

COMPARING STRENGTHS OF DIRECTIONAL SELECTION: HOW STRONG IS STRONG?

JOE HEREFORD,^{1,2} THOMAS F. HANSEN,^{1,3} AND DAVID HOULE^{1,4}

¹*Department of Biological Science, Florida State University, Tallahassee, Florida 32306-1100*

²*E-mail: hereford@bio.fsu.edu*

³*E-mail: thomas.hansen@bio.fsu.edu*

⁴*E-mail: dhoule@bio.fsu.edu*

Abstract.—The fundamental equation in evolutionary quantitative genetics, the Lande equation, describes the response to directional selection as a product of the additive genetic variance and the selection gradient of trait value on relative fitness. Comparisons of both genetic variances and selection gradients across traits or populations require standardization, as both are scale dependent. The Lande equation can be standardized in two ways. Standardizing by the variance of the selected trait yields the response in units of standard deviation as the product of the heritability and the variance-standardized selection gradient. This standardization conflates selection and variation because the phenotypic variance is a function of the genetic variance. Alternatively, one can standardize the Lande equation using the trait mean, yielding the proportional response to selection as the product of the squared coefficient of additive genetic variance and the mean-standardized selection gradient. Mean-standardized selection gradients are particularly useful for summarizing the strength of selection because the mean-standardized gradient for fitness itself is one, a convenient benchmark for strong selection. We review published estimates of directional selection in natural populations using mean-standardized selection gradients. Only 38 published studies provided all the necessary information for calculation of mean-standardized gradients. The median absolute value of multivariate mean-standardized gradients shows that selection is on average 54% as strong as selection on fitness. Correcting for the upward bias introduced by taking absolute values lowers the median to 31%, still very strong selection. Such large estimates clearly cannot be representative of selection on all traits. Some possible sources of overestimation of the strength of selection include confounding environmental and genotypic effects on fitness, the use of fitness components as proxies for fitness, and biases in publication or choice of traits to study.

Key words.—Breeder's equation, elasticity, natural selection, phenotypic selection, selection gradient.

Received March 4, 2004. Accepted July 30, 2004.

Lande's realization that directional selection could be approximated as a selection gradient (Lande 1979), together with his proposal of multiple regression as an appropriate technique for estimation of selection gradients (Lande and Arnold 1983), resulted in an explosive increase in empirical estimates of directional selection. As long as some measure of fitness of individuals is available, gradients can be estimated for any quantifiable trait in any taxon. Selection gradients have been used to describe the pattern of selection for estimation of adaptive landscapes (e.g., Arnold et al. 2001) and to test for differences in the pattern of selection in different environments (e.g., by Dudley 1996).

In many of these examples, testing the validity of different hypotheses requires that the magnitudes of selection in different environments or on different traits be compared. This comparison necessitates quantifying selection on a common scale. Almost all such published comparisons have relied on variance-standardized selection gradients or selection intensities, which measure selection in units of standard deviations. For example, Kingsolver et al. (2001) and Hoekstra et al. (2001) reviewed gradients and compared them on a variance-standardized scale (see also Endler 1986). While this is a useful approach to standardization, it has some important disadvantages. The first of these is that variance-standardized selection gradients are functions of population variation and therefore cannot be viewed as descriptors of the fitness function or of the adaptive landscape (sensu Simpson 1944). Second, the variance-standardized gradient offers no clear criteria for determining whether selection is strong. One of the conclusions of Kingsolver et al.'s (2001) review was that

selection was generally not strong, but they did not discuss any criterion for identifying strong selection. Finally, exclusive dependence on one standardization makes it difficult to separate differences due to the quantity of interest (the selection gradient) from those due to the standardizing factor itself.

Our purpose in the present paper is twofold. First we call attention to the advantages of mean-standardized directional selection gradients (Morgan 1999; van Tienderen 2000). Mean-standardized gradients better reflect the shape of the selective surface and provide a natural scale for assessing the strength of selection, although they are not appropriate for all traits. Second, we survey published estimates of directional selection, convert them to mean-standardized gradients, and compare the strength of selection on different types of traits and on different fitness components. This reveals a picture of the strength of selection very different from that suggested previously (Hoekstra et al. 2001; Kingsolver et al. 2001).

Standardized Measures of Selection

Lande (1979) paved the way for a new approach to the measurement of directional selection by decomposing the response to selection as $\mathbf{R} = \mathbf{G}\boldsymbol{\beta}$, where \mathbf{R} is the change in mean of a (vector) trait, \mathbf{G} is the additive genetic variance matrix, and $\boldsymbol{\beta}$ is the selection gradient. Following up on a forgotten suggestion by Pearson, Lande and Arnold (1983) subsequently showed how the selection gradient could be estimated as a set of partial regression coefficients from a

regression of fitness on the trait vector. This method made it possible to assess the strength of direct selection acting on a single trait while controlling for selection on correlated traits. This aspect of Lande’s formulation represented a considerable practical advance, and it has been employed in many empirical studies.

Another fundamentally important benefit of using Lande’s equation is less appreciated: It separates a measure of natural selection (the selection gradient) from the ability of the population to respond. This distinction is easiest to grasp from the univariate version of the Lande equation:

$$R = \sigma_A^2 \beta, \tag{1}$$

where σ_A^2 is the additive genetic variance in the selected trait, and β is the selection gradient, the regression slope of relative fitness, w (fitness standardized to a mean of one), on trait value, z (Lande 1979). The shape of the fitness landscape in the neighborhood the population inhabits is captured by β , whereas σ_A^2 determines the population’s ability to respond to selection, the evolvability. As a regression coefficient, β consists of the ratio

$$\beta = \frac{COV_{w,z}}{\sigma_z^2}, \tag{2}$$

where $COV_{w,z}$ is the covariance between relative fitness and trait z , and σ_z^2 is the phenotypic variance of trait z (Lande and Arnold 1983). Note that the phenotypic variance is the sum of the additive genetic variance and all other sources of variation, which we label the residual variance (σ_R^2). $COV_{w,z}$ is also the selection differential, S , the change in average phenotype within a generation due to selection (Price 1970).

Comparisons of responses to selection are often difficult on the sole basis of equation (1). The response, R , is in trait units and σ_A^2 in trait units squared, whereas β gives the change in relative fitness for a unit change in the trait and thus has units of one per trait. Different traits may be measured in different units, making direct comparisons of R , σ_A^2 , or β fruitless. Even when traits are measured in the same units, the means and variances may be very different, again making comparisons difficult. The obvious solution is to standardize each quantity to remove the units of measure before comparison. Two such standardization schemes are available.

The most familiar standardization uses the standard deviation of the trait. Equation (1) can be rearranged into the familiar breeder’s equation as

$$R = \frac{\sigma_A^2}{\sigma_z^2} COV_{w,z} = h^2 S, \tag{3}$$

where h^2 is the dimensionless narrow-sense heritability and S is the covariance between the trait and relative fitness. Standardization is completed by dividing both sides by the phenotypic standard deviation σ_z

$$\frac{R}{\sigma_z} = \frac{\sigma_A^2}{\sigma_z^2} \frac{COV_{w,z}}{\sigma_z} = h^2 \beta_\sigma. \tag{4}$$

Here, $\beta_\sigma = COV_{w,z}/\sigma_z = \sigma_z \beta$ is the variance-standardized selection gradient (Lande and Arnold 1983), which gives the change in relative fitness for a change of one standard deviation in the trait mean. It also corresponds to i , the intensity

of selection (Falconer and Mackay 1996), the number of standard deviations by which selection changes the trait mean within a generation.

This standardization has one major disadvantage. Although the Lande equation cleanly separates its measure of the fitness landscape from the properties of the population, the breeder’s equation does not. The natural measure of evolvability is the additive genetic variance, σ_A^2 , but the standardization factor σ_z is itself a function of σ_A^2 . Using σ_z as a yardstick is like standardizing leg length by dividing by body length; the resulting quantity may be interesting, but is no longer a measure of leg length. Heritability is thus a singularly misleading measure of evolvability (Houle 1992; Hansen et al. 2003). Similarly, multiplying a measure of directional selection by a function of evolvability yields a hybrid of the two, calling into question the general significance of β_σ . Although the variance-standardized breeder’s equation does have its uses, as discussed below, separation of evolvability and selection is not one of them.

A simple example helps to make the problem clear. Imagine two populations that are at the same point on a linear selective landscape where $\beta = 0.1$ and that have the same residual variance, $\sigma_R^2 = 10$, but where population 1 has a larger σ_A^2 than population 2, say $\sigma_{A,1}^2 = 10$, $\sigma_{A,2}^2 = 5$. Population 1 will respond more to the given selection pressure, and the Lande equation clearly shows why: $\sigma_{A,1}^2 > \sigma_{A,2}^2$. However, one would be misled by interpreting the variance-standardized equation in the same way. The heritability of population 1 (0.5) is greater than that of population 2 (0.33), as expected, but population 1 has larger σ_z than population 2. The two populations therefore differ in the standardization factor, precisely *because* of the difference in σ_A^2 . This difference leads to a larger estimate of selection in population 1 ($\beta_{\sigma,1} = 0.45$) than in population 2 ($\beta_{\sigma,2} = 0.39$), even though this example is constructed to have the same strength of selection. If the additive variances were the same, but the residual variances differed, the confusion generated would be complete. In this second case, both the evolvability and the landscape are the same, but the standardization would cause both h^2 and β_σ to differ. Thus, what the Lande equation has clearly separated the standardized breeder’s equation conflates.

Alternatively, the response to selection can be standardized by the trait mean (Johnson et al. 1955; Houle 1992; Morgan 1999; van Tienderen 2000; Hansen et al. 2003). Recalling that β has units of one per trait, the directional selection gradient can be standardized by multiplying by the trait mean to yield the mean-standardized selection gradient

$$\beta_\mu = \bar{z} \beta = \bar{z} \frac{COV_{z,w}}{\sigma_z^2}. \tag{5}$$

The mean-standardized gradient is the increase in relative fitness for a proportional change in the trait z and thus is an elasticity (Caswell 1989; van Tienderen 2000). It is particularly useful that β_μ is the natural measure of selection in a mean-standardized version of the Lande equation

$$\frac{R}{\bar{z}} = \frac{\sigma_A^2}{\bar{z}^2} \bar{z} \beta = I_A \beta_\mu, \tag{6}$$

where I_A is the square of the additive genetic coefficient of

variation ($CV_A = \sigma_A/\bar{z}$). The symbol I_A (Houle 1992) was chosen by analogy to the opportunity for selection, I (Crow 1958; Arnold and Wade 1984). The opportunity for selection is the squared coefficient of variation of fitness $I = \sigma_w^2/\bar{W}^2$, and gives the maximum possible response of fitness to selection, that is, the response of fitness if its heritability was one. Replacing the phenotypic variance with the additive genetic variance to yield I_A gives the actual response of fitness to selection. In the general case where the trait is not fitness I_A measures the opportunity for response to selection.

Although equation (6) is valid for all traits with nonzero means, the quantities I_A and β_μ only have a natural interpretation on a true ratio scale, where the origin of the scale is not arbitrary. Examples of true ratio traits are fecundity and any measure of size, where a value of zero sets a natural origin for the scale of measurement. An example in which the origin is arbitrary is measurement of timing of biological events in calendar days. In such a case, the trait is not true ratio.

The use of β_μ has several major advantages as a measure of the standardized strength of selection. First, the trait mean is a better standardization factor than the trait variance because neither of the quantities to be compared is a direct function of the mean. This statement must be qualified, in that variances are often correlated with means, and the strength of directional selection will generally depend on the position on an adaptive landscape. Furthermore, the use of any standardization factor will introduce a covariance between evolvability and selection.

Second, the strength of selection on fitness provides a useful benchmark from which to judge the strength of directional selection on other traits. When the selected trait is fitness, it is easy to show that $\beta = 1/\bar{W}$, so the mean-standardized selection gradient for fitness ($\beta_\mu = \bar{W} \times [1/\bar{W}]$) is one (Hansen et al. 2003). This clarifies why I is the opportunity for selection and I_A the actual response of fitness to selection. The β_σ for fitness is equal to the coefficient of variation in fitness ($\beta_\sigma = \sigma_w/\bar{W}$), rather than a constant. The minimum amount of directional selection is obviously no directional selection in either system, which leads to $\beta_\mu = \beta_\sigma = 0$. Thus the mean-standardized system provides a benchmark for both strong selection and the absence of directional selection; in the variance-standardized system the magnitudes of nonzero β_σ values have so far only been interpreted intuitively (e.g. Endler 1986; Kingsolver et al. 2001).

It is important to realize that $\beta_\mu = 1$ is not a maximum strength of selection, but a benchmark for selection that everyone can agree is strong. The fitness landscape itself can of course have areas that are arbitrarily steep, for example, a threshold below which individuals die and above which they survive. In such cases, the estimated gradient will depend on the mean and variance in the measured trait. If the variation is small and the mean is centered on the steepest part of the landscape, β_μ can be much greater than one. As trait variance becomes larger, the rate of change in fitness over the range of the data will become smaller. An area with a $\beta_\mu > 1$ can only occur when the fitness function is nonlinear over the entire landscape, as in the threshold case. If the fitness function is linear with an intercept of zero, then β_μ

= 1 is indeed a maximum, but this is a special case that is unlikely to hold for traits that are not fitness components.

Both standardized measures of selection do have maximum values set by the correlation between fitness and the trait (Arnold 1986). The correlation coefficient, $r_{w,z}$, must have a value between -1 and 1 . Starting from the inequality $|r_{w,z}| = |COV_{w,z}/\sigma_w\sigma_z| \leq 1$, it is straightforward to show that $|\beta_\mu| \leq \bar{z}\sigma_w/\sigma_z = CV_w/CV_z$ and $|\beta_\sigma| \leq \sigma_w = CV_w$. Note that these inequalities imply that there are limits to the proportion of fitness explained by variation in the trait.

The relationships between the dimensionless quantities in equations (4) and (6) are simple functions of the phenotypic coefficient of variation, $CV_z = \sigma_z/\bar{z}$. For the standardized measures of selection, $\beta_\sigma = CV_z\beta_\mu$. For the measures of genetic variation, $I_A = h^2(CV_z)^2$. Therefore, if traits have different phenotypic coefficients of variation, comparisons based on equation (4) and (6) may lead to different conclusions.

For simplicity, our derivations have all been for the univariate case, where selection is estimated on a single trait. Both of these standardization systems extend to the multivariate case, where selection is measured on more than one potentially correlated trait. In such cases, β is a partial regression coefficient (Lande and Arnold 1983), which expresses the projected change in relative fitness for a change in the trait while all other traits are held constant. The interpretation of standardized partial coefficients and their standardized equivalents is otherwise the same as that of univariate regression coefficients. They differ only in the number of other traits for which the effects of indirect selection have been removed—none for univariate estimates and a small number for multivariate estimates. We expect that in most cases a large number of unmeasured traits are directly selected and therefore can exert indirect selection on the measured traits.

Note that we have used terminology and symbols slightly different from those used by Kingsolver et al. (2001). They reserved the term “selection gradient” and the symbol “ β ” for variance-standardized multiple-regression coefficients, that is, estimates derived from datasets in which more than one predictor trait was measured. Kingsolver et al. termed all estimates of gradients based on univariate regressions (regressions with only a single predictor trait) “standardized selection differentials,” which they symbolized i , following Falconer and Mackay (1996). As outlined above, $\beta_\sigma = i$ in the univariate case.

We have not considered the case of nonlinear selection here. Extension to higher-order terms of the fitness function is straightforward. For example, quadratic terms can be converted to mean-standardized forms by multiplying by \bar{z}^2 .

METHODS

We surveyed studies of phenotypic selection published from 1984 through 2003 in *American Journal of Botany*, *American Naturalist*, *Evolution*, *Ecology*, *Heredity*, *International Journal of Plant Sciences*, and *Journal of Evolutionary Biology*. We used the capabilities of ISI’s Web of Science to augment our search. Subsequent to our survey, we obtained the list of studies in Kingsolver et al.’s (2001) review (J. G.

Kingsolver, pers. comm.) and examined all of the studies on it. Data were included in our dataset if the study met several criteria. First, we included studies that estimated univariate or multivariate selection. In the case of univariate selection, we included studies that measured selection by direct comparison of trait means before and after selection and those that calculated a univariate selection gradient by linear regression. These measures comprise the univariate dataset. We also included studies that estimated multivariate selection using the standardized selection gradient of Lande and Arnold (1983). Second, studies had to be conducted in the field and use natural populations; studies that used phenotypic manipulations were disqualified. Third, so that we could calculate mean-standardized gradients from regression coefficients and selection gradients, studies had to include the means and standard deviations of the trait of interest for the sample used to calculate the regressions. In a few cases, unpublished means and variances were obtained directly from the authors. Finally, we excluded traits measured on scales that were not true ratio, with the exception of traits measured on a log scale. To a first approximation, the (univariate) selection gradient on $\ln(z)$ is equal to the mean-standardized gradient of z divided by $1 + (I_p/2)$, where I_p is the mean-standardized phenotypic variance of z (and I_p itself is approximately equal to the variance of $\ln[z]$, assuming a locally linear fitness function). We located three studies that both measured selection for a trait on a log scale and provided the other necessary statistics.

To calculate mean-standardized selection gradients, we multiplied regression coefficients of trait on fitness by the trait mean. For studies that only reported the β_w , we multiplied by \bar{z}/σ_z to obtain β_w . Where univariate selection was measured, we sometimes obtained the selection differentials directly from the means and variances provided in the articles and calculated gradients from them. We analyzed the data for univariate and multivariate selection separately and then looked for differences between the two measures. We analyzed the data in this way because of the differences in interpretation of univariate and multivariate selection and to allow comparison of the strength of selection when these measures were used. Although on average univariate selection measures will be larger than multivariate ones when traits are correlated, in some cases they will be smaller. To check the legitimacy of estimates of β , we calculated the correlation between fitness and the trait as $r_{wz} = \beta\sigma_z/\sigma_w$.

We recorded several additional variables for each estimate. The fitness component used to measure selection was classified as viability, fecundity, or sexual selection. Only survival or survivorship was considered viability, and fecundity was limited to the number of offspring. Two studies included a combination of viability and fecundity as the fitness component. These studies were not included in one- and two-way analyses of differences in selection measures or groupings by type of fitness component. Female choice, sperm and pollen competition, and male-male competition were all included as sexual selection. In some studies, different measures of fitness were regressed on the same sets of traits in the same population or year. For these studies, we only included the estimates of selection that used the most comprehensive measure of fitness, to make sure that the same episodes of se-

lection were not repeatedly analyzed. The traits involved were coded as morphological or life-history traits. Only life span, timing of reproduction, and number of flowers were considered life-history traits. We did not include characters made up of dates of events such as flowering or germination because dates are not on a ratio scale. One study included behavioral traits, but these were not included in analyses of type of trait because they were from a single study and comprised only four estimates. For each estimate we recorded the sample size, the level of statistical significance, and the standard error of the estimate where they were available.

For our purposes, the signs of selection estimates are arbitrary, so we analyzed the absolute values of the estimates. Two-way analysis of variance was performed in the SAS procedure GLM (SAS, ver. 8.1; SAS Institute, Cary, NC) using type of trait and fitness components as independent variables and the selection measure as the dependent variable. Significance of these analyses was judged relative to the distribution of F -ratios from randomized datasets. The estimates and sample sizes were randomly shuffled 1000 times relative to the independent variables, preserving the pairing of estimate and sample size. Each randomized dataset was analyzed with GLM and the F -ratio obtained. The SAS programs of Cassell (2002) were used to perform the randomization tests, after some modification. These tests are all approximate, as they all treat each selection estimate as independent. In fact, each set of estimates for the same population at the same time are correlated with each other, but because the vast majority of comparisons are between estimates from different populations and studies, we suspect that this inaccuracy is small.

Correcting for Bias in Absolute Values

When comparing the strengths of selection on different traits and in different studies, we naturally wanted to compare the absolute value of selection. Unfortunately the absolute values of selection gradients are biased upward whenever the confidence limits of the estimate overlap zero. To see this, imagine studying a set of traits that are not under directional selection. With estimation error, the estimates of β will not equal zero, although they can be expected to average zero. When the absolute values of the resulting standardized estimates are calculated, they will average more than zero, as they must in this case be nonnegative, giving a misleading impression of the strength of selection. Omitting estimates whose confidence intervals overlap zero is not a solution, as doing so would also cause an upward bias in the estimates.

To correct for this bias we began with the assumption that the estimated selection slope, β , is normally distributed with a mean, b , equal to the true value of the selection gradient and a variance, s^2 , that we assume to be equal to the square of the reported standard error. It therefore follows that the variable $|\beta|/s$ follows a chi distribution (i.e., the distribution of the square root of a chi-square-distributed variable) with one degree of freedom and noncentrality parameter $(b/s)^2$ (Evans et al. 2000). On this basis we can compute the expected value of $|\beta|$ as

$$E[|\beta|] = s \left\{ \sqrt{\frac{2}{\pi}} \text{Exp} \left[\frac{-(b/s)^2}{2} \right] + \left| \frac{b}{s} \right| \text{Erf} \left(\frac{|b|}{s\sqrt{2}} \right) \right\}, \quad (7)$$

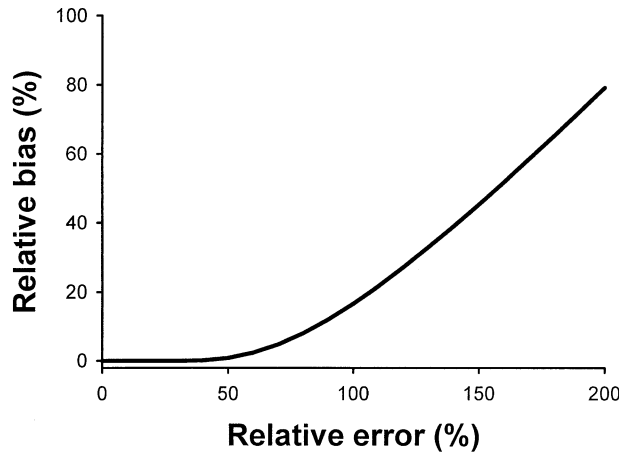


FIG. 1. Relative bias of absolute values of selection estimates, shown as a percentage of the estimate, as a function of the relative error of the regression slope (the standard error of the estimate as a percentage of the estimate).

where $\text{Erf}[x] = (2/\sqrt{\pi}) \int_0^x \text{Exp}[-t^2] dt$ is the error function. If $b = 0$, then $E[|\beta|] = s\sqrt{2/\pi}$. If the relative error, $s/|b|$, is small, $E[|\beta|] \approx b$.

Figure 1 shows the relative bias (bias measured in percentage of the mean) plotted against the relative error of the regression slope. The bias is very large when the relative error exceeds 100% (i.e., when the standard error is as large as the estimated slope) but decreases rapidly at lower values. When the relative error is 50%, the bias is less than 1% of the mean, and it decreases very rapidly below that. Therefore, as a rule of thumb, the bias is very small for selection gradients that are significantly different from zero.

Many of the estimated selection gradients in our sample have relative errors in excess of 100%, and in many cases these estimates are quite large. In these instances the bias has a genuine impact on our results. As a partial correction for the bias, we substituted the estimated β for b in equation (7) to get an estimate of the bias as

$$\text{bias} = E[|\beta|] - |\beta| \tag{8}$$

for those studies that gave the necessary standard error. This bias is underestimated, as the true b is smaller in expectation than the estimated β used to derive the bias. This remaining bias cannot be removed in any obvious way, as further iteration will not converge and may lead to an overestimate of the bias. Before computing standardized gradients, we corrected the absolute values of the selection gradients by subtracting the bias as computed by equation (8). In cases where the bias exceeded 100% of the mean, we adjusted the absolute selection gradient to zero. This procedure can be justified by Bayesian reasoning, as it is reasonable to use a prior with zero weight on negative selection strength.

RESULTS

We were able to locate only 38 studies that included all the information necessary for calculation of mean-standardized gradients and met our other criteria for inclusion (see Appendix available online at <http://dx.doi.org/10.1554/04.147.1.s1>). They furnished a total of 340 multivariate es-

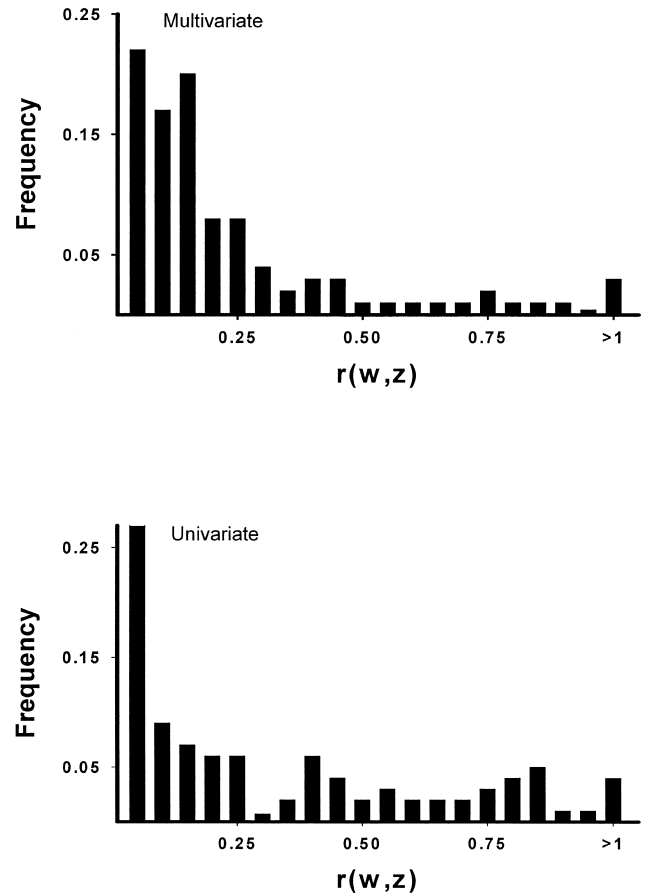


FIG. 2. Absolute values of the correlations between fitness measures and trait values. Top, partial correlations based on multivariate estimates; bottom, correlations based on univariate selection estimates.

timates of selection (multivariate defined as estimates derived by multiple regression with at least two traits) and 240 univariate estimates. This is less than a third of the number that Kingsolver et al. (2001) reported, despite our inclusion of more recent work. The difference arises primarily because published estimates often do not include the means and variances necessary for calculation of mean-standardized gradients, β_μ .

For 81% of the multivariate estimates and 62% of the univariate estimates, the variance in relative fitness was also published, allowing us to calculate the correlations between fitness and the trait, r_{wz} . Figure 2 shows the absolute values of these correlations or partial correlations. The median multivariate r_{wz} was 0.11, and the median univariate r_{wz} was 0.23. Six univariate r_{wz} exceeded the maximum possible value of one, as did nine of the multivariate estimates. These results could only be due to errors in calculating selection estimates, so we eliminated these estimates before further analysis. The square of r_{wz} is the proportion of variation in fitness explained by each trait.

Mean-Standardized Gradients

The distribution of absolute values of mean-standardized selection gradients, β_μ , is shown in Figure 3. Both the uni-

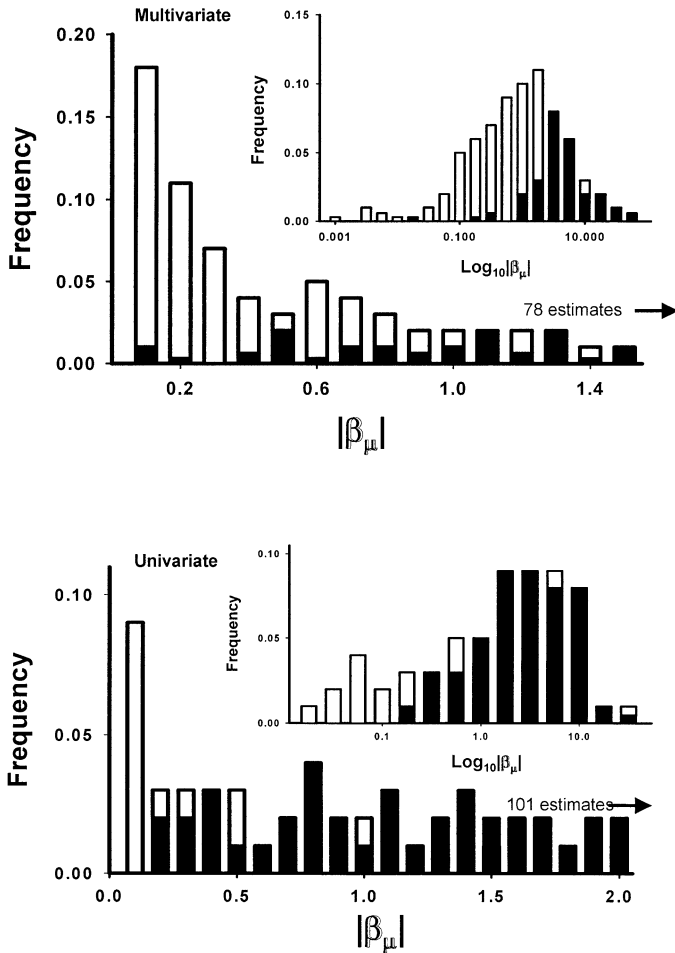


FIG. 3. Distribution of absolute values of mean-standardized gradients ($|\beta_\mu|$). Top, multivariate $|\beta_\mu|$; bottom, univariate $|\beta_\mu|$. Large panels plot the data on an arithmetic scale; insets plot the same data on a \log_{10} scale. Open bars denote estimates that are not significantly different from zero; closed bars denote significant values.

variate and multivariate distributions had modes near zero and very long tails of large estimates. The median uncorrected multivariate β_μ was 0.54. Therefore, more than half the traits studied were under selection at least half as strong as selection on fitness itself. Twenty-five percent of the multivariate β_μ were greater than 1.34; the maximum had an absolute value of 25.18. Univariate β_μ were substantially larger, with a median of 1.45, suggesting selection much stronger than the selection on fitness; the 75% quantile was 3.72. To get a better idea of the effect of controlling for at least some correlated traits, we compared estimates for which both a univariate and a multivariate estimate were available. The multivariate and univariate medians were similar (1.00 and 1.06, respectively).

The very large sizes of the median β_μ raise the possibility that the estimates were dominated by estimates with large standard errors and upwardly biased absolute values. Correcting for bias due to use of absolute values decreases the multivariate β_μ substantially. Taking only studies where the standard errors of the gradients are available, the median multivariate β_μ was 0.50 before correction (close to that of

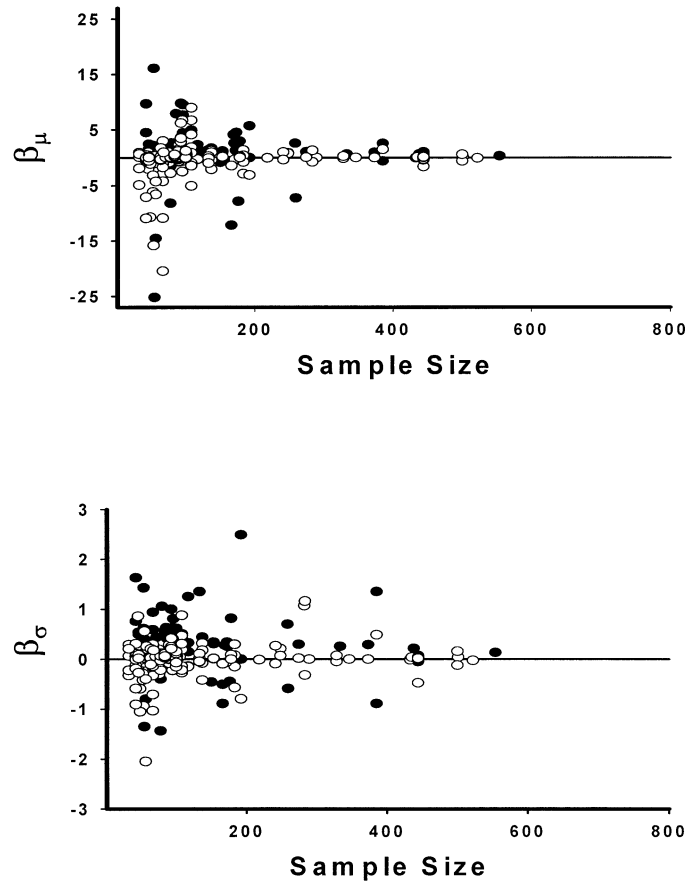


FIG. 4. Plots of sample size versus selection gradient. The reference line is plotted at zero. Top, plot of sample size and multivariate β_μ ; bottom, that of sample size and univariate β_μ .

all estimates) and only 0.31 after correction, a decrease of 38%. The vast majority of the difference was due to estimates that were not significantly different from zero (shown as open symbols in Fig. 3). The long tail of extremely large values, most of which are significant, was less affected by the bias correction. The larger quantiles of the bias-corrected distribution were thus less affected; the 75% quantile for example was 1.04, only 22% less than the value based on all the multivariate estimates. Less than 20% of the univariate estimates had the standard errors necessary for calculation of a bias correction, and these estimates were substantially smaller than the other univariate estimates, having a median value of only 0.48. Bias correction resulted in a much smaller reduction of only 6%.

Another window on this phenomenon is provided by funnel plots of multivariate β_μ values as a function of sample size, shown in the upper part of Figure 4. The magnitude of estimates is reduced to some extent for estimates of large sample size, but many of the estimates for the largest sample size, consistent with this result, the sample-size-weighted mean multivariate β_μ is 0.97, considerably larger than the median.

Examination of Table 1 suggests that life-history traits have values of β_μ almost twice as large as those of morphological traits in the multivariate dataset, but the difference is

TABLE 1. Median absolute value of mean-standardized selection gradients, $|\beta_\mu|$, categorized by the type of trait studied (life history or morphological) and by the fitness measure used (fecundity, mating success, or viability). Sample sizes are given in parentheses after each estimate.

	Raw $ \beta_\mu $			Bias corrected $ \beta_\mu $		
	Mult. ¹	Univ. ²	Combined ³	Mult.	Univ.	Combined
Life history	0.86 (42)	0.28 (30)	0.43 (61)	1.19 (13)	0.18 (21)	0.66 (33)
Morphological	0.53 (296)	1.90 (206)	0.72 (383)	0.28 (189)	0.89 (28)	0.29 (190)
Fecundity	0.38 (191)	0.74 (54)	0.37 (215)	0.25 (128)	0.31 (27)	0.23 (148)
Mating	0.76 (107)	1.93 (146)	1.59 (177)	0.39 (56)	0.75 (21)	0.39 (56)
Viability	0.48 (24)	0.70 (32)	2.83 (38)	0.49 (17)	5.70 (1)	0.57 (18)
All estimates	0.54 (340)	1.45 (240)	0.68 (448)	0.31 (193)	0.48 (49)	0.30 (227)

¹ Denotes estimates based on multiple regressions.

² Univariate estimates.

³ The combined dataset uses a single estimate for each population/trait combination available, whether univariate or multivariate. If both multivariate and univariate estimates are available for the same trait and population, the multivariate estimate with the most covariate traits is used.

almost fourfold in the opposite direction in the univariate dataset. Sexual (mating) selection appears stronger than fecundity or viability selection, whereas viability selection seems strongest in the combined dataset. Despite the large sizes of some of the differences, our randomization tests sug-

gest that none of these differences are statistically significant, except with regard to trait type in the combined dataset. There morphological traits had significantly larger elasticities than life-history traits ($P = 0.03$).

Variance-Standardized Gradients

The distribution of variance-standardized selection gradients, β_σ , is shown in Figure 5. Like the β_μ , selection gradients have modes at zero and long tails of large values. The median values across the dataset are shown in Table 2. The median uncorrected value of multivariate β_σ was 0.09, and the 75% quantile was 0.29. This median is substantially lower than the comparable value of 0.16 reported by Kingsolver et al. (2001). Our requirement for more published information resulted in a sample of studies showing weaker selection. The univariate estimates had a median of 0.19. When the 95 observations with both univariate and multivariate estimates were compared directly, the univariate median was only slightly lower (0.17 vs. 0.21). Correcting for bias decreased the multivariate gradients by 33%, from 0.09 to 0.06. Bias correction of the relatively small number of univariate β_σ with the necessary data resulted in a median of 0.15.

A funnel plot of multivariate β_σ values is shown in the lower part of Figure 4. There are many larger estimates with large sample sizes. Consistent with this, the sample-size-weighted mean β_σ is 0.16, larger than the median estimate (0.09).

Variance-standardized gradients showed inconsistent relationships with trait type and fitness measure depending on whether the multivariate, univariate, or combined datasets were used, as was the case for the β_μ above. The randomization tests showed that the interaction between trait type and fitness measure was just significant at $P = 0.03$ for the uncorrected multivariate dataset. No other effects were significant across all six data partitions.

Comparison of Mean- and Variance-Standardized Gradients

The relationship between multivariate β_μ and β_σ is shown in Figure 6. The values of β_μ and β_σ had modest Pearson's correlations that were significantly different from both zero and one by bootstrapping ($r_p = 0.63$, $N = 340$, upper 95% limit 0.70, lower 95% limit 0.57) but larger rank correlation ($r_s = 0.87$, upper 95% limit 0.89, lower 95% limit 0.84). The

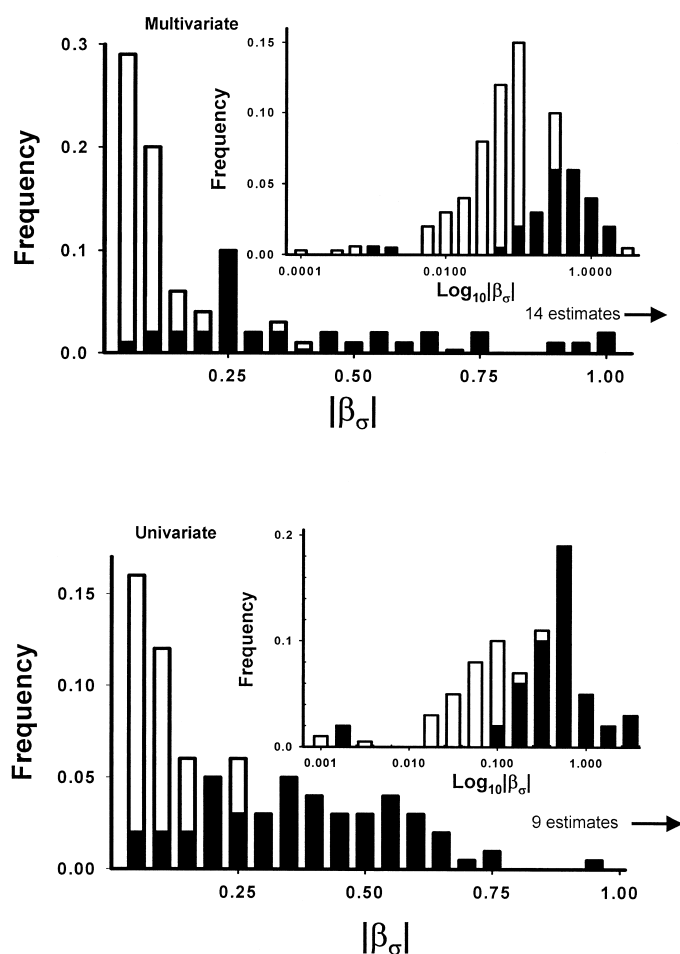


FIG. 5. Distribution of absolute values of variance-standardized selection gradients, $|\beta_\sigma|$. Top, multivariate $|\beta_\sigma|$; bottom, univariate $|\beta_\sigma|$. Large panels plot the data on an arithmetic scale; insets plot the same data on a \log_{10} scale. Open bars denote estimates that are not significantly different from zero; closed bars denote significant values.

TABLE 2. Median absolute value of variance-standardized selection gradients, $|\beta_\sigma|$, categorized by the type of trait studied and by the fitness measure used. Presentation is as in Table 1.

	Raw $ \beta_\sigma $			Bias corrected $ \beta_\sigma $		
	Mult.	Univ.	Combined	Mult.	Univ.	Combined
Life history	0.25 (42)	0.30 (30)	0.18 (61)	0.35 (13)	0.09 (21)	0.30 (33)
Morphological	0.09 (296)	0.19 (206)	0.11 (383)	0.06 (190)	0.17 (28)	0.06 (191)
Fecundity	0.08 (191)	0.21 (54)	0.08 (215)	0.06 (128)	0.14 (27)	0.06 (148)
Mating	0.18 (107)	0.19 (146)	0.19 (177)	0.15 (56)	0.15 (21)	0.15 (56)
Viability	0.10 (24)	0.07 (32)	0.19 (38)	0.11 (18)	0.18 (1)	0.14 (19)
All estimates	0.09 (340)	0.19 (240)	0.12 (448)	0.06 (207)	0.15 (49)	0.07 (228)

correlations of the 240 univariate estimates were somewhat lower but still different from both zero and one ($r_p = 0.28$, upper 95% limit 0.38, lower 95% limit 0.20; $r_s = 0.71$, upper 95% limit 0.75, lower 95% limit 0.63).

DISCUSSION

Standardized Measures of Selection

Our purpose in the present paper is twofold. First we call attention to two standardized forms of the Lande equation for the response to selection (eq. 1, $R = \sigma_A^2 \beta$), one based on standardizing the fundamental parameters using the trait variance (the breeder's equation, $R/\sigma = h^2 \beta_\sigma$) and the other on standardizing by the trait mean ($R/\bar{z} = I_A \beta_\mu$). Second, we review estimates of directional selection using the mean-standardized selection gradient, β_μ . Mean-standardized gradients give proportional changes in fitness for a proportional change in trait value and are thus elasticities. Our review makes clear that, if these estimates are taken at face value, the average strength of directional selection observed in nature is extremely strong. These estimates are in aggregate so large that we doubt that they can be considered typical for reasons that we discuss below.

Previous reviews of the strength of directional selection have used the variance-standardized selection gradient, β_σ ,

the measure of directional selection appropriate to the breeder's equation. We outlined why we believe that β_σ is not an appropriate general measure of the strength of selection, although it is informative in other ways, as demonstrated by the reviews of Kingsolver et al. (2001; see also Hoekstra et al. 2001). For example, if one is interested in predicting the number of standard deviations that a trait will change given an estimate of selection, β_σ is ideal. The variance-standardized selection gradient is also directly related to statistical power. The availability of β_σ values allowed Kingsolver et al. (2001) to show that the power of most studies to estimate typical strengths of selection was very low.

What the variance-standardized selection gradient does measure is the rate of change in fitness per standard deviation change in the trait. Thus, the higher β_σ is, the bigger the fitness difference predicted between extreme individuals in the population and the higher the selection load. We therefore suggest that β_σ be referred to as the "population strength" of selection to emphasize its applicability to a particular population on a selective landscape.

In contrast, mean-standardized gradients, β_μ , tell us about the fitness consequences of proportional changes in trait means, without regard to the variation within the population. This procedure makes biological sense for traits measured on a true ratio scale, where zero is a meaningful limit to phenotypic value, denoting a true absence. A key advantage of mean-standardized gradients is that the β_μ of relative fitness is one, a useful benchmark from which to judge the strength of directional selection on other traits. Mean-standardized gradients are thus likely to be on a biologically meaningful scale for any true ratio trait. We suggest that β_μ be referred to as the "landscape strength" of selection to emphasize that it measures the slope of the fitness landscape at the population mean in proportional units.

Mean-standardized gradients have some clear advantages. One is ease of interpretation, as we have emphasized. Second, selection on characters that have drastically different distributions such as multinomially distributed characters, which might take on only a few values, and continuous characters, which can take on an infinite number of values, can be compared on a proportional scale. Comparing selection on these types of characters is more difficult in units of standard deviation because the variances of these distributions are drastically different. Third, means are more easily estimated than variances, so the error introduced by standardizations themselves is reduced. Mean-standardized gradients can be interpreted as elasticities (van Tienderen 2000), linking them to

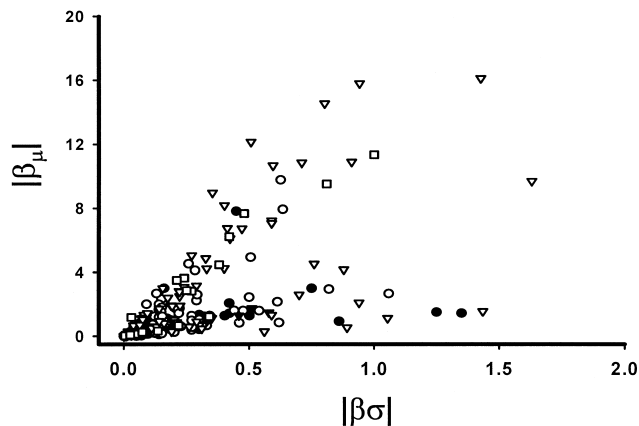


FIG. 6. Scatter plot of the relationship between multivariate β_μ and multivariate β_σ , grouped by type of trait and fitness measure. Solid circles, fecundity selection on life-history traits; open circles, fecundity selection on morphological traits; solid triangles, sexual selection on life-history traits; open triangles, sexual selection on morphological traits; solid squares, viability selection on life-history traits; open squares, viability selection on morphological traits.

a large literature on demographic elasticities (Caswell 1989). Finally, mean-standardized gradients provide a link between selection and evolvability, measured as mean-scaled additive genetic variances (I_A ; Houle 1992; Hansen et al. 2003), all on compatible scales.

The relative usefulness of the two standardizations rests in part on the investigator's intuition about whether a change in trait means or trait variances is more biologically significant or more likely, for example, due to environmental changes. Consider a series of populations subject to the same simple linear fitness function, $w(z) = a + bz$. On the one hand, populations could be compared in which the mean is the same but phenotypic variances differ. In this case, populations with small variances will have low β_σ values, which might be taken to mean that the fitness function itself differs, when in fact only the population variance does. On the other hand, if populations are compared that have different means but the same variance, populations with smaller means will have lower β_μ , again perhaps giving the impression of differences in the fitness function when none exists. Any standardization scheme depends on both the properties of interest (in this case the strength of selection) and the differences in the standard used.

The above scenario is of course unrealistic in many respects. Real differences between populations are likely to include differences in both means and variances, as these are often related to each other. Similarly, fitness functions will rarely be linear, so the slope of the fitness function is likely to change with population means. No simple standardization scheme will serve all purposes or address all problems. Simultaneous use of more than one approach helps to guard against overinterpretation of results based on any one measure.

Strength of Selection

Our review of mean-standardized gradients, β_μ , is notable for the extremely strong selection observed overall. The median β_μ of 0.54 means that doubling the value of the average trait increases fitness by 54%. Judged against the scale of the strength of selection on fitness, it is also 54% as strong as selection on fitness. This median is biased upward because of our dependence on the absolute value of β_μ as a summary measure of selection. Even when this bias is corrected to the extent possible, the median β_μ is still 0.28, or 28% of the strength of selection on fitness.

This conclusion is in sharp contrast to that of Kingsolver et al. (2001; see also Hoekstra et al. 2001), who concluded that directional selection on most traits is weak based on their summary of variance-standardized selection gradients, β_σ . Kingsolver et al.'s conclusion that selection is usually weak is not explicitly justified by them, but close reading of their paper suggests that it is based at least partially on the fact that the estimates have a mode at zero. The difference between their conclusions and ours does not result from our use of a different and smaller set of studies. The median β_σ that we observed was 0.09 (before bias correction), almost 50% less than the value of 0.16 reported by Kingsolver et al. This difference implies that, had we been able to include

all the estimates of selection in their review, our median β_μ would have been even larger.

Consideration of β_σ estimates in light of what they actually measure—the rate of change in fitness within the range of the population—suggests that they can also be seen as large. Most modest-sized populations range over at least four standard deviations, so it is reasonable to compare the fitness of individuals this many standard deviations apart. With fitness standardized to a mean of one, the median β_σ of 0.16 indicates that these extreme individuals will differ by 0.64 in relative fitness. This means that the least fit individuals will be at least 50% less fit relative to the best phenotype in the population, and that mean fitness is reduced by about 25%. Our best estimate of the median β_σ is considerably lower at 0.06, but it still indicates a range of fitnesses of 20% due to the typical trait and a reduction in mean fitness due to each trait of 11%.

These estimates of selection are so large that they call into question the representativeness or meaning of this sample of estimates. The typical study in our sample addressed only a modest number of traits, averaging 3.5. If these are considered to be a random sample of traits, then it is very difficult to see how a population could sustain directional selection of this magnitude on even 50 such traits, much less the nearly infinite number of characters into which an individual phenotype could be decomposed. This paradox could be resolved in several nonexclusive ways. First, the choice of study systems may be biased toward those in which strong directional selection is occurring. Second, the choice of traits for study might be skewed toward those under the strongest selection. Third, there may be publication bias. Fourth, the dimensionality of the phenotype may be low enough that selection acts on only a small number of causally independent traits. Fifth, the estimates that are obtained may be biased upward as a result of covariance between environment and phenotype. Sixth, the estimates are based on fitness components that may be poorly correlated with fitness. If any of these potential explanations is correct, one cannot directly infer a typical strength of selection from the available data.

The first three potential explanations involve a preference on the part of investigators for studying (or publishing) systems and traits under strong selection and suggest that a randomly chosen trait in a randomly chosen population would typically be under much weaker directional selection. We attempted to determine which traits were chosen for study “at random” and which because of some a priori expectation that they would be under selection. In many cases, an a priori expectation was clearly evident. For example, many studies of sexual selection looked at ornaments that clearly imply an expectation of mate choice. In the majority of studies, however, authors did not clearly state whether they considered some a priori prediction. A few studies estimated selection on a very large number of traits, so the investigators seem very unlikely to have had an expectation about selection on each one. It seemed likely to us that a preference for studying strong selection is at least part of the explanation for the very high average strength of selection, but contrary to this expectation, Kruskal-Wallis tests revealed no significant differences based on the a priori expectations expressed in the

published papers. This result may reflect a failure to express those expectations rather than a lack of investigator choice.

Kingsolver et al. (2001) reported some modest evidence for publication bias for linear selection gradients, that is a failure to publish nonsignificant results. The low power of most studies of phenotypic selection (Kingsolver et al. 2001; Hersch and Phillips 2004) certainly gives a great deal of opportunity for such bias to operate. However, the vast majority of the estimates in our dataset are from studies that estimate selection on many traits or in many populations, lessening the opportunity for the significance of any particular estimate to influence its publication.

The fourth possible explanation—that effectively only a few axes of independent phenotypic variation are important to organism fitness—implies that even a fairly haphazard choice of traits to measure could capture variation along all these axes without study of large numbers of traits. The multiple-regression approach would then spread the impact of selection over those traits that indicate each axis. In this case, we predict that the total strength of directional selection would continue to fall as more and more traits were added. In an extremely simple hypothetical example, all real selection might be due to variation in organism size. The size of almost any morphological character would then detect this selection, but if more and more morphological traits were measured, on average, less and less selection would appear to affect each one. There is evidence that this occurs between estimates based on one trait (the univariate estimates), and those based on more than one (Tables 1, 2). Within the multivariate studies, however, there is no evidence for this effect. The rank correlations of number of traits in a study with strength of selection are actually positive although not significantly different from zero (0.11 for β_μ and 0.20 for β_σ).

The fifth potential explanation for the very strong selection detected is that the estimates are themselves biased upward. A likely source of such a bias is environmentally induced covariance between the trait and fitness (Rausher 1992). Environmental covariance can artificially increase estimates of phenotypic selection in the following way: If sites or territories vary in their suitability to support growth, and therefore in the fitnesses of individuals that inhabit them, the environment and fitness covary, which can create covariance between the phenotype and fitness that will be detected as directional selection by the methods of Lande and Arnold (1983). There is experimental evidence that such biases can be substantial (Stinchcombe et al. 2002; Winn 2004). Methods for eliminating environmental covariance have been described (Price et al. 1988; Rausher 1992; Scheiner et al. 2002), but these are neither perfect nor commonly used. Environmental covariance is likely to have influenced some of the estimates of selection reviewed here.

Finally, we note that selection is measured relative to a fitness component and not to true fitness. Fitness components, such as viability or fertility, represent only a part of the total selection and are usually measured in ways that cover only a small part of the life history of the organism. Furthermore, there are theoretical reasons to believe that fitness components are sometimes negatively genetically correlated with each other (Robertson 1955; Charlesworth 1990; Reeve et al. 1990; Houle 1991), such that the effects of selection gen-

erated through one component are reduced or nullified by selection on other components. The empirical evidence is, however, equivocal, and on the phenotypic level, positive correlations may dominate (for reviews see Houle 1991; Roff 1992; Stearns 1992). In any event, the correlation between fitness components and true fitness is far from perfect, and if estimated mean-standardized gradients or intensities are to be interpreted as measures of total selection acting on the trait, rather than as measures of the selection generated by the individual fitness component in question, they are severe overestimates. To get a feel for this bias, we may use the fact that elasticities follow a chain rule; the elasticity with respect to total fitness is the product of the elasticity with respect to the fitness component multiplied with the elasticity of the fitness component with respect to total fitness:

$$\frac{d \ln W}{d \ln z} = \sum_i \frac{\partial \ln W}{\partial \ln F_i} \frac{d \ln F_i}{d \ln z}, \quad (9)$$

where W is fitness, F_i are fitness components, z is the trait, and the sum is over all fitness components affected by the trait. The fitness components used in the studies we have reviewed may differ considerably in their effect on total fitness.

In their review of elasticities of avian life-history variables with respect to population growth rate (λ , computed from life tables), which we can take as a better estimator of total fitness, Sæther and Bakke (2000) found that the elasticity of adult survival to λ was in the range 0.35 to 0.95, whereas the elasticity of fecundity to λ ranged from 0.05 to 0.65. Multiplying our median bias-corrected β_μ (for combined multivariate and univariate estimates) with respect to viability, 0.57, by the approximate mean elasticity of viability with respect to λ from Sæther and Bakke (0.6, estimated from their fig. 1) produces an average β_μ with respect to λ of approximately 0.40. More dramatically, multiplying our median bias-corrected β_μ for fecundity, 0.23, by Sæther and Bakke's value of 0.25 produces a β_μ of 0.06 with respect to λ . These numbers tell us that typical selection on the trait, as mediated through viability and fecundity, respectively, would be 40% and 6% as strong as selection on fitness. Although this selection can still be characterized as moderately strong, the numbers seem much more realistic than the naive, uncorrected numbers of 283% and 37%, respectively, that we get if both the bias from use of fitness components and the estimation bias were ignored.

On the other hand, if each selective episode in the life cycle results in selective death, the load induced by selection may still be quite substantial, even when net selection is weak. This is the sort of selection that favors a plastic response to environmental conditions. Clearly, the task of inferring a typical strength of selection is fraught with pitfalls. Careful consideration must be given to the choice of study populations, traits to study, the covariances among traits, and the measure of fitness, as well as to the causes of fitness variation.

Finally, we note that the data gathered in this review relied on the publication of trait means, variances, and clearly described methods of data analysis. Most published studies are lacking some of these fundamental details. We urge authors, reviewers, and journal editors to insist on the publication of

these important details along with the main conclusions of a study. As is the case in so many other areas of evolution and ecology, haphazard standards of reporting and the lack of attention to the meaning of estimates greatly impairs our ability to generalize on the basis of past work.

ACKNOWLEDGMENTS

We thank D. J. Fairbairn, J. Maad, and A. A. Winn for sharing unpublished data; J. G. Kingsolver for furnishing his compilation of gradient estimates; M. Morgan for making available his unpublished manuscript; and C. Fenster, J. G. Kingsolver, F. Knappczyk, M. Morgan, J. Travis, A. A. Winn, and two anonymous reviewers for comments on the manuscript. This work was supported in part by National Science Foundation grant DEB-0129219 to DH.

LITERATURE CITED

- Arnold, S. J. 1986. Limits on stabilizing, disruptive, and correlational selection set by the opportunity for selection. *Am. Nat.* 128:143–146.
- Arnold, S. J., and M. J. Wade. 1984. On the measurement of natural and sexual selection: theory. *Evolution* 38:709–719.
- Arnold, S. J., M. E. Pfrender, and A. G. Jones. 2001. The adaptive landscape as a conceptual bridge between micro- and macro-evolution. *Genetica* 112/113:9–32.
- Cassell, D. L. 2002. A randomization-test wrapper for SAS PROCs. SAS Users' Group International Proceedings 27:251. <http://www2.sas.com/proceedings/sugi27/p251-27.pdf>. Accessed 27 February 2004.
- Caswell, H. 1989. Matrix population models. Sinauer Associates, Sunderland, MA.
- Charlesworth, B. 1990. Optimization models, quantitative genetics, and mutation. *Evolution* 44:520–538.
- Crow, J. F. 1958. Some possibilities for measuring selection intensities in man. *Hum. Biol.* 30:1–13.
- Dudley, S. A. 1996. Differing selection on plant physiological traits in response to environmental water availability: a test of adaptive hypotheses. *Evolution* 50:92–102.
- Endler, J. A. 1986. Natural selection in the wild. Princeton Univ. Press, Princeton, NJ.
- Evans, M., N. Hastings, and B. Peacock. 2000. Statistical distributions. 3rd ed. Wiley, New York.
- Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Longman, Essex, U.K.
- Hansen, T. F., C. Pélabon, W. S. Armbruster, and M. L. Carlson. 2003. Evolvability and genetic constraint in *Dalechampia* blossoms: components of variance and measures of evolvability. *J. Evol. Biol.* 16:754–766.
- Hersch, E. I., and P. C. Phillips. 2004. Power and potential bias in field studies of natural selection. *Evolution* 58:479–485.
- Hoekstra, H. E., J. M. Hoekstra, D. Berrigan, S. N. Vignieri, A. Hoang, C. E. Hill, P. Beerli, and J. G. Kingsolver. 2001. Strength and tempo of directional selection in the wild. *Proc. Natl. Acad. Sci. USA* 98:9157–9160.
- Houle, D. 1991. Genetic covariance of fitness correlates: what genetic correlations are made of and why it matters. *Evolution* 45:630–648.
- . 1992. Comparing evolvability and variability of quantitative traits. *Genetics* 130:195–204.
- Johnson, H. W., H. F. Robinson, and R. E. Comstock. 1955. Estimates of genetic and environmental variability in soybeans. *Agron. J.* 47:314–318.
- Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gilbert, and P. Beerli. 2001. The strength of selection phenotypic selection in natural populations. *Am. Nat.* 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.
- Morgan, M. T. 1999. Comparisons of selection and inheritance: evolutionary elasticities. Unpubl. ms., available via <http://www.wsu.edu/~mmorgan/working/Elasticities>. Accessed 27 February 2004.
- Price, G. R. 1970. Selection and covariance. *Nature* 227:520–521.
- Price, T., M. Kirkpatrick, and S. J. Arnold. 1988. Directional selection and the evolution of breeding date in birds. *Science* 240:798–799.
- Rausher, M. D. 1992. The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. *Evolution* 46:616–626.
- Reeve, R., E. Smith, and B. Wallace. 1990. Components of fitness become effectively neutral in equilibrium populations. *Proc. Natl. Acad. Sci. USA* 87:2018–2020.
- Robertson, A. 1955. Selection in animals: synthesis. Cold Spring Harbor Symp. Quant. Biol. 20:225–229.
- Roff, D. A. 1992. The evolution of life histories: theory and analysis. Chapman and Hall, New York.
- Sæther, B. E., and Ø. Bakke. 2000. Avian life-history variation and contribution of demographic traits to the population growth rate. *Ecology* 81:642–645.
- Scheiner, S. M., K. Donohue, L. A. Dorn, S. J. Mazer, and L. M. Wolfe. 2002. Reducing environmental bias when measuring natural selection. *Evolution* 56:2156–2167.
- Simpson, G. G. 1944. Tempo and mode in evolution. Columbia Univ. Press, New York.
- Stearns, S. C. 1992. The evolution of life histories. Oxford Univ. Press, Oxford, U.K.
- Stinchcombe, J. R., M. T. Rutter, D. S. Burdick, P. Tiffin, M. D. Rausher, and R. Mauricio. 2002. Testing for environmentally induced bias in phenotypic estimates of natural selection: theory and practice. *Am. Nat.* 160:511–523.
- van Tienderen, P. H. 2000. Elasticities and the link between demographic and evolutionary dynamics. *Ecology* 81:666–679.
- Winn, A. A. 2004. Natural selection, evolvability and bias due to environmental covariance in the field in an annual plant. *J. Evol. Biol.* 17:1073–1083.

Corresponding Editor: C. Fenster