of this phenomenon and its evolutionary consequences. Consider a hypothetical population of plants that exhibits genetic variation for concentration of nitrogen-based defensive alkaloids. In this population, alkaloid concentration has no direct effect on fitness. However, alkaloid concentration is influenced by an environmental factor, soil nitrogen content, with the result that plants growing in nitrogen-rich soil produce higher concentrations of alkaloids than plants growing in nitrogen-poor soil. Moreover, in this population, plants growing in nitrogen-rich soil have higher fitness than plants growing in nitrogen-poor soil because nitrogen is limiting to plant growth and reproduction.

Because soil nitrogen level influences both fitness and alkaloid concentration, the phenotypic values of these two traits are positively correlated. Consequently, the Lande-Arnold method would indicate that directional selection acts to increase alkaloid concentrations. Moreover, since there is genetic variation for alkaloids, an increase in alkaloid concentration in the next generation would be predicted. However, because alkaloid concentration does not affect fitness (in this hypothetical example), genotypes that differ in alkaloid concentration do not differ in fitness, and the genetic covariance between alkaloid concentration and fitness will equal 0. Alkaloid concentration will thus not evolve because an evolutionary change in a quantitative character requires that there be an additive genetic covariance between the character and fitness (Fisher 1958; Price 1970; Crow and Nagylaki 1976). The environmentally induced covariance between alkaloid concentration and fitness thus results in misleading predictions of evolutionary change.

Environmentally induced covariances between a trait and fitness are likely to be common in nature, especially for physiological traits or other traits, including fitness, that vary depending on the "condition" of the organism (Mauricio and Mojonnier 1977). For example, environmental covariances caused by effects of nutritional state on both metric traits and fitness have been proposed to mask the pattern of selection on several traits in birds, including breeding date (Price et al. 1988), clutch size (Price and Liou 1989), and tarsus length (Alatalo et al. 1990).

One method for removing bias introduced by environmentally induced covariances was developed by Rausher (1992). This method is identical to the multivariate regression method proposed by Lande and Arnold (1983) except that estimates to additive genetic or breeding values (or genotypic values in the case of highly selfing species) are used in place of phenotypic values. By using breeding value regression coefficients (which can be approximated by family-mean regressions) as estimates of selection, we can average out the effects of environmental factors on trait values of individuals. Consequently, covariances between the means of fitness and trait values, and the regression coefficients on which they are based, should reflect selection gradients unaffected by environmentally induced bias (Mauricio and Mojonnier 1997). Estimates of selection that are based on breeding values should therefore provide more accurate predictors of evolutionary change within a population.

Estimates of selection based on breeding values are not, however, a universal solution. First, as described by Rausher (1992), the breeding values method will also produce a biased estimate of selection if correlated traits are not included. For the breeding values method, however, this bias will be limited to cases where genetically correlated traits are omitted from the analysis, in contrast to phenotypic methods that produce biased estimates if either environmentally or genetically correlated traits are omitted. Furthermore, as Rausher (1992) noted, this method can only be applied to traits for which there is significant genetic variation. Thus, the method is not applicable if selection has already eliminated genetic variation, as has been predicted for traits that are closely related to fitness (e.g., life-history traits). Nevertheless, abundant evidence accumulated over the past 30 yr indicates that most traits, including life-history traits closely linked to fitness, are usually genetically variable (e.g., Roff 1992), suggesting that this potential limitation may seldom be a significant problem.

Although the breeding values method does provide a solution to the problem of environmentally induced biases in selection analyses, it may not be readily applicable for many field biologists. Given that the data required for the standard Lande-Arnold phenotypic analysis are much eas-

ier to obtain, it would be convenient if environmentally induced biases were rare and of small magnitude. To date, however, there exists little empirical evidence indicating the magnitude of such biases or whether they tend to occur in certain types of traits but not others. It is thus unclear whether evolutionary biologists need to be concerned with such biases when estimating the magnitude of selection. The objective of the analyses presented here is to address this question by formally evaluating the magnitude of environmentally induced biases in three large data sets we have collected. In doing so, we also describe a new statistical procedure for determining whether environmental covariances will bias phenotypic estimates of natural selection. This corrects an error in the tests used by Rausher and Simms (1989) and later described by Rausher (1992).

Methods

Comparing Standardized Phenotypic and Genotypic Selection Gradients

To determine whether phenotypic selection gradients are biased by environmental covariances between traits and fitness, one must quantitatively compare estimates of selection based on genetic and phenotypic data. Unfortunately, making direct comparisons between selection gradients that are estimated using phenotypic and genotypic methods introduces several statistical issues. One issue involves whether to compare standardized or unstandardized selection gradients. Although the dynamical equation for evolutionary change in quantitative characters (eq. [7a] in Lande 1979) employs the unstandardized selection gradient, the phenotypic and genotypic selection gradients to be compared have typically been estimated from data in which trait values have been standardized to a mean of 0 and a variance of 1. However, the variance in a sample of family means and the variance in a sample of phenotypic data will not be equal. Standardizing the phenotypic and genetic data separately can lead to very different estimates of the magnitude of selection, even in the absence of environmentally induced bias. This can be seen from the formula for a standardized selection gradient:

$$\beta' = \beta \times \sigma, \tag{1}$$

where β' is the standardized selection gradient, β is the unstandardized selection gradient, and σ is the standard deviation of the trait of interest (Lande and Arnold 1983). Thus, even if β is identical for phenotypic and genetic selection analyses, differences in σ will produce different estimates of β'. In general, if the phenotypic variance in a trait is at least partially determined by environmental variance, then differences in the standard deviations of phenotypic and genotypic values will result in differences in the estimated standardized phenotypic and genotypic selection gradients, even in the absence of bias caused by environmentally induced covariances. Consequently, comparisons of standardized genetic and phenotypic selection gradients are not necessarily evidence of environmentally induced biases in the pattern of selection; rather, such differences may simply reflect differences in the magnitudes of phenotypic and genotypic trait variance.

Comparing Unstandardized Phenotypic and Genotypic Selection Gradients

Theoretical work has shown that in the absence of environmentally induced covariances between traits and fitness, estimates of selection calculated using phenotypic and breeding values are equal (Rausher 1992). Accordingly, in this article, we interpret any statistical evidence that selection gradients calculated with these two types of data differ as evidence of a bias caused by an environmental covariance (Rausher and Simms 1989; Rausher 1992). The statistical approach we use to compare selection gradients is a correction of one that was suggested by Rausher and Simms (1989) and Rausher (1992).

Our statistical approach is motivated by the following decomposition of a phenotypic regression coefficient, which is given by

$$\beta_p = \frac{\mathrm{cov}_p(W, Y)}{\mathrm{var}_p(Y)}, \tag{2}$$

where W indicates fitness and Y indicates the trait of interest. Partitioning the phenotypic covariance and variance into their standard components (e.g., Falconer and MacKay 1996) yields

$$\beta_p = \frac{\mathrm{cov}_a(W, Y)}{\mathrm{var}_a(Y)} \times h^2 + \frac{\mathrm{cov}_e(W, Y)}{\mathrm{var}_e(Y)} \times (1 - h^2), \tag{3}$$

$$\beta_p = \beta_a \times h^2 + \beta_e \times (1 - h^2), \tag{4}$$

where β_p, β_a, and β_e are the phenotypic, additive genetic, and environmental regression coefficients, respectively, and $h^2 = \mathrm{var}_a(Y)/\mathrm{var}_p(Y)$. Here the nonadditive genetic effects are combined with the environmental effects in the second term (e.g., Lande 1979).

In the absence of environmentally induced correlations between a trait and fitness, $\beta_p = \beta_a$ (Rausher 1992). Substituting β_a into equation (4) for β_p and rearranging yields $\beta_a = \beta_e$. In other words, in the absence of environmentally induced bias, the breeding value regression coefficient equals the environmental regression coefficient (the coefficient of a regression of environmental deviations for

fitness on environmental deviations for the character). Intuitively, this equality indicates that environmentally induced variation in the trait has the same effect on fitness as genetically induced variation in the trait. As the following paragraphs describe, our statistical approach to testing whether genotypic and phenotypic regression coefficients are equal is actually to test the equivalent hypothesis that the genotypic and environmental selection gradients are equal.

For a single character, the statistical model for the genotypic selection analysis is

$$w = \beta_1 X + \text{error}, \tag{5}$$

where w denotes relative fitness of an individual, X is the trait breeding value for that individual's family, and β_1 is the genotypic selection gradient. The deviations of individual phenotypic values from breeding values ($Y_i - X$, where Y_i indicates the phenotypic value of an individual) are the net effect of dominance, epistatic, and environmentally induced deviations. Because breeding values are not correlated with any of these component deviations (Kempthorne 1969; Falconer and MacKay 1996), breeding values, X, are not correlated with the net phenotypic deviations from the breeding values ($Y_i - X$). Consequently, equation (5) yields the same estimate of β_1 as the model

$$w = \beta_1 X + \beta_2 (Y_i - X) + \text{error}, \tag{6}$$

where β_2 is the selection gradient for $Y_i - X$. Although $Y_i - X$ contains nonadditive genetic effects, for simplicity in usage we refer to this term as the environmental deviation and to β_2 as the environmental selection gradient. Because $Y_i - X$ is not correlated with X, equations (5) and (6) yield the same estimate for the genotypic selection gradient, β_1. Note that if the phenotypic values themselves (Y_i) rather than the environmental deviations ($Y_i - X$) were used in equation (6), there would be a strong colinearity of the independent variables, hindering both hypothesis testing and biological interpretation (e.g., Mitchell-Olds and Shaw 1987).

The statistical model for the standard Lande-Arnold phenotypic selection analysis is

$$w = \beta_3 Y_i + \text{error}, \tag{7}$$

where w is relative fitness, Y_i is the phenotypic value of traits, and β_3 is the phenotypic selection gradient. Equations (5) and (7) are identical except that equation (7) uses phenotypic values rather than estimates of breeding values. Therefore, when $\beta_1 = \beta_3$, the phenotypic and genotypic approaches will yield the same estimate of selection. Finding the conditions under which this equality holds is equivalent to finding the conditions under which the models in equations (6) and (7) are equivalent. Equations (6) and (7) are equivalent when $\beta_2 = \beta_1$, because then equation (6) reduces to

$$w = \beta_1 Y_i + \text{error}, \tag{8}$$

which is the same form as equation (7), the phenotypic selection analysis. Therefore, equation (6) provides a framework for detecting environmentally induced bias in phenotypic selection gradients: bias is present if β_2 does not equal β_1.

To test the hypothesis that $\beta_1 = \beta_2$, standard contrasts for linear models (e.g., Searle 1971) can be used; these contrasts utilize an F-test to evaluate the equality of the two regression coefficients. To evaluate whether environmentally induced bias exists in stabilizing/disruptive selection gradients, we can use a similar approach to test the equality of second-order coefficients in a quadratic multiple regression of fitness on the characters.

Relationship to Previous Approaches

The analysis presented above (specifically eq. [6]) corrects a mistake in the approach previously suggested by Rausher and Simms (1989) and Rausher (1992). In their test, the statistical model is identical to ours except that the deviations have the opposite sign:

$$w = \beta_1 X + \beta_2 (X - Y_i) + \text{error}. \tag{9}$$

Expanding the second term and assuming that there is no environmentally induced covariance between the trait of interest and fitness (i.e., that $\beta_1 = \beta_2$), we reduce equation (9) to

$$w = 2\beta X - \beta Y_i + \text{error}. \tag{10}$$

Equation (10) is not a straightforward expression for the phenotypic analysis of selection (i.e., it is not equivalent to eq. [7]). Therefore, testing the hypothesis that $\beta_1 = \beta_2$ in equation (9), as suggested by Rausher and Simms (1989) and Rausher (1992), does not test the null hypothesis that phenotypic estimates of selection are unbiased by environmentally induced covariances between traits and fitness.

Determining the Appropriate Degrees of Freedom for Hypothesis Testing

Since half- or full-sibs from the same family are not statistically independent (because individuals cannot be randomized to families as they are to experimental treat-

ments), the degrees of freedom used by standard linear contrasts to test whether $\beta_1 = \beta_2$ may not be appropriate. In this section, we present procedures for determining the appropriate degrees of freedom for this linear contrast. We start with the multiple regression equation described above:

$$w = \beta_1 X + \beta_2 (Y_i - X) + \text{error}. \tag{11}$$

Estimating β_1 and β_2 from equation (11) utilizes two degrees of freedom. Testing the hypothesis that $\beta_1 = \beta_2$ with a standard linear contrast pools the remaining degrees of freedom that are associated with X and $Y_i - X$ in the error term as the denominator of the F-test. Because genetically related individuals are used whenever genotypic selection gradients are estimated, this pooling may not be valid and could artificially elevate the degrees of freedom associated with the error term. The improper pooling of degrees of freedom in the denominator of the F-test thus results in an evaluation of environmentally induced bias that is too powerful. As such, in all cases where environmentally induced bias is indicated by a standard linear contrast test of whether $\beta_1 = \beta_2$ in equation (11), the appropriateness of using pooled degrees of freedom should be evaluated.

Determining whether one should adjust the degrees of freedom used when testing for bias can be assessed directly with the following statistical model:

$$w = \beta_1 X + \text{FAMILY} + \beta_2 (Y_i - X) + \text{error}, \tag{12}$$

where FAMILY is a categorical variable containing each individual's family identification. A significant FAMILY term in an ANCOVA based on equation (12) indicates the presence of genetic variation for fitness after the effects of the trait of interest are considered, presumably because of other traits under selection. Therefore, a significant FAMILY term suggests that pooling the degrees of freedom associated with the family effect with the error degrees of freedom for the linear contrast is invalid. Conversely, a nonsignificant FAMILY term indicates that there is no need to adjust the degrees of freedom used in a standard linear contrast test. When pooling is not justified, parameter estimates and mean squares from equations (11) and (12) can be used to determine whether estimates of selection are unbiased, using an alternative method that does not involve pooling. This alternative method utilizes the mean squares estimated from the statistical models described by equations (11) and (12) in the following manner.

First, one calculates F statistics for β_1 and β_2 that are not contaminated by pooling the family and error degrees of freedom. In other words, one calculates an F statistic for $\beta_1 (F_1)$ using only genetic data and, likewise, an F statistic for $\beta_2 (F_2)$ using only phenotypic data:

$$F_1 = \frac{\text{Type I mean square for } X \text{ from eq. (11)}}{\text{Type I mean square for FAMILY from eq. (12)}}, \tag{13}$$

$$F_2 = \frac{\text{mean square for } Y_i - X \text{ from eq. (11)}}{\text{mean square error from eq. (12)}}. \tag{14}$$

In equation (13), the family effect is used as the denominator for the F statistic associated with the family means parameter estimate—the genetic selection gradient. In equation (14), the error term from the model that prevents pooling is used as the denominator for the F statistic associated with the environmental deviations parameter estimate—the environmental selection gradient.

To calculate these F statistics, the effects of variable X must be removed before the mean square for FAMILY is calculated using equation (12); in other words, the sum of squares (SS) for FAMILY is calculated as a reduction SS, which is equal to the difference between the SS with and without FAMILY in the model (Searle 1971). The mean square calculated from this reduction SS is equivalent to the Type I mean square of common statistical software (e.g., PROC GLM of SAS). Because reduction sums of squares are calculated sequentially in the order they appear in a statistical model, if FAMILY were to be included in the statistical model for equation (12) before X, all the between-family variation would have already been accounted for before the effects of X were incorporated into the model. Including FAMILY before X in the statistical model for equation (12) would therefore cause the sum of squares for X to be a meaningless 0 with 0 degrees of freedom. (This is exactly what happens to the Type III sum of squares for both X and FAMILY in eq. [12] in this context.) To calculate F_2, the mean square error from the analysis described by equation (12) is utilized because the FAMILY term, as a categorical variable, prevents the pooling of the family and error degrees of freedom.

The F statistics calculated using equations (13) and (14) can be used to calculate corrected standard errors for β_1 and β_2 and, therefore, a t statistic for testing whether $\beta_1 = \beta_2$. Because an F statistic associated with a regression parameter is equal to the square of the parameter estimate divided by the square of its standard error (Draper and Smith 1966), β_1, F_1, β_2, and F_2 can be used to calculate corrected standard errors for β_1 and β_2:

$$\text{SE for } \beta_1 = \sqrt{\frac{\beta_1^2}{F_1}}, \tag{15a}$$

$$\text{SE for } \beta_2 = \sqrt{\frac{\beta_2^2}{F_2}}. \tag{15b}$$

Given these estimates of β_1, β_2, and their standard errors,

the hypothesis that $\beta_1 = \beta_2$ can be tested using a *t*-test for the equality of two regression coefficients:

$$t = \frac{\beta_1 - \beta_2}{\sqrt{(\text{SE for } \beta_1)^2 + (\text{SE for } \beta_2)^2}}. \quad (16)$$

The significance of the *t* statistic calculated in equation (16) should be tested with degrees of freedom equal to the number of families minus 2.

Description of the Data

We compared phenotypic and genotypic selection gradients from three field studies of annual plants, all of which examined the pattern of selection on ecologically important traits (Mauricio and Rausher 1997; Tiffin and Rausher 1999; Stinchcombe and Rausher 2001). The studies are similar in many respects. All took place in agricultural fields near Durham, North Carolina. Within each field, a set of family lines was planted in a randomized block design. Traits of interest (table 1) and estimates of fitness (total number of viable seeds for *Ipomoea*, fruit production for *Arabidopsis*) were measured on all individuals, thus creating a set of phenotypic values. We calculated relative fitness by dividing individual fitness by the population mean fitness. In each experiment, all individuals were generated by a half- or self-sib breeding design, which allowed the calculation of family means for each trait. These family means were used to estimate genotypic selection gradients, which represent breeding value regression coefficients for the predominantly outcrossing *Ipomoea purpurea* and genotypic value regression coefficients for the highly selfing *Ipomoea hederacea* and *Arabidopsis thaliana*. Because the genotypic selection analysis is only applicable to genetically variable traits, we only analyzed traits that exhibited significant genetic variation, as determined by a significant family effect in an ANOVA (Mauricio and Rausher 1997; Tiffin and Rausher 1999; Stinchcombe and Rausher 2001).

In the studies of *I. purpurea* and *I. hederacea*, seeds were planted in the field, while in the study of *A. thaliana*, plants were transplanted as seedlings. In the experiments in which there was an insecticide treatment, previous work had determined that the pattern of selection differed between treatments (i.e., the presence of insects altered the pattern of selection). As a result, we analyzed the two treatments separately. Finally, in *A. thaliana*, trichome density and glucosinolate concentration were genetically correlated (Mauricio and Rausher 1997), so we used a joint analysis that included both characters to assess whether the pattern of phenotypic selection on these traits was biased by environmental covariances.

Results

Effects of Differential Standardization

As noted above, differential standardization of phenotypic and genotypic selection gradients has the potential to confound direct comparisons of the two types of gradients. To illustrate the expectation that standardized selection gradients will differ, in tables 2–4 we present standardized and unstandardized phenotypic and genotypic selection gradients along with the results of our analysis of the presence of bias according to equation (6) described above.

Inspection of tables 2–4 reveals that in many cases, the standardized phenotypic and genotypic selection gradients appear markedly different and without overlapping confidence intervals, yet the unstandardized selection gradients were not found to be significantly different. For example, the standardized phenotypic directional selection gradient on resistance to seed capsule damage in *Ipomoea*

Table 1: Description of the three studies that were the source of the data presented in this article

	Arabidopsis thaliana	*Ipomoea purpurea*	*Ipomoea hederacea*
Number of individuals	1,728	1,440	1,440
Number of families[a]	144	25	18
Insecticide treatment	Yes	No	Yes
Traits measured:			
Leaf area	Yes	Yes	Yes
Resistance to leaf damage[b]	Yes	Yes	Yes
Resistance to capsule damage[b]	No	Yes	No
Glucosinolate concentration	Yes	No	No
Trichome density	Yes	No	No

Note: Further information can be found in Mauricio and Rausher (1997; *A. thaliana*), Tiffin and Rausher (1999; *I. purpurea*), and Stinchcombe and Rausher (2001; *I. hederacea*).

[a] Inbred lines were used in the *A. thaliana* and *I. hederacea* experiments, and paternal half-sib families were used in the *I. purpurea* experiment.

[b] Resistance to leaf damage and capsule damage were measured for as 1 minus percent leaf area missing and 1 minus percent capsules damaged.

Table 2: Summary of the statistical analyses of bias caused by environmental covariances in phenotypic estimates of the pattern of natural selection in *Ipomoea hederacea*

	Insects absent		Insects present	
Trait	Directional	Quadratic	Directional	Quadratic
Total leaf area:				
1.	β'_p = **.79*** **(.03)**	γ'_p = −.081* (.02)	β'_p = .044 (.05)	γ'_p = −.016 (.03)
2.	β'_g = **.018 (.06)**	γ'_g = .022 (.09)	β'_g = .02 (.05)	γ'_g = −.04 (.05)
3.	β_p = **.000007*** **(.0000003)**	γ_p = −.00000001* (.00)	β_p = .0000004 (.0000004)	γ_p = .00 (.00)
4.	β_g = **.0000009 (.000003)**	γ_g = .00 (.00)	β_g = .0000012 (.000003)	γ_g = .00 (.00)
5.	t = **2.09**, df = **16**, P = **.052**	t = .1, df = 16, P = .92	F = .09, df = 1,614, P = .76	F = .08, df = 1,614, P = .77
Resistance to leaf damage:				
1.	β'_p = **.41*** **(.04)**	γ'_p = .16* (.03)	β'_p = .50* (.04)	γ'_p = **.13*** **(.03)**
2.	β'_g = **−.005 (.06)**	γ'_g = −.13 (.08)	β'_g = .026 (.05)	γ'_g = **.0028 (.06)**
3.	β_p = **1.85*** **(.20)**	γ_p = 3.14* (.052)	β_p = 2.14* (.18)	γ_p = **2.38*** **(.48)**
4.	β_g = **−.12 (1.5)**	γ_g = −71.35 (43.9)	β_g = .53 (.99)	γ_g = **12.16 (23.84)**
5.	t = **2.12**, df = **16**, P = **.049**	t = .14, df = 16, P = .89	F = 3.29, df = 1,614, P = .07	F = **3.8**, df = **1,614**, P = **.0516**

Note: Standardized (β', γ') and unstandardized (β, γ) estimates of linear and quadratic selection gradients estimated using phenotypic or genetic data (p and g subscripts, respectively). SEs are in parentheses. Underneath these entries, an F statistic, degrees of freedom, and P value are presented for the statistical analysis of whether the phenotypic pattern of selection is biased by environmental covariances. For cases in which utilizing a linear contrast would have artificially elevated the degrees of freedom associated with the error term of the F-test, we present a t statistic (see text). Significant cases of bias are indicated by cells in boldface font.

* $P < .05$ (statistically significant selection gradient).

purpurea ($\beta'_p = 0.30 \pm 0.03$) appears to be greater than the standardized genotypic directional selection gradient ($\beta'_g = 0.06 \pm 0.03$); however, estimates of these selection gradients made from unstandardized data are nearly identical ($\beta_p = 1.58 \pm 0.17$; $\beta_g = 1.56 \pm 0.90$). In like fashion, the standardized phenotypic directional selection gradient on total leaf area in *I. purpurea* appears greater than the standardized genotypic selection gradient ($\beta'_p = 0.84 \pm 0.02$; $\beta'_g = 0.13 \pm 0.03$), but estimates from unstandardized data are also nearly identical ($\beta_p = 0.002 \pm 0.0005$; $\beta_g = 0.002 \pm 0.0004$). These apparent discrepancies result from differences in standardization and demonstrate that estimates of environmentally induced bias based on direct comparison of standardized phenotypic and genotypic selection gradients can be misleading. Because of the potentially misleading effects of standardization, in the remaining sections we consider unstandardized gradients only.

Qualitative Agreement between Methods

In many cases, it is of interest to determine only the direction of selection, not its precise magnitude. We therefore first consider how frequently the phenotypic and genotypic analyses of selection differ in the sign of estimated selection gradients. For approximately 86% (26 of 30) of the directional and quadratic selection gradients examined, the two methods agreed in the sign of the gradient (tables 2–5). Of the remaining four gradients for which the two methods disagreed in sign, in each case, one or both methods produced a gradient that was not significantly different from 0, indicating that there is little statistical support for a difference in the sign of selection gradients. Thus, for all 30 gradients examined, the two methods gave statistically consistent estimates of the direction of selection; that is, there is not detectable bias in this qualitative aspect of selection.

Another way of assessing qualitative agreements between the methods is to ask whether the two methods disagree in whether selection was acting. For approximately one-third of the gradients examined (11 of 30), the two analyses agreed in whether a gradient was significant (tables 2–5; in none of these 11 cases did the two methods disagree on the sign of the gradient). Conversely, in 63% of the cases (19 of 30), one approach indicated a statistically nonzero gradient while the other did not, suggesting that the two methods potentially yielded different estimates of whether selection was acting. However, in 15 of these 19 cases, it was the phenotypic selection gradient that differed significantly from 0, and of these 15 significant phenotypic selection gradients, 12 had the same sign as the nonsignificant genotypic selection gradient. Rather

Table 3: Summary of the statistical analysis of bias caused by environmental covariances in phenotypic estimates of the pattern of natural selection in *Ipomoea purpurea*

Trait	Directional	Quadratic
Total leaf area:		
1.	$\beta'_p = .84^*$ (.02)	$\gamma'_p = -.08^*$ (.007)
2.	$\beta'_g = .13^*$ (.03)	$\gamma'_g = -.04$ (.02)
3.	$\beta_p = .002^*$ (.00006)	$\gamma_p = -.0000005^*$ (.00000004)
4.	$\beta_g = .002^*$ (.0004)	$\gamma_g = -.000008$ (.000006)
5.	$F = .34$, df = 1,1146, $P = .55$	$F = 1.50$, df = 1,1146, $P = .22$
Resistance to leaf damage:		
1.	$\beta'_p = .14^*$ (.03)	$\gamma'_p = .001$ (.01)
2.	$\beta'_g = .02$ (.03)	$\gamma'_g = .02$ (.02)
3.	$\beta_p = 3.05^*$ (.75)	$\gamma_p = .6975$ (4.7)
4.	$\beta_g = 2.79$ (3.8)	$\gamma_g = 256.88$ (262.44)
5.	$F = .05$, df = 1,1142, $P = .83$	$F = .04$, df = 1,1142, $P = .84$
Resistance to capsule damage:		
1.	$\beta'_p = .30^*$ (.03)	$\gamma'_p = -.059^*$ (.02)
2.	$\beta'_g = .06$ (.03)	$\gamma'_g = -.043$ (.03)
3.	$\beta_p = 1.59^*$ (.18)	$\gamma_p = -1.64^*$ (.68)
4.	$\beta_g = 1.56$ (.90)	$\gamma_g = -33.14$ (23.54)
5.	$F = .007$, df = 1,1126, $P = .93$	$F = .04$, df = 1,1126, $P = .85$

Note: Notations and symbols are the same as in table 2.

* $P < .05$ (statistically significant selection gradient).

than indicating disagreement about whether selection was acting, these data are more consistent with the interpretation that the difference in significance was due to a difference in statistical power: the larger number of significant phenotypic selection gradients suggests that the genotypic analysis has less power because of smaller sample sizes. However, the appreciable number of cases with significant genotypic but not phenotypic selection gradients indicates that this conclusion is not universal.

Quantitative Agreement between Methods

Although there appears to be substantial qualitative agreement between phenotypic and genotypic methods in assessing the direction of selection, the methods differ in quantitative estimates of the magnitude of selection. Our analyses revealed that approximately 25% (seven of 30 comparisons; boldface in tables 2 and 4) of phenotypic selection gradients were biased by environmental covariances between traits and fitness. Because 30 gradients were compared, we would expect an average of 1.5, and at most three, false positives, using an overall significance level of $P < .05$. It thus appears that four to six of the nominally significant differences are true differences, indicating that between 13% and 20% of the phenotypic selection gradients we estimated were biased. Our 30 gradients themselves represent a finite sample of comparisons. Assuming that they represent a random sample of traits, the true proportion of traits with biased phenotypic selection gradients could be as high as 38% (upper confidence limit based on binomial probability).

Among the seven cases of bias that we detected, the sign of the gradient differed in only two cases (directional selection on resistance to leaf damage in *Ipomoea hederacea* and quadratic selection on glucosinolates in *Arabidopsis thaliana*). In both of these cases, one of the estimated gradients is essentially 0, indicating no statistical support for a true difference in sign. This result reinforces the conclusion of the previous section that the two methods yield qualitatively similar results even though they may differ quantitatively.

Directional bias is not evident in our data set; in two of the seven cases of bias, the magnitude of selection was stronger in the phenotypic analysis. Similarly, there was no evident effect of type of selection gradient; three of seven biased estimates were quadratic selection gradients. Finally, there was no clear difference in the frequency of bias between selection acting on plant size (one of 10 estimates were biased) compared with resistance traits (six of 20 estimates; $P = .24$ by Fisher's Exact test).

Discussion

Environmentally Induced Bias and Biological Interpretation

Although a number of authors have suggested that environmentally induced covariances between traits and fitness can sometimes bias estimates of selection gradients derived

Table 4: Summary of the statistical analysis of bias caused by environmental covariances in phenotypic estimates of the pattern of natural selection in *Arabidopsis thaliana*

	Pests present		Pests absent	
Trait	Directional	Quadratic	Directional	Quadratic
Total leaf area:				
1.	$\beta'_p = .08^*$ (.002)	$\gamma'_p = .014^*$ (.001)	$\beta'_p = .10^*$ (.003)	$\gamma'_p = .015^*$ (.002)
2.	$\beta'_g = .36^*$ (.02)	$\gamma'_g = .016$ (.013)	$\beta'_g = .42^*$ (.023)	$\gamma'_g = .007$ (.015)
3.	$\beta_p = .0022^*$ (.00005)	$\gamma_p = .00001^*$ (.000001)	$\beta_p = .003^*$ (.00008)	$\gamma_p = .000009^*$ (.000001)
4.	$\beta_g = .016^*$ (.008)	$\gamma_g = .00003$ (.00003)	$\beta_g = .02^*$ (.001)	$\gamma_g = .000014$ (.0003)
5.	$F = 1.29$, df = 1,832, $P = .26$	$F = 2.35$, df = 1,832, $P = .13$	$F = 1.08$, df = 1,842, $P = .29$	$F = 1.75$, df = 1,842, $P = .18$
Resistance to leaf damage:				
1.	$\beta'_p = -.017^*$ (.003)	$\gamma'_p = -.006^*$ (.003)	$\beta'_p = -.008$ (.005)	$\gamma'_p = -.0055^*$ (.003)
2.	$\beta'_g = -.069$ (.04)	$\gamma'_g = -.092^*$ (.03)	$\beta'_g = -.42$ (.27)	$\gamma'_g = -1.39$ (1.28)
3.	$\beta_p = -.07^*$ (.01)	$\gamma_p = -.09^*$ (.04)	$\beta_p = -.17$ (.1)	$\gamma_p = -2.66^*$ (1.36)
4.	$\beta_g = -.54$ (.28)	$\gamma_g = -5.72^*$ (2.07)	$\beta_g = -3.3$ (2.11)	$\gamma_g = -86.53$ (79.42)
5.	$t = .006$, df = 142, $P = .99$	$t = .0001$, df = 142, $P = .99$	$F = .21$, df = 1,842, $P = .64$	$F = .21$, df = 1,842, $P = .64$
Glucosinolates				
1.	$\beta'_p = \mathbf{-.004\ (.003)}$	$\gamma'_p = \mathbf{.0002\ (.0007)}$	$\beta'_p = -.009$ (.005)	$\gamma'_p = -.0013$ (.001)
2.	$\beta'_g = \mathbf{-.093^*\ (.05)}$	$\gamma'_g = \mathbf{-.078^*\ (.038)}$	$\beta'_g = -.06^*$ (.025)	$\gamma'_g = -.036^*$ (.013)
3.	$\beta_p = \mathbf{-.002\ (.002)}$	$\gamma_p = \mathbf{.00007\ (.0002)}$	$\beta_p = -.003$ (.001)	$\gamma_p = -.00013$ (.0001)
4.	$\beta_g = \mathbf{-.076^*\ (.04)}$	$\gamma_g = \mathbf{-.052^*\ (.03)}$	$\beta_g = -.05^*$ (.025)	$\gamma_g = -.024^*$ (.008)
5.	$F = \mathbf{34.9}$, df = **1,832**, $P = \mathbf{.0001}$	$F = \mathbf{20.62}$, df = **1,832**, $P = \mathbf{.0001}$	$F = .0072$, df = 1,826, $P = .93$	$F = .0007$, df = 1,826, $P = .98$
Trichomes:				
1.	$\beta'_p = \mathbf{-.027^*\ (.003)}$	$\gamma'_p = \mathbf{.003^*\ (.001)}$	$\beta'_p = -.044^*$ (.005)	$\gamma'_p = .0079^*$ (.002)
2.	$\beta'_g = \mathbf{-.13^*\ (.03)}$	$\gamma'_g = \mathbf{.004\ (.024)}$	$\beta'_g = -.20^*$ (.04)	$\gamma'_g = -.033$ (.023)
3.	$\beta_p = \mathbf{-.003^*\ (.0003)}$	$\gamma_p = \mathbf{.00003^*\ (.00001)}$	$\beta_p = -.003^*$ (.0003)	$\gamma_p = .00004^*$ (.000009)
4.	$\beta_g = \mathbf{-.024^*\ (.006)}$	$\gamma_g = \mathbf{.00013\ (.0008)}$	$\beta_g = -.035^*$ (.006)	$\gamma_g = -.001$ (.0007)
5.	$F = \mathbf{37.8}$, df = **1,832**, $P = \mathbf{.0001}$	$F = \mathbf{19.49}$, df = **1,832**, $P = \mathbf{.0001}$	$F = .0000$, df = 1,826, $P = .99$	$F = .0007$, df = 1,826, $P = .98$

Note: Notations and symbols are the same as in table 2. Significant cases of bias are indicated by cells in boldface font.

* $P < .05$ (statistically significant selection gradient).

from the standard Lande-Arnold phenotypic selection analysis (Price et al. 1988; van Noordwijk et al. 1988; Price and Liou 1989; Alatalo et al. 1990; Schluter et al. 1991; Rausher 1992; van Tienderen and de Jong 1994; Mauricio and Mojonnier 1997), few data exist on the prevalence of this type of bias. Our results suggest that such bias may be common: approximately one-quarter of the selection gradients we examined exhibited evidence of environmentally induced bias, and our data are consistent with an actual proportion as high as 38% for similar characters. Moreover, it is likely that these figures are underestimates since the small number of families used in our experiments almost certainly reduced our statistical power to detect significant differences between genotypic and phenotypic selection gradients.

Evolutionary biologists generally use estimates of selection for at least two purposes. One goal is to predict the long-term outcome of natural selection, that is, whether increases or decreases in the values of quantitative traits are expected. For such predictions, only qualitative characterization of the pattern of selection is necessary. The signs of the directional and quadratic selection gradients determine the expected long-term evolutionary equilibrium for the traits. The results of our analysis suggest that for this purpose, phenotypic selection analysis may be preferred because there appears to be little bias in the signs of the gradients it estimates that would offset the advantages of this method. The agreement in the sign of selection gradients estimated using phenotypic and genotypic data, however, should be viewed with some caution until confirmed by further studies.

A second purpose of measuring selection gradients is

Table 5: Number of pairs of unstandardized phenotypic and genotypic selection gradients categorized by sign and statistical significance

	Number of gradient estimates significantly different from 0	
Sign of phenotypic and genotypic gradients	1	0 or 2
Same	15	11
Different	4	0

to predict the rate at which quantitative traits will evolve. For a given genetic architecture, the rate of evolution of a particular trait will generally be proportional to the magnitude of the selection gradient corresponding to that trait (e.g., Lande 1979). Our analysis suggests that for this purpose, environmentally induced covariances between traits and fitness may often introduce biases into the magnitudes of selection gradients estimated by the phenotypic approach. Such biases are not universal; if taken at face value, our analysis suggests that they may exist only about 25% of the time. However, when they exist, they are large. For the seven gradients identified with bias (tables 2, 4), the phenotypic and genotypic selection gradients differ on average by a factor of 17.9 (geometric mean of ratio of larger gradient to smaller gradient). For the four directional selection gradients with bias, the average difference is a factor of 13.8, while for the three quadratic selection gradients with bias, it is 25.4. Moreover, standardization exacerbates the problem of bias. For the seven cases of bias, the standardized gradients differ on average by a factor of 26.7 (the four directional selection gradients differ on average by a factor of 25.2, while for the three quadratic selection gradients with bias, it is 28.9). These results suggest that conclusions and inferences based on quantitative estimates of standardized phenotypic selection gradients (e.g., Conner 2001; Kingsolver et al. 2001) should be made with caution.

Given the potential for environmental factors to affect both phenotypic traits and fitness, it is not surprising to find quantitative differences in the estimates of selection obtained from phenotypic and genotypic methods. The traits we investigated here, plant size and resistance traits, are all likely to be affected by many environmental factors, including light and soil nutrient availability. Moreover, these same environmental factors are also likely to affect fitness in plants. In general, phenotypic selection gradients are least likely to be biased by environmental covariances for situations in which phenotypic variation in traits and fitness is mainly due to genetic rather than environmental variation (i.e., when $V_g \gg V_e$). If environmental variation produces phenotypic variation in either the trait or fitness (but not both), we would expect the power to detect selection gradients to be diminished. However, in this case, the direction and magnitude of phenotypic selection gradients would be largely unbiased by environmental covariances. In contrast, if most phenotypic variation is due to environmental rather than genetic variation ($V_e \gg V_p$), there is a greater opportunity for phenotypic selection gradients to be biased and to reflect the correlation between environment and fitness.

Phenotypic versus Genotypic Selection Analysis: Which to Choose?

In deciding which method to use for estimating the direction and magnitude of natural selection, several competing considerations must be weighed, including ease of application, power to detect selection, and accuracy of the estimated selection gradients. In most cases, the Lande-Arnold phenotypic approach will be favored on the ease-of-application criterion, since it requires no information on family relationships among individuals and thus can be applied to natural populations without genealogical data. Our analyses also suggest that the Lande-Arnold method often has greater power because of the larger number of data points included in the analysis. In addition, because a collection of family means typically has reduced variation compared with the phenotypic data from which they are calculated, the Lande-Arnold method will typically have more power simply because it includes greater variation in the independent variable (the trait of interest). These factors and the absence of environmentally induced bias affecting the sign of phenotypic selection gradients suggest that the Lande-Arnold approach will be the preferred method in many cases. Methods based on phenotypic data also allow investigation of the adaptive value of traits that are closely related to fitness but are not genetically variable, although this approach assumes that the estimated selection gradients are free of environmentally induced bias (see below).

In particular cases, however, the genotypic method is clearly more appropriate. For example, the Lande-Arnold method is not applicable for the analysis of selection acting on reaction norms, a trait that cannot be measured on a single individual (e.g., tolerance to herbivore damage; Mauricio et al. 1997; Tiffin and Rausher 1999; Stinch-

combe 2001). Furthermore, if genetic variation in the trait of interest exists (which can often be detected given sufficient sample sizes), explorations of the adaptive value of traits with genetic data may be superior. For instance, in the case of the alkaloid example discussed in the introduction, phenotypic selection analysis would provide a misleading estimate of the adaptive value of alkaloids. Our analyses, as well as this alkaloid example, present a case for conducting a joint analysis of selection on breeding values and environmental deviations, as described by equation (6). The pattern of selection on environmental deviations (β_2 from eq. [6]) provides investigators with a quantitative estimate of the strength and direction of environmentally induced covariances between traits and fitness. An assessment of the relative strength (and direction) of genetic and environmental covariances between a trait and fitness thus provides a more accurate picture of the adaptive value of traits.

Perhaps most important, our results suggest that whenever estimating the magnitude of selection is critical, the genotypic approach may be more appropriate. Although most estimates of selection may not be biased by environmentally induced covariances, one often cannot determine ahead of time whether this is likely to be a problem for a given trait. If, for example, the magnitude of selection needs to be estimated without bias for three crucial traits, the probability of doing this with the phenotypic analysis is, based on our results, only $(0.75)^3$, or 0.42—less than half. (This probability becomes even less favorable as the number of traits increases.) Ultimately, the difference between the two methods reflects a trade-off between quantitative accuracy and power to detect significance, and investigators will have to determine in each case which method is more appropriate.

Alternative Methods

Several alternative methods exist for minimizing the influence of environmental covariances on estimates of selection, including alternative uses of genetic data and techniques in experimental design and statistical analysis (reviewed by Mauricio and Mojonnier 1997). Among these alternatives, the methods described by van Noordwijk et al. (1988) and Schluter et al. (1991) also depend on the presence of genetic data and thus have many of the same problems of sample size and feasibility associated with genotypic selection analysis.

In the absence of genetic data, environmentally induced covariances can sometimes be minimized by using a spatially replicated experimental design (Mitchell-Olds and Shaw 1987). In many instances, however, the source of microenvironmental variation will be unknown or the relevant spatial scale of variation too small for replication to eliminate environmental covariances (Rausher 1992). Indeed, all of the studies analyzed here used spatially replicated designs, and environmentally induced bias was still present. If the environmental effects on the traits of interest are sufficiently understood, it might also be possible either to include hypothesized environmental factors as covariates in the analysis (e.g., ANCOVA or path analysis) or to manipulate the environmental factors to determine whether both the trait and fitness are affected. Alternatively, it is sometimes possible to manipulate the trait itself (e.g., tail length in birds: Møller 1988; wing color patterns in insects: Kingsolver 1996; Grether 1997), thereby minimizing environmental covariances between the trait and fitness with experimental treatments.

Another potential method of eliminating environmentally induced bias is to include traits in the selection analysis (or path analysis of selection; Scheiner et al. 2000) that one predicts will respond to microenvironmental variation, and thus potentially control for environmentally induced biases statistically. For many studies, this typically involves including a measure of organismal size in the analysis. Obviously, this approach will not work if analyzing selection on size is a goal of the study. Furthermore, we noted several instances where the relationship between plant size and fitness was unbiased by environmental covariances, yet the relationship between other traits and fitness was biased by environmental covariances (e.g., table 4). Thus, the assumption that the relationship between size and fitness is particularly sensitive to environmental covariances and therefore capable of accounting for such covariances statistically is not always valid. Although including additional covariates in analyses increases the likelihood that an environmentally induced covariance will be controlled for statistically, this could come at the expense of experimental effort, statistical power (as more parameters need to be estimated), and uncertainty that the covariates will actually eliminate environmentally induced bias.

Conclusions

It is not clear how generalizable our results are to other systems; the data we have analyzed are not a random sample of all traits in all organisms. We do know that our analyses are in one sense atypical in that they are based on data that come from manipulative field experiments rather than completely natural populations. One sign of the possible peculiarity of our data is that we detected evidence for quadratic selection much more commonly than other studies have found in natural populations (described in Kingsolver et al. 2001). This may simply reflect a greater ease in detecting quadratic selection when environmental effects are randomized in experiments. Nev-

ertheless, we believe it is likely that the conditions under which our experiments were conducted actually minimized deviations between genotypic and phenotypic estimates because they were conducted in environments that had relatively uniform environmental conditions. The relatively uniform experimental conditions in which plants were grown would be expected to minimize environmentally induced variance in general as well as environmentally induced correlations between traits and fitness in particular. By contrast, many of the investigations that have estimated selection gradients acting on ecologically important traits have been conducted in populations living in environments that have not been manipulated to minimize environmental heterogeneity (reviewed by Kingsolver et al. 2001). Organisms measured under these conditions are likely to be exposed to much greater environmental heterogeneity and thus might be expected to exhibit stronger bias in phenotypic selection gradients.

Although our conclusions cannot immediately be generalized to other systems, they do reveal two intriguing patterns relevant to the issue of how selection gradients should be measured: phenotypic and genotypic estimates provide similar qualitative assessments of selection, but the magnitude of phenotypic selection gradients can be substantially biased. Analysis of whether these patterns are characteristic of other systems should eventually allow some conclusion regarding the risks and benefits of each approach.

Acknowledgments

We thank three anonymous reviewers for helpful comments on the manuscript. J. Darrigo, T. Darrigo, L. Reichold, P. Reichold, J. L. Stinchcombe, and W. Stinchcombe provided computer and logistical support. J.R.S. was supported by the Harry S. Truman Scholarship Foundation, M.T.R. was supported by a Wainwright instructorship from Duke University, and P.T. was supported by a National Research Initiative Competitive Grants Program/ USDA award (99-35301-8076).

Literature Cited

Alatalo, R. V., L. Gustafsson, and A. Lundberg. 1990. Phenotypic selection on heritable size traits: environmental variance and genetic response. American Naturalist 135: 464–471.

Arnold, S. J., and M. J. Wade. 1984*a*. On the measurement of natural and sexual selection: applications. Evolution 38:720–734.

———. 1984*b*. On the measurement of natural and sexual selection: theory. Evolution 38:709–719.

Brodie, E. D. III, A. J. Moore, and F. J. Janzen. 1995. Visualizing and quantifying natural selection. Trends in Ecology & Evolution 10:313–318.

Bumpus, H. C. 1899. The elimination of the unfit as illustrated by the introduced sparrow, *Passer domesticus*. Pages 209–226 *in* Biological lectures from the Marine Biological Laboratory of Wood's Hole, Mass. 1898. Ginn, Boston.

Conner, J. K. 2001. How strong is natural selection? Trends in Ecology & Evolution 16:215–217.

Crow, J. F., and T. Nagylaki. 1976. The rate of change of a character correlated with fitness. American Naturalist 110:207–213.

Draper, N. R., and H. Smith. 1966. Applied regression analysis. Wiley, New York.

Endler, J. A. 1986. Natural selection in the wild. Princeton University Press, Princeton, N.J.

Falconer, D. S., and T. F. C. MacKay. 1996. Introduction to quantitative genetics. Longman, Essex.

Fisher, R. A. 1958. The genetical theory of natural selection. 2d ed. Dover, New York.

Grether, G. F. 1997. Survival cost of an intrasexually selected ornament in a damselfly. Proceedings of the Royal Society of London B, Biological Sciences 264:207–210.

Janzen, F. J., and H. S. Stern. 1998. Logistic regression for empirical studies of multivariate selection. Evolution 52: 1564–1571.

Kempthorne, O. 1969. An introduction to genetic statistics. Iowa State University Press, Ames.

Kingsolver, J. G. 1996. Experimental manipulation of wing pigment pattern and survival in western white butterflies. American Naturalist 147:296–306.

Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. American Naturalist 157:245–261.

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

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

Mauricio, R., and L. E. Mojonnier. 1997. Reducing bias in the measurement of selection. Trends in Ecology & Evolution 12:433–436.

Mauricio, R., and M. D. Rausher. 1997. Experimental manipulation of putative selective agents provides evidence for the role of natural enemies in the evolution of plant defense. Evolution 51:1435–1444.

Mauricio, R., M. D. Rausher, and D. S. Burdick. 1997. Variation in the defense strategies of plants: are resistance and tolerance mutually exclusive? Ecology 78: 1301–1311.

Mitchell-Olds, T., and R. G. Shaw. 1987. Regression anal-

ysis of natural selection: statistical inference and biological interpretation. Evolution 41:1149–1161.

Møller, A. P. 1988. Female choice selects for male sexual tail ornaments in the monogamous swallow. Nature 332: 640–642.

Phillips, P. C., and S. J. Arnold. 1989. Visualizing multivariate selection. Evolution 43:1209–1222.

Price, G. R. 1970. Selection and covariance. Nature 227: 520–521.

Price, T., and L. Liou. 1989. Selection on clutch size in birds. American Naturalist 134:950–959.

Price, T., M. Kirkpatrick, and S. J. Arnold. 1988. Directional selection and the evolution of breeding date in birds. Science (Washington, D.C.) 240:798–799.

Rausher, M. D. 1992. The measurement of selection on quantitative traits: biases due to the environmental covariances between traits and fitness. Evolution 46: 616–626.

Rausher, M. D., and E. L. Simms. 1989. The evolution of resistance to herbivory in *Ipomoea purpurea*. I. Attempts to detect selection. Evolution 43:563–572.

Roff, D. A. 1992. The evolution of life histories: theory and analysis. Chapman & Hall, New York.

Scheiner, S. M., R. J. Mitchell, and H. S. Callahan. 2000. Using path analysis to measure natural selection. Journal of Evolutionary Biology 13:423–433.

Schluter, D. 1988. Estimating the form of natural selection on a quantitative trait. Evolution 42:849–861.

Schluter, D., and D. Nychka. 1994. Exploring fitness surfaces. American Naturalist 143:597–616.

Schluter, D., T. D. Price, and L. Rowe. 1991. Conflicting selection pressures and life history trade-offs. Proceedings of the Royal Society of London B, Biological Sciences 246:11–17.

Searle, S. R. 1971. Linear models. Wiley, New York.

Stinchcombe, J. R. 2001. Evolutionary ecology of deer resistance and tolerance in the Ivyleaf morning glory, *Ipomoea hederacea*. Ph.D. diss. Duke University, Durham, N.C.

Stinchcombe, J. R., and M. D. Rausher. 2001. Diffuse selection on resistance to deer herbivory in the Ivyleaf morning glory, *Ipomoea purpurea*. American Naturalist 158:376–388.

Tiffin, P., and M. D. Rausher. 1999. Genetic constraints and selection acting on tolerance to herbivory in the common morning glory *Ipomoea purpurea*. American Naturalist 154:700–716.

van Noordwijk, A. J., J. H. van Balen, and W. Scharloo. 1988. Heritability of body size in a natural population of the great tit (*Parus major*) and its relation to age and environmental conditions during growth. Genetical Research 51:149–162.

van Tienderen, P. H., and G. de Jong. 1994. A general model of the relation between phenotypic selection and genetic response. Journal of Evolutionary Biology 7: 1–12.

Wade, M. J., and S. Kalisz. 1990. The causes of natural selection. Evolution 44:1947–1955.

Weldon, W. F. R. 1901. A first study of natural selection in *Clausilia laminata* (Montagu). Biometrika 1:109–124.

Willis, J. H. 1996. Measures of phenotypic selection are biased by partial inbreeding. Evolution 50:1501–1511.

Editor: Joseph Travis