



## Reducing Environmental Bias When Measuring Natural Selection

Samuel M. Scheiner; Kathleen Donohue; Lisa A. Dorn; Susan J. Mazer; Lorne M. Wolfe

*Evolution*, Vol. 56, No. 11. (Nov., 2002), pp. 2156-2167.

Stable URL:

<http://links.jstor.org/sici?sici=0014-3820%28200211%2956%3A11%3C2156%3AREBWMN%3E2.0.CO%3B2-T>

*Evolution* is currently published by Society for the Study of Evolution.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ssevol.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

## REDUCING ENVIRONMENTAL BIAS WHEN MEASURING NATURAL SELECTION

SAMUEL M. SCHEINER,<sup>1,2</sup> KATHLEEN DONOHUE,<sup>3,4,5</sup> LISA A. DORN,<sup>3,6</sup> SUSAN J. MAZER,<sup>7,8</sup> AND LORNE M. WOLFE<sup>9,10</sup>

<sup>1</sup>Division of Environmental Biology, Room 635, National Science Foundation, 4201 Wilson Boulevard, Arlington, Virginia 22230

<sup>2</sup>E-mail: sscheine@nsf.gov

<sup>3</sup>Department of Ecology and Evolutionary Biology, Box G-W, Brown University, Providence, Rhode Island 02912

<sup>5</sup>E-mail: kdonohue@oeb.harvard.edu

<sup>6</sup>E-mail: Lisa.Dorn@brown.edu

<sup>7</sup>Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, California 93106

<sup>8</sup>E-mail: mazer@lifesci.ucsb.edu

<sup>9</sup>Department of Biology, Georgia Southern University, Statesboro, Georgia 30460

<sup>10</sup>E-mail: wolfe@gasou.edu

**Abstract.**—Crucial to understanding the process of natural selection is characterizing phenotypic selection. Measures of phenotypic selection can be biased by environmental variation among individuals that causes a spurious correlation between a trait and fitness. One solution is analyzing genotypic data, rather than phenotypic data. Genotypic data, however, are difficult to gather, can be gathered from few species, and typically have low statistical power. Environmental correlations may act through traits other than through fitness itself. A path analytic framework, which includes measures of such traits, may reduce environmental bias in estimates of selection coefficients. We tested the efficacy of path analysis to reduce bias by re-analyzing three experiments where both phenotypic and genotypic data were available. All three consisted of plant species (*Impatiens capensis*, *Arabidopsis thaliana*, and *Raphanus sativus*) grown in experimental plots or the greenhouse. We found that selection coefficients estimated by path analysis using phenotypic data were highly correlated with those based on genotypic data with little systematic bias in estimating the strength of selection. Although not a panacea, using path analysis can substantially reduce environmental biases in estimates of selection coefficients. Such confidence in phenotypic selection estimates is critical for progress in the study of natural selection.

**Key words.**—*Arabidopsis thaliana*, *Impatiens capensis*, natural selection, path analysis, *Raphanus sativus*, regression analysis, selection gradient.

Received March 27, 2002. Accepted July 23, 2002.

Crucial to understanding the process of natural selection is characterizing phenotypic selection, the within-generation process of differential survival and reproduction based on individual phenotypes. Combined with genetic variation, phenotypic selection results in evolutionary changes in the population mean phenotype and in the frequency of alternative genotypes. Phenotypic selection can be measured as the covariance between a trait and fitness (see Fig. 1A; Robertson 1966; Price 1970; Crow and Nagylaki 1976). However, this measure suffers from the problem of confounding correlations. Lande and Arnold (1983; see also Pearson 1903) pointed out that selection estimates can be confounded by selection on traits that are phenotypically correlated with the trait of interest. They solved this problem by the use of multiple regression, which partitions total selection on a trait ( $s$ ) into direct ( $\beta$ ) and indirect components. The result is a better understanding of which traits are the direct targets of selection and, thus, the ecological processes responsible for selection. Their paper led to a flurry of studies over the past 15 years devoted to the measurement of phenotypic selection in natural populations (Kingsolver et al. 2001).

The paper of Lande and Arnold (1983) addresses two separate issues: understanding the causal components of phenotypic selection and predicting the response to selection. The partition of total selection into direct and indirect components addresses the first issue. Combining these selection

components with information on genetic covariances among traits addresses the second issue. We highlight this distinction because it is often not appreciated that they are separate, if related, questions. More importantly, we emphasize that our paper addresses the first issue, understanding the causal components of phenotypic selection. The relationship between the selection components presented here and predicting the response to selection is still unresolved (Scheiner et al. 2000). Our efforts are similar to those of others that have focused on understanding phenotypic selection, for example, the spline technique of Schluter (1988) or the use of logistic regression as suggested by Janzen and Stern (1998).

An important caution about measuring phenotypic selection was raised by Mitchell-Olds and Shaw (1987), Price et al. (1988), and Rausher (1992), who noted that environmental variation among individuals could create biases in selection estimates (see Fig. 1B), both in the sign and magnitude of selection coefficients and by disguising stabilizing selection as directional selection or the converse. Imagine that selection is measured on leaf size in a population of plants in a field that varies in nutrient levels. Plants in pockets of high nutrients may have both larger leaves and greater seed set (one component of female fitness) because both are directly affected by nutrient level. An estimate of selection on leaf size would erroneously conclude that it is under selection. We might expect such environmental bias to be especially prevalent in sessile organisms such as plants because mobile organisms might tend to average out environmental heterogeneity. Social interactions, however, can create environ-

<sup>4</sup> Present address: Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138.

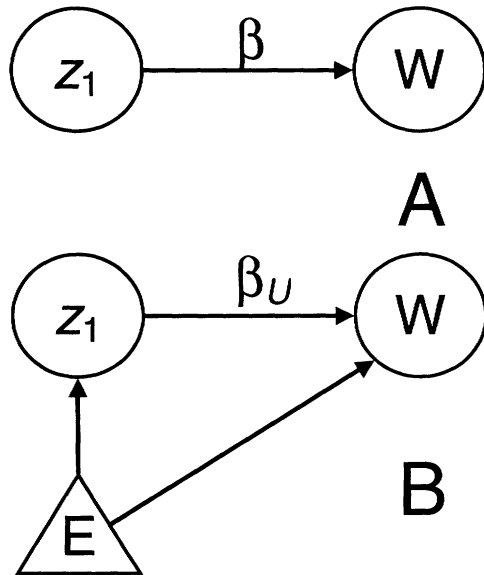


FIG. 1. (A) A simple model for the regression of fitness ( $W$ ) on the trait of interest ( $z_1$ ), where  $\beta$  is the regression coefficient. (B) A more complex model that includes an environmental factor ( $E$ ) that effects both the trait and fitness.  $E$  may represent a complex of variables that have their effects through different factors and are correlated with each other. In this model the unbiased selection coefficient ( $\beta_U$ ) could be estimated only if the environmental factor were directly measured. For simplicity, variation due to error is not included.

mental differences even among mobile individuals, so environmental bias might exist in many types of situations. See Rausher (1992) for a formal derivation of this problem and Stinchcombe et al. (2002) for a method for identifying such biases.

Rausher's (1992) solution involves estimating selection on trait breeding values. The response to selection is based on the covariance between breeding values and fitness. If one's goal is to predict the response to selection, breeding values should be used. Phenotypic values accurately estimate the covariance of breeding values and fitness only when there is no environmental covariance or if the environmental covariance happens to mimic the genetic covariance (see Fig. 2; Rausher 1992).

However, if one's goal is to accurately characterize phenotypic selection, genotypic values can also be used to eliminate bias due to environmental covariance. To implement this method, one raises genetic replicates (e.g., clones, full-sibs, or half-sibs with the last preferred because they provide the best estimates of additive genetic values) and estimates selection using genet or family means. Environmental bias is eliminated as long as the replicates are distributed randomly with respect to environmental factors. In essence, the use of genotypic means averages environmental heterogeneity across replicates.

This method has a critical limitation, however; it can only be used in cases where it is possible to manipulate genetic or family replicates. A noncomprehensive survey of the literature for the period January 1996 to December 2000 found only five studies that analyzed phenotypic selection using genotypic data (Mauricio and Rausher 1997; Mauricio et al.

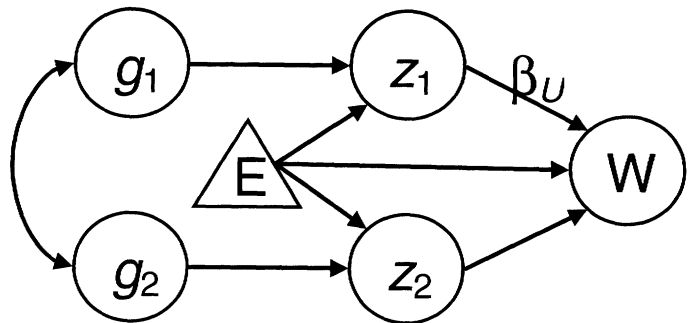


FIG. 2. A model for the analysis of selection using breeding values. Breeding values ( $g_i$ ) are linked to phenotypes by heritabilities and may be genetically correlated with each other.  $E$  may represent a complex of variables that have their effects through different factors and are correlated with each other. For simplicity, variation due to error is not included.

1997; Tiffin and Rausher 1999; Juenger and Bergelson 2000; Shonle and Bergelson 2000). In contrast, in the period January 1999 to December 2000 we identified 33 studies that used phenotypic data. In most cases an analysis of genotypic data could not have been performed. The studies that analyzed genotypic data were all done with plants where manipulation of siblings or clones is relatively simple due to, for example, the ease of controlling breeding, the rapidity of the life cycle, or the ease of creating clonal replicates.

In this paper we present an alternative method for characterizing patterns of selection that avoids the limitations of genotypic data by allowing the use of phenotypic data. This alternative method assumes that we can reduce bias in selection estimates by using one or more traits (other than those of primary interest) to estimate the overall condition of an individual. Fitness (e.g., number of grandoffspring) is rarely measured. Instead, we typically measure one or more fitness surrogates or components, such as survival or fecundity. These fitness measures depend on the composite state of the individual. That is, the nutrient variation previously described does not likely affect fecundity directly, but rather affects a suite of traits that in turn determine fecundity. We term this suite of traits "condition," in keeping with typical usage in vertebrate biology. In some species there are routine measures of condition, such as body mass in vertebrates. For other species we can identify similar traits, such as measures of vegetative size just prior to flowering.

Now we can recast the environmental covariance problem into one involving condition. The environment may affect both our trait of interest ( $z_1$ ) and an individual's condition (see Fig. 3A). If we have a measure of condition, including it in a path analysis of selection on  $z_1$  will produce an unbiased estimate of selection on that trait ( $\beta_U$ ). The model assumes that the environment affects fitness only through its effect on condition rather than directly, and that the trait of interest is not causal on condition. We can now substitute a model in which the environment is not directly measured and only condition and our other traits are included (see Fig. 3B). This model is now a path analysis, rather than a multiple regression, because we have specified that condition is causal on  $z_1$ . It does not matter that the environment affects both the trait of interest and condition, the path model accounts

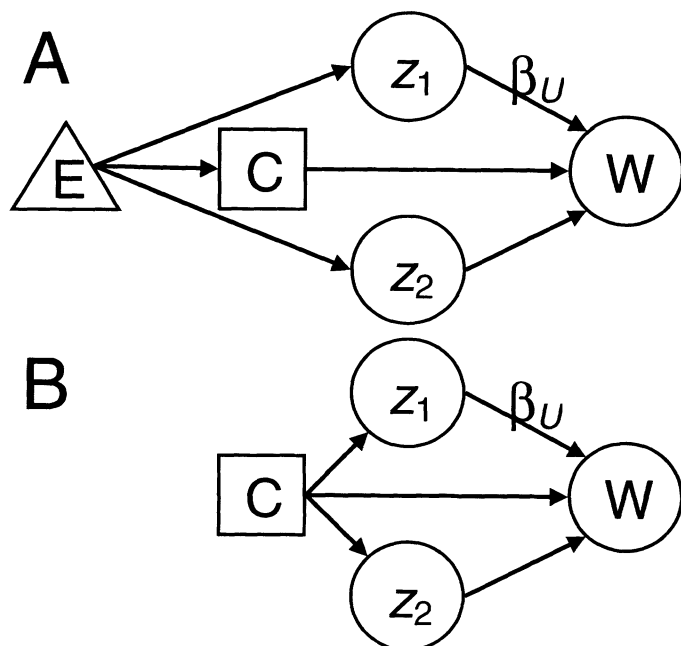


FIG. 3. (A) A path model that includes the effect of the environmental factor on both the trait of interest ( $z_1$ ) and condition (C). This model will produce an equivalent estimate of the model in (B), although the interpretation of the path between the trait and condition would differ. For simplicity, variation due to error is not included.

for any such correlation between the trait and condition. It does mean, though, that the estimate of selection on condition will be biased. This corrective effect of condition is referred to as mediation in the structural equation literature (e.g., MacKinnon and Dwyer 1993). See Shipley (2000) and references therein for a detailed description of the use of path analysis.

The cause of the correlation between condition and the trait of interest does not matter (an environmental covariance in this case). Other causes affecting both the trait and condition, such as a genetic correlation, are similarly accounted for. To put it another way, although the analysis presented here is motivated by the issue of environmental bias, our solution is general. We may, in fact, be reducing bias from causes other than the environment. It does not matter to our conclusion—path analysis acts to reduce bias—that we may fail to identify what the source(s) of that bias is (are). The result is a better understanding of the causal components of phenotypic selection. We note, however, that our method will not work for some traits, for example, early life traits that might be causal on condition.

To test the efficacy of path analysis to reduce environmental bias we compared it with multiple regression analysis using both phenotypic and genotypic data. We use genotypic data here not because it can make a better prediction of the response to selection, but because it provides a bias-free measure of phenotypic selection. We present three examples, each highlighting different strengths of our method. None of these datasets were collected with the use of path analysis in mind. Despite this limitation, we show that path analysis works

well in capturing the structure of natural selection in the phenotypic data.

#### MATERIALS AND METHODS

The first dataset was from an experiment designed to understand patterns of selection on traits and trait plasticities in *Impatiens capensis* (Balsaminaceae). This experiment included a number of vegetative and flowering traits measured on multiple inbred family lines. Other analyses showed that microenvironmental variation in light and soil moisture to be important in determining trait values and fitness. See Donohue et al. (2000a,b, 2001) for complete details of the experiment, results, and history of this system. For the purposes of the current paper, we examined only one of the four treatments: the woodland site at low density. This treatment was chosen a priori, as were the other treatments analyzed in this paper, based either on having the largest sample size or on previous analyses indicating strong trait selection.

Seeds for this experiment were initially collected from two populations, one in an open area and one beneath an oak-hickory canopy. Inbred lines were maintained through single-seed descent of self-pollinated plants for six generations. To start the experiment, seeds were germinated in a cold frame and planted into the treatment as seedlings. Seedlings in the low-density treatment analyzed here were planted in a  $7 \times 16$  array, 15 cm apart, giving a density of 53 plants/m<sup>2</sup>. Initial seed mass and germination date were recorded. Two weeks after the seedlings were planted into the treatments, they were measured for length of the first and second internodes; height; number of nodes; length of the largest leaf; and the number of axillary flowers, branches, and quiescent buds at each node of the main stem as indicators of meristem allocation. Twice a week plants were censused for the presence of cleistogamous and chasmogamous flowers and the date of first flowering was determined. Total lifetime fitness was measured as the estimated total number of seeds produced during the lifetime of the individual by counting the total number of fruits and multiplying by the number of seeds per fruit from non-experimental plants in the surrounding population. For this experiment, leaf length was the measure of condition. The final sample size consisted of 258 individuals from 35 full-sib families.

The second dataset was from an experiment designed to examine the effects of light quality and quantity on plasticity of *Arabidopsis thaliana* (Brassicaceae). This experiment measured a limited number of vegetative and flowering traits in a highly inbred species. The experiment was done in a greenhouse, where environmental heterogeneity is expected to be low. See Dorn et al. (2000) for complete details of the experiment and results. For the purposes of the current paper, we examined only one of the four treatments: foliage shade.

The lines used in this experiment were derived from seeds collected in the spring of 1995 from four natural populations. From each population, nine inbred full-sib families were randomly selected. Seeds were planted in plastic flats in  $6 \times 6$  grids, 2 cm apart. Within each flat, seeds were randomly assigned a position, with one seed per line per flat. The light treatment consisted of a reduced red:far-red ratio (0.5 compared to 1.1 for natural sunlight) and photosynthetically ac-

tive radiation of 25–30% of ambient levels. The following traits were measured: bolting date, the number of days between bolting and flowering, the number of rosette leaves and length of the longest leaf at bolting, and at senescence the number of branches growing from the main inflorescence (inflorescence branches) and from basal leaf axils (basal branches). Fitness was measured as the total number of fruits. The number of rosette leaves was the measure of condition for this dataset. The final sample size consisted of 289 individuals from 36 full-sib families.

The third dataset was from an experiment designed to examine density-mediated maternal effects on seed size in *Raphanus sativus* (Brassicaceae), to detect genetic variation in phenotypic plasticity for fitness-related traits, and to determine the effects of plant density on the expression of additive genetic variation in fitness-related traits. This experiment used a sire-dam mating design and, thus, allowed examination of differences between half-sib and full-sib analyses of selection. It also included two fitness measures, maternal and paternal. In this experiment, environmental heterogeneity was substantial as indicated by a significant block effect. See Mazer and Wolfe (1998, unpubl. ms.) for complete details of the experiment. For the purposes of the current paper, we examined only one of the three treatments: low density.

The parental generation plants that were hand-pollinated to produce the seeds used in this experiment were derived from 95 seeds (each representing a different maternal family) collected from naturally pollinated plants in August 1990. Of these plants, 19 were used as pollen donors, which were each mated to four pollen recipients. From each pollen recipient, 12 seeds were sown in each treatment. The low-density treatment consisted of three blocks. Each block consisted of a  $10 \times 30$  grid of 300 seeds positioned randomly with respect to family membership and sown 20 cm apart ( $\sim 30$  plants/m<sup>2</sup>). Plants were measured for initial seed mass, germination date, number of days to flowering, number of leaves at flowering, number of pollen grains in a single flower, pollen grain volume, style length, total petal area, number of ovules per flower, total number of flowers, total number of fruits, mean seed number per fruit, mean individual seed mass, lifetime fecundity (number of fruits  $\times$  mean seed number per fruit), maternal fitness (lifetime fecundity  $\times$  mean seed mass), and paternal fitness (total number of flowers  $\times$  number of pollen grains/flower). Paternal half-sib family means were calculated as the mean of the four full-sib means nested within each pollen donor. The number of leaves was the measure of condition in this dataset. Final sample sizes consisted of 484 individuals, 76 full-sib families, and 19 half-sib families.

### Statistical Analyses

Evidence for environmental bias in phenotypic selection estimates was determined using the method of Stinchcombe et al. (2002). The method involves a comparison of total selection on phenotypic and genotypic values using a method first suggested by Rausher (1992), but correcting an error in the formula presented in that paper. In brief, the method involves a multiple regression of trait values and deviations of those trait values from the breeding value (for details, see Stinchcombe et al. 2002). Although Rausher (1992) and

Stinchcombe et al. (2002) emphasize this method with regard to environmental bias, actually it indicates whether bias exists, regardless of the sources. Other sources include maternal effects and genetic correlations due to linkage.

Natural selection was measured two ways—with multiple regression ( $\beta$ , also referred to as the selection gradient) and with path analysis ( $\beta^*$ )—using both phenotypic and genotypic data. We emphasize that the genotypic analyses are being used as a bias-free measure of phenotypic selection and not as predictors of the response to selection. The *Impatiens* and *Arabidopsis* data each had one measure of fitness and, thus, two estimates of each selection measure for each trait. The *Raphanus* data had two measures of fitness and two genotypic measures (full-sib and half-sib) and, thus, six estimates of each selection measure for each trait. The selection gradient ( $\beta$ ) was calculated as the partial regression coefficient from a multiple regression of fitness on all traits. For the *Raphanus* data, separate multiple regressions were done for the maternal and paternal fitness measures. All multiple regressions were done with SYSTAT 6.1 for Windows (SPSS, Inc., Chicago, IL). Statistical significance for all regression coefficients was based on ordinary least-squares parametric analyses. The second measure of direct selection ( $\beta^*$ ) was calculated using path analysis where direct selection is the total effect of a trait on fitness. See Scheiner et al. (2000) for details of this calculation and a discussion of path analysis terminology. For the *Raphanus* data, both fitness measures could be included in a single analysis. To determine the comparative importance of environmental correlations acting through plant condition, path analyses with and without the condition variables were compared. All path analyses were done with AMOS (Arbuckle and Wothke 1999). Statistical significance was determined using a maximum-likelihood bootstrap.

In the current instance we are not interested in testing the path model. Instead we take the path model as a given and use it to estimate selection. The path models were built by a combination of logic and a priori information about each system. For example, most of the traits were measured at different times or represent different points in a developmental sequence. Logic tells us that later traits can never be causal on earlier traits, whereas earlier traits may or may not be causal on later traits. We assumed, however, that simultaneous traits (e.g., floral traits in *Raphanus*, see below) were not causal on each other, with correlations due to the joint effects of earlier traits.

In such a use of a path model as ours, it is incorrect to adjust the model based on statistical tests, for example, removal of a path because it is not statistically significant. This is similar to what is typically done in multiple regression analyses of selection, where traits with nonsignificant selection coefficients are retained in the analysis. We emphasize that the structure of the model is important; an incorrect structure can increase bias. However, it is outside the scope of this paper to test the model structures used here. Because an incorrect structure would predispose our analyses against reducing bias, our conclusions concerning the power of path analysis to reduce bias are conservative and, thus, robust.

Agreement between the phenotypic and genotypic analyses was assessed by calculating a correlation between sets of

TABLE 1. Analyses of natural selection on *Impatiens capensis*. Values shown in bold are statistically significant ( $P < 0.05$ ). See Figure 4 for the path diagram.

Trait	Bias test $P <$	Multiple regression ( $\beta$ )		Path analysis ( $\beta^*$ )	
		Phenotypic	Full-sib	Phenotypic	Full-sib
Seed mass	0.0001	2.71	22.31	1.43	4.58
Germination date	0.11	0.03	-0.11	<b>-0.05</b>	-0.03
Leaf length	0.14	0.18	0.14	<b>0.37</b>	0.21
Internode 1 length	0.26	0.10	0.19	<b>0.10</b>	0.13
Internode 2 length	0.75	0.24	-0.18	<b>0.07</b>	0.05
Number of nodes	0.86	-0.03	-0.47	0.06	-0.03
Number of branches	0.0001	<b>0.22</b>	<b>0.58</b>	<b>0.20</b>	<b>0.62</b>
Height	0.11	-0.07	-0.08	<b>0.08</b>	0.09
Number of flowers	0.01	<b>0.18</b>	0.11	<b>0.47</b>	<b>0.58</b>
Number of buds	0.28	<b>0.07</b>	-0.001	0.03	-0.05
Flowering interval	0.01	<b>-0.03</b>	-0.02	<b>-0.02</b>	<b>-0.03</b>
<i>N</i>		258	35	258	35

selection estimates (e.g.,  $\beta$  calculated for the phenotypic and genotypic data). If the path analysis method provides a less biased estimate ( $\beta^*$ ) than the multiple regression estimate ( $\beta$ ), then the correlation between the phenotypic and genotypic data for the  $\beta^*$  estimates will be higher than the correlation for the  $\beta$  estimates. It would be incorrect to compare the phenotypic  $\beta^*$  estimates with the genotypic  $\beta$  estimates because they are based on different causal models. Thus, we are assuming that the genotypic path model provides a bias-free measure of phenotypic selection. See Scheiner et al. (2000) for a discussion of the issue of different causal models of phenotypic selection.

Phenotypic estimates of selection could be biased in two ways that can be assessed by examining the slope and intercept of a linear regression of the genotypic  $\beta^*$  estimates on the phenotypic  $\beta^*$  estimates. This regression asks how well the phenotypic estimates predict the genotypic estimates. If there is no bias, the regression will have a slope of one and an intercept of zero. An intercept different from zero would indicate a systematic tendency toward larger or smaller selection coefficients at all magnitudes. A slope different from one (with a zero intercept) would indicate a systematic tendency toward overestimating or underestimating the absolute magnitude of the selection coefficient. An alternative measure of this second bias is the sum of the squared deviations of the phenotypic and genotypic coefficients

$$\sum_{i=1}^n \frac{(\beta_i^P - \beta_i^G)^2}{|\beta_i^G|}, \quad (1)$$

where  $n$  is the number of traits and  $\beta^G$  and  $\beta^P$  are the genetic-

based and phenotypic-based selection coefficients, respectively. The two methods were compared using the difference in this index.

For all of these tests of bias, statistical significance cannot be calculated because the estimates are not independent of each other. Clearly the phenotypic and genotypic estimates for a single trait are not independent. Also, the estimates of selection of different traits within a path analysis are not independent because the paths for some are contained within others. Thus, we used our measures—correlation, slope, intercept, and squared deviation—for comparative purposes only.

## RESULTS

For all three datasets, selection was detected for most traits (Tables 1–3). Because our purpose here is to examine the efficacy of path analysis to reduce environmental bias, we forego a detailed analysis of patterns of selection. See the original publications for those analyses. All three species showed evidence for environmental bias in the phenotypic selection estimates: four of 11 traits in *Impatiens*, five of six traits in *Arabidopsis*, three of 14 traits for maternal fitness in *Raphanus*, and two of 10 traits for paternal fitness in *Raphanus*. For the multiple regression analyses, correlations among the phenotypic, full-sib, and half-sib selection coefficients ( $\beta$ ) ranged from 0.13 to 0.99, with the lowest correlations for the *Raphanus* phenotypic-half sib comparisons and the highest correlation for the *Impatiens* comparison (Table 4).

TABLE 2. Analyses of natural selection on *Arabidopsis thaliana*. Values shown in bold are statistically significant ( $P < 0.05$ ). See Figure 5 for the path diagram.

Trait	Bias test $P <$	Multiple regression ( $\beta$ )		Path analysis ( $\beta^*$ )	
		Phenotypic	Full-sib	Phenotypic	Full-sib
Leaf length	0.02	0.02	0.02	<b>0.01</b>	0.01
Number of rosette leaves	0.02	-0.02	-0.04	<b>-0.11</b>	<b>-0.12</b>
Bolting date	0.33	<b>-0.05</b>	-0.08	0.000002	-0.001
Flowering interval	0.0003	-0.003	-0.01	0.00001	-0.004
Number of inflorescence branches	0.0001	0.04	0.14	0.0003	0.11
Number of basal branches	0.004	0.04	-0.02	<b>0.10</b>	0.09
<i>N</i>		289	36	289	36

TABLE 3. Analyses of natural selection on *Raphanus sativus*. Values shown in bold are statistically significant ( $P < 0.05$ ). See Figure 6 for the path diagram.

A. Maternal fitness analyses							
Trait	Bias test $P <$	Multiple regression ( $\beta$ )			Path analysis ( $\beta^*$ )		
		Phenotypic	Full-sib	Half-sib	Phenotypic	Full-sib	Half-sib
Seed mass	0.05	−0.003	−0.005	0.006	−0.0007	−0.0001	− <b>0.006</b>
Germination date	0.80	−0.01	−0.02	−0.10	− <b>0.04</b>	−0.008	− <b>0.12</b>
Days to flowering	0.06	0.003	0.004	0.05	<b>0.03</b>	<b>0.05</b>	<b>0.08</b>
Number of leaves	0.25	−0.01	0.001	0.02	<b>0.15</b>	<b>0.13</b>	<b>0.13</b>
Style length	0.91	−0.003	−0.001	−0.16	−0.0007	−0.02	0.05
Petal area	0.22	0.00003	−0.001	0.01	−0.0009	−0.007	−0.006
Pollen number/flower	0.03	0.0000001	−0.0000002	−0.00004	<b>0.000002</b>	<b>0.000003</b>	0.0000009
Pollen volume	0.29	<b>0.00005</b>	−0.00002	−0.0002	−0.00003	0.0001	−0.0002
Ovule number/flower	0.14	−0.01	0.003	−0.05	<b>0.07</b>	0.08	<b>0.09</b>
Number of flowers	0.06	<b>0.002</b>	0.002	0.0009	<b>0.01</b>	<b>0.009</b>	<b>0.007</b>
Seeds/fruit	0.91	−0.005	0.05	0.23	<b>0.18</b>	<b>0.18</b>	<b>0.14</b>
Offspring seed mass	0.97	<b>0.09</b>	<b>0.10</b>	0.05	<b>0.10</b>	<b>0.11</b>	<b>0.12</b>
Number of fruits	0.30	−0.003	0.002	0.02	<b>0.03</b>	<b>0.03</b>	<b>0.02</b>
Fecundity	0.05	<b>0.005</b>	<b>0.004</b>	0.002	<b>0.005</b>	<b>0.005</b>	<b>0.004</b>
<i>N</i>		484	76	19	484	76	19
B. Paternal fitness analyses							
Trait	Bias test $P <$	Multiple regression ( $\beta$ )			Path analysis ( $\beta^*$ )		
		Phenotypic	Full-sib	Half-sib	Phenotypic	Full-sib	Half-sib
Seed mass	0.01	−0.0001	−0.01	−0.01	−0.0008	−0.0002	− <b>0.005</b>
Germination date	0.45	0.003	0.001	−0.02	− <b>0.05</b>	−0.01	−0.12
Days to flowering	0.08	−0.005	−0.01	−0.002	<b>0.02</b>	0.04	<b>0.06</b>
Number of leaves	0.79	−0.005	0.01	0.01	<b>0.14</b>	<b>0.12</b>	<b>0.15</b>
Style length	0.56	0.0001	−0.01	−0.05	−0.0007	−0.02	0.07
Petal area	0.01	− <b>0.001</b>	−0.002	−0.001	−0.0009	−0.01	−0.008
Pollen number/flower	0.97	<b>0.000008</b>	<b>0.00001</b>	<b>0.000007</b>	<b>0.00001</b>	<b>0.00001</b>	0.000009
Pollen volume	0.51	0.00001	0.00003	0.0001	−0.00004	0.0002	−0.0003
Ovule number/flower	0.08	−0.01	−0.01	−0.02	0.03	0.05	0.09
Number of flowers	0.48	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>
<i>N</i>		484	76	19	484	76	19

The phenotypic and genotypic path analyses of the *Impatiens* data were very similar, with generally small changes in the magnitudes of the path coefficients (Fig. 4). Only one path coefficient, the path from the number of buds to fitness, changed sign, and the magnitudes of those coefficients were small. For *Arabidopsis*, again the signs of most coefficients were the same in the two analyses, with a tendency toward somewhat larger magnitudes in the genotypic analysis (Fig. 5). The most notable exception was the path from number of rosette leaves to number of basal branches, which changed from negative and of small magnitude to positive and of large

magnitude. For *Raphanus* there was also a tendency toward path coefficients of larger magnitudes in the full-sib analysis compared to the phenotypic analysis, and greater still in the half-sib analysis (Fig. 6). In addition, six path coefficients changed sign among the analyses, all involving phenology or flower morphology traits. Changes in sign were generally associated with coefficients of small magnitude.

Correlations between the phenotypic and genotypic selection estimates from the path analyses ( $\beta^*$ ) were higher than those from the multiple regressions ( $\beta$ ), with the exception of the *Impatiens* analyses (Table 4). (Similar results [not

TABLE 4. Comparisons of selection measures of phenotypic and full-sib data (*Impatiens* and *Arabidopsis*) or phenotypic, full-sib, and half-sib data (*Raphanus*): correlations ( $r$ ) of regression analyses ( $\beta$ ) and path analyses ( $\beta^*$ ) and slopes and intercept of regressions of genotypic  $\beta^*$  on phenotypic  $\beta^*$  values. The squared deviation shows the difference of the sum of the square deviations of the multiple regression and path analysis regression coefficients; a positive value indicates that the phenotypic multiple regression estimates deviate more from their genotypic estimates than do the path analysis estimates.

<i>Raphanus</i>								
Comparison	<i>Impatiens</i>	<i>Arabidopsis</i>	Maternal			Paternal		
			Pheno-full	Pheno-half	Full-half	Pheno-full	Pheno-half	Full-half
$r_\beta$	0.993	0.703	0.870	0.170	0.499	0.520	0.125	0.729
$r_{\beta^*}$	0.963	0.812	0.975	0.896	0.844	0.936	0.855	0.757
Slope	3.087	0.670	0.929	0.974	0.961	0.794	1.220	1.273
Intercept	−0.211	−0.010	0.005	−0.00001	−0.002	0.005	0.006	0.002
Squared deviation	24.922	0.076	0.024	0.474	0.317	−0.081	−0.072	−0.162

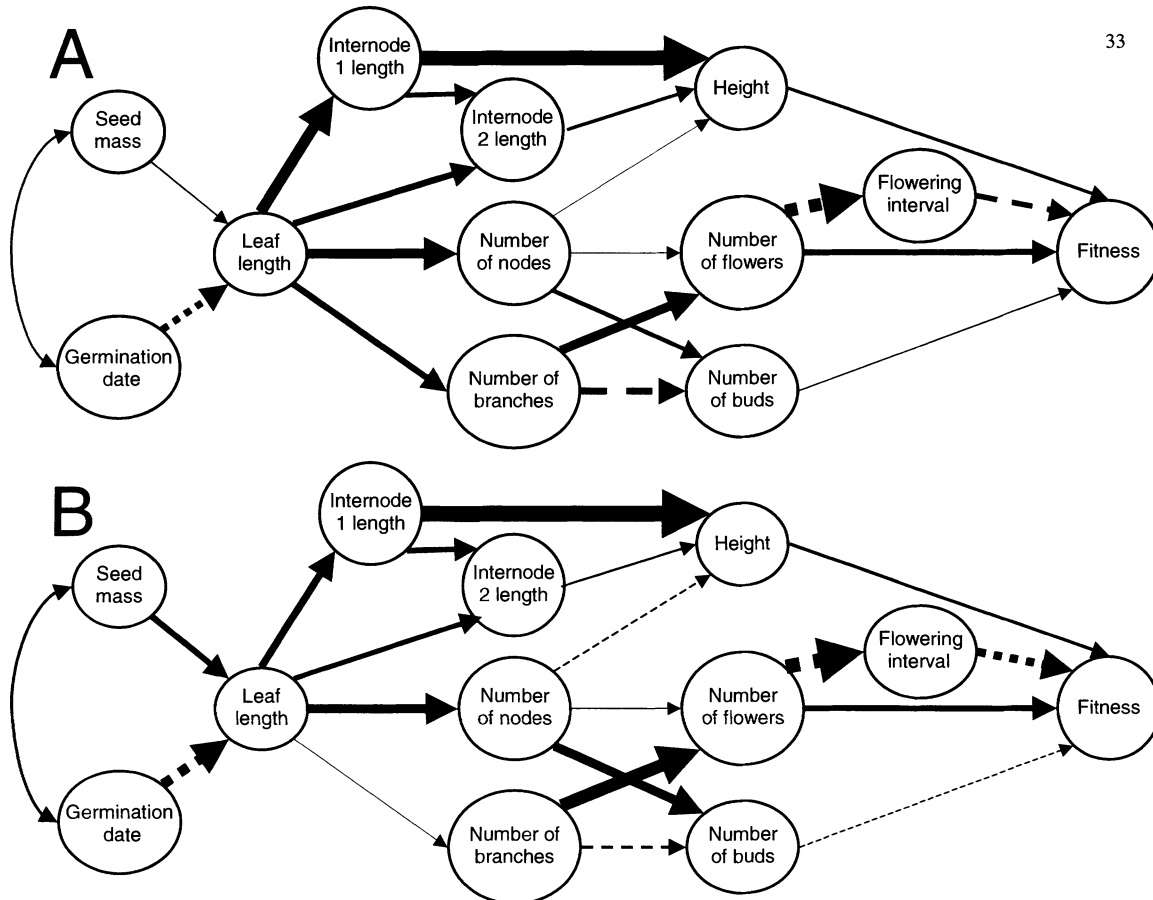


FIG. 4. Phenotypic (A) and full-sib (B) path analyses of *Impatiens*. Arrow thickness is proportional to the magnitude of the standardized path coefficient. Solid arrows denote positive coefficients and dashed arrows denote negative coefficients. For simplicity, variation due to error is not included. See Table 1 for the direct selection coefficients.

shown] were obtained using standardized selection coefficients [ $\beta'$  of Lande and Arnold 1983]. The conclusions about the *Impatiens* analyses was strongly influenced by a single trait, seed mass. If that trait is removed from the analyses, the correlations decrease, but substantially more so for the multiple regression analysis ( $r = 0.53$ ), than for the path analysis ( $r = 0.78$ ). The squared deviations indicated a better the fit for the path analyses for *Impatiens*, *Arabidopsis*, and *Raphanus* maternal fitness, but worse for *Raphanus* paternal fitness.

For all comparisons, the agreement between the phenotypic and genotypic estimates from the path analyses was quite good (0.76–0.96). The intercepts of the regressions of genotypic  $\beta^*$  estimates on phenotypic  $\beta^*$  estimates all clustered closely around zero, with the exception of *Impatiens*. The slopes for the *Impatiens* and *Raphanus* half-sib paternal regressions were greater than one (Tables 1, 3B), indicating that the phenotypic path analyses tended to underestimate the absolute strength of selection, particularly at large values of  $\beta$ . In comparison, for *Arabidopsis* and the *Raphanus* full-sib paternal comparisons, the slopes were less than one, indicating that the phenotypic path analyses tended to overestimate the strength of selection (Tables 2, 3B). The *Raphanus* maternal fitness-half sib comparisons had slopes very close to one.

Because these correlations could be dominated by a few traits closely related to fitness, we also examined the correlations between the phenotypic  $\beta^*$  and genotypic  $\beta^*$  estimates for the set of traits that were intermediate between the early vegetative traits and final fitness components. For *Impatiens* the traits were internode lengths, number of nodes, and number of branches and the correlation ( $r = 0.992$ ) was even greater than that of the entire set of traits. In contrast, for *Arabidopsis*, after elimination of leaf length and number of leaves, the correlation ( $r = 0.443$ ) dropped substantially. For *Raphanus* the traits were number of pollen grains, pollen grain volume, style length, total petal area, and number of ovules. The correlations were somewhat smaller than that for the entire set of traits: 0.750 to 0.982 for maternal fitness and 0.460 to 0.949 for paternal fitness, with the lower values being associated with half-sib comparisons.

Another test of possible environmental effects is a re-analysis with a change in the path model, either adding a path from a measure of condition directly to fitness or deleting all condition measures. One then compares selection estimates for traits that are downstream (i.e., to the right in the path diagram) of the measure of condition. The comparison consisted of a correlation of the selection coefficients estimated from the two models (with and without condition traits). For *Impatiens* the best measure of condition in this



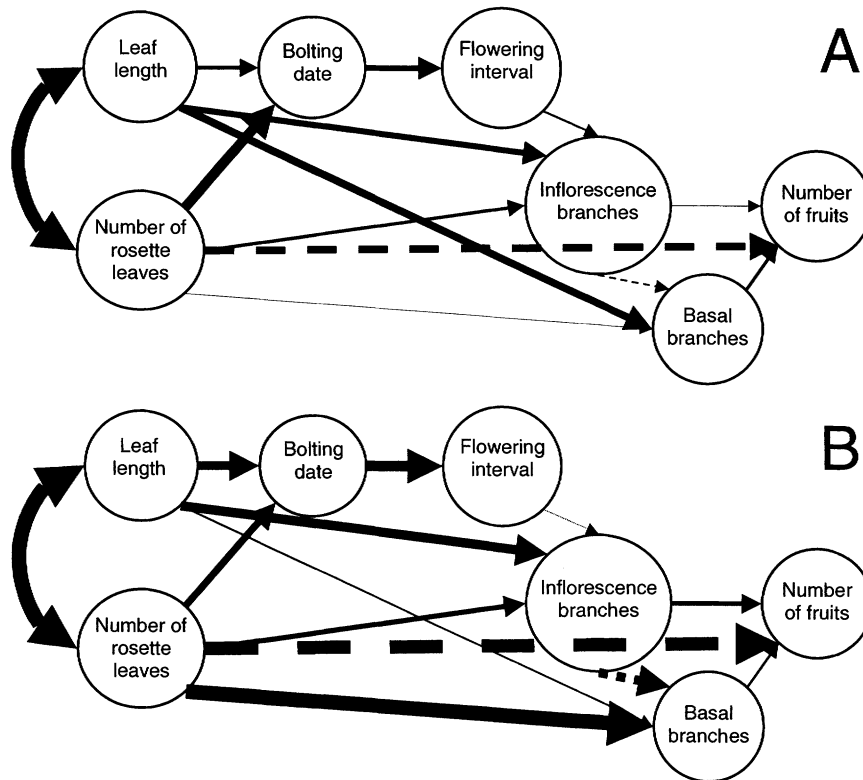


FIG. 5. Phenotypic (A) and full-sib (B) path analyses of *Arabidopsis*. Arrow thickness is proportional to the magnitude of the standardized path coefficient. Solid arrows denote positive coefficients and dashed arrows denote negative coefficients. For simplicity, variation due to error is not included. See Table 2 for the direct selection coefficients.

experiment is leaf length. Adding a path directly from leaf length to fitness had almost no change in the estimates of selection on the phenotypic ( $r = 0.846$ ) or genotypic ( $r = 0.999$ ) analyses. Similarly deleting seed mass, germination data, and leaf length traits resulted in no change in the selection estimates ( $r = 1.000$  for both). We also examined the correlation between the phenotypic and genotypic estimates with the new model, which was very similar to the same correlation with the original model ( $r = 0.988$  for the analysis with the traits removed compared with  $r = 0.992$  for the same traits in the original analysis).

For *Arabidopsis* the best measure of condition in this experiment is the number of rosette leaves. In this case, eliminating the path from number of leaves to number of fruits decreased the concordance between the phenotypic and genotypic estimates ( $r = 0.678$ ). Comparing the selection coefficients estimated from the new and old models found large differences for both the phenotypic analysis ( $r = 0.231$ ) and, especially, the genotypic analysis ( $r = -0.656$ ).

For *Raphanus* the best measure of condition in this experiment is the number of leaves at flowering. Adding paths directly from leaf number to maternal and paternal fitness made no change in the selection estimates ( $r = 1.000$  for all comparisons of new with old models). We also analyzed a model deleting the preflowering traits (seed mass, germination date, days to flowering, and number of leaves). For comparisons with this model we focused on the floral morphology traits (number of pollen grains, pollen grain volume, style length, total petal area, and number of ovules). For maternal

fitness, the model change resulted in a substantial reduction in the correlation for the phenotypic–half sib comparison ( $r = 0.855$  to  $r = -0.097$  with and without, respectively) and the full sib–half sib comparison ( $r = 0.749$  to  $r = -0.688$ ), and a smaller reduction for the phenotypic–full sib comparison ( $r = 0.982$  to  $r = 0.766$ ). These differences were primarily due to a large discrepancy between the two models for the estimates of selection for the half-sib data ( $r = 0.168$ ). A similar pattern was found for paternal fitness.

## DISCUSSION

Selection coefficients estimated using path analysis ( $\beta^*$ ) with phenotypic data were generally close to those estimated from genotypic data. Correlations between the phenotypic and genotypic estimates ranged from 0.76 upward. Bias in our estimates tended to be expressed as a tendency to under- or overestimate the strength of selection for the most extreme values, especially in *Impatiens*. This tendency went in both directions, with no systematic bias in one direction or the other. The intercepts of the regressions were very close to zero, again except for *Impatiens*. This conclusion is similar to that of Stinchcombe et al. (2002), who found that environmental bias tended to effect the magnitude of selection estimates, rather than the sign. However, bias indicated as discrepancies in selection estimates were not related to the chance of a significant result in the environmental bias test of Stinchcombe et al. (2002).

The correlations between the phenotypic and genotypic  $\beta^*$ s

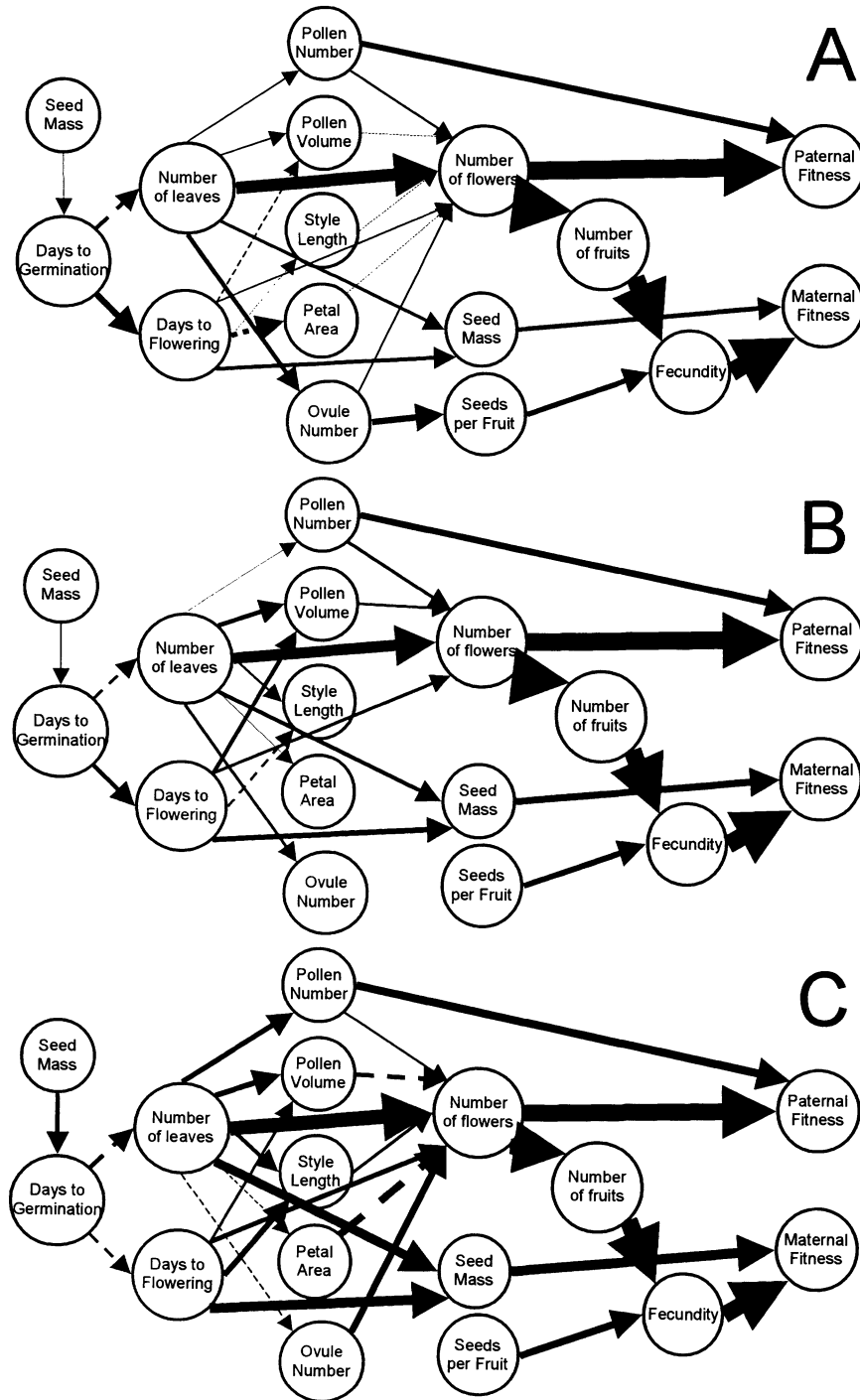


FIG. 6. Phenotypic (A), full-sib (B), and half-sib (C) path analyses of *Raphanus*. Arrow thickness is proportional to the magnitude of the standardized path coefficient. Solid arrows denote positive coefficients and dashed arrows denote negative coefficients. For simplicity, variation due to error is not included. See Table 3 for the direct selection coefficients.

tended to be stronger than those for the  $\beta_s$ , suggesting that bias, environmental or otherwise, existed in these datasets, but was reduced by path analysis. Additional evidence for environmental bias in two of the datasets comes from the comparison of path analyses with and without the condition variables. In the case of *Impatiens*, analyses that changed the path model by eliminating the condition trait had little effect

on the selection coefficients. This consistency suggests that environmental effects on fitness were not strong, but could also indicate that leaf size is not an accurate indicator of condition because it was measured on young plants. In contrast, in the case of *Raphanus* the accuracy of the phenotypic  $\beta^*s$  was likely due to the ability of the path model to account for environmental effects. As expected if environmental ef-

fects were present, the two genotypic estimates using the full-sib and half-sib data were more highly correlated with each other than either were with the phenotypic data. More importantly, removing the vegetative traits through which environmental effects likely manifest themselves significantly degraded the agreement between the phenotypic and genotypic estimates. Most interesting was the case of *Arabidopsis* where we expected environmental effects to be minimal. Instead, removal of the path between the measure of condition and fitness substantially degraded the accuracy of the phenotypic selection analysis. So, either there was substantial uncontrolled environmental heterogeneity in the experiment or leaf number has an important direct effect on fruit set independently of its effect through flowering traits. Of our three datasets, this experiment showed the greatest evidence of environmental bias by the test of Stinchcombe et al. (2002).

Our results are consistent with a recent comprehensive literature survey of natural selection (Kingsolver et al. 2001). They found that estimates of direct selection ( $\beta$ ) from phenotypic regression analyses were highly correlated with total selection estimates ( $s$ ), indicating that indirect selection was small. Because environmental effects would most likely manifest as changes in indirect selection (Figs. 1, 2), indirect selection will likely be weaker when environmental effects are minimal.

#### Use of Latent Variables

In this paper we used path models that only included measured variables. Condition can be treated as a latent variable, however, one that is not directly measured but is a function of measured variables. Such an approach has several advantages when we wish to focus only on selection of a subset of the traits. In such a case the form of the model relating the other traits to condition does not matter. For example, it does not matter if one of these other traits effects the target trait directly and not through condition. Such an error in the model will mean that the path coefficients between condition and the target trait and between condition and fitness are incorrect. But we would pay no attention to these values. See Crespi and Bookstein (1989) and Pugesek and Tomer (1996) for other examples of the use of latent variables in selection analysis; in those examples it is selection on the latent variable that is of interest, in contrast to the current case. For the data analyzed here, we only had one or two traits that represented measures of condition, obviating the need for a latent variable.

#### Phenotypic versus Genotypic Analyses

A. Winn (unpubl. ms.) looked for evidence of environmental bias in an analysis of selection on leaf traits in *Dicerandra linearifolia* (Lamiaceae). She found that selection gradients from genotypic analyses were never statistically significant and that the magnitude of the phenotypic gradients were larger than the genotypic gradients for four of seven traits; she concluded that environmental bias may be more common than previously thought. The datasets presented here likewise showed evidence of environmental bias, but the path analyses that included a measurement of condition were able to decrease that bias. Winn measured only leaf morphological

traits and had no measure of condition, so the environmental bias acting through condition was not controlled for in her phenotypic selection analyses. Yet, for her estimates of linear selection the correlation between her phenotypic and genotypic estimates was 0.997, whereas the slope and intercept were 4.791 and  $-0.487$ , respectively, similar to what we found for *Impatiens*. Stinchcombe et al. (2002) also concluded that environmental bias tends to effect the magnitude, rather than the sign, of selection coefficients.

If the causal structure for an organism is as depicted in Figure 3, then even a multiple regression will reduce bias for selection coefficients of the trait of interest ( $z_1$ ). This reduction would not hold for all traits, especially for more complex causal structures (e.g., in Figs. 4–6). However, it would explain why the phenotypic multiple regression analyses were often in reasonable agreement with the parallel genotypic analyses, even in the face of apparent environmental bias (e.g., *Arabidopsis*; Tables 1, 4). Thus, we might have guarded confidence that previous selection analyses (see summary in Kingsolver et al. 2001) are at least qualitatively correct, with the caveat that our analyses did not explore the accuracy of stabilizing/disruptive selection estimates.

Winn's observation of nonsignificant genotypic selection gradients was likewise reflected in the analyses presented here. In our data (Tables 1–3),  $\beta$  and  $\beta^*$  were statistically significant much less often for analyses with genotypic data even when the magnitudes of the coefficients were larger (e.g., estimates of  $\beta$  for bolting date in *Arabidopsis* and  $\beta^*$  for days to flowering and paternal fitness in *Raphanus*). This concurrence results from the inevitable fact that the power of analyses based on genotypic data will always be lower than analyses based on phenotypic data. Selection studies must contend with both environmental bias in phenotypic analyses and the difficulty of executing a genotypic design that nonetheless has low statistical power. A path analysis approach has the potential to alleviate both problems. We also note that in a direct comparison of a path analysis and a multiple regression using the same data (e.g., phenotypic data) the path analysis will have greater power because the total model estimates fewer coefficients. That difference is not always appreciated because we typically do not consider the covariances among the independent variables in a multiple regression as part of the estimated parameters.

Path analysis is not a panacea, however. There may be environmental factors that can directly affect fitness that are not captured by measures of condition. For example, nitrogen may be limiting for production of both pollen and herbivore defense compounds. How common such examples are is an empirical question. Early-life traits might have causal effects on condition, thus not allowing one to factor out environmental biases. In addition, one must have knowledge of which traits accurately indicate condition. However, although genotypic estimates are less biased, they require manipulated individuals. Because any manipulation potentially changes the environment, such experiments themselves may bias estimates of selection coefficients. Thus, path analysis provides a measurement context that may be less intrusive and more accurate.

A complementary approach to that outlined in this paper is to directly measure the environment and include environ-

mental parameters in the analysis (e.g., Donohue 1999). As in this paper, a path analytic framework can be used to incorporate environmental effects, although, as always, one can never be sure that all relevant parameters have been measured. Observed effects may be due to some unmeasured, correlated parameter. Some of these problems can be addressed by experimental manipulation of environmental parameters that, again, can be incorporated into a path analysis.

In this paper we examined whether a phenotypic analysis using path analysis is sufficient for assessing the most important qualities of natural selection—such as whether selection is strong or weak, positive or negative, stabilizing, directional or disruptive—and concentrated on the entire pattern, not just those coefficients that were statistically significant. As pointed out by A. Winn (unpubl. ms.), genotypic analyses are both valuable and difficult. If we rely on them alone and simultaneously commit ourselves to a level of statistical significance that is unnecessary and unachievable, we are potentially trapping ourselves into an inability to ever say anything about patterns of natural selection. We do not deny the importance of statistical inference in guiding scientific inquiry. Rather we stress that a scientific inquiry needs to consider the total evidence when reaching any conclusions (Scheiner 2003).

### Conclusions

We found that estimates of phenotypic selection calculated using path analysis were not compromised by environmental bias when compared to genotypic estimates. Our conclusions give us confidence that we can rely on phenotypic estimates in other instances where genotypic estimates are difficult or impossible to obtain.

In the current paper, the path model was assumed *a priori* and then used to estimate selection coefficients. Because of the strong dependence of results on the hypothesized causal structure, it is important that one has confidence in the path model. The model can be built from a combination of logic, biological knowledge, and experiments (but for another approach, see Shipley 1997), as well as formal model testing (Shipley 2000). Model building and testing is an important exercise that takes place prior to the issues that we raise in this paper (Shipley 2000).

Path analysis is a powerful technique for incorporating causal relationships among traits into selection analyses and showing how those causal relationships create the linkage between a trait and fitness (Scheiner et al. 2000 and references therein). Our use of path analysis in this paper does not fully explore its possibilities. Analyzing selection in a path-analytic framework promises to provide a better understanding of natural selection by more accurately indicating when during the life cycle selection is operating (Arnold 1983). Scheiner et al. (2000) showed how the discrepancy between estimates of total selection ( $s$ ) and the predicted covariance ( $s^*$ ) provides clues about the possible importance of unmeasured traits and how stabilizing selection can be analyzed by path analysis. Path analysis and multiple regression assume different causal structures and the use of path-analysis-generated selection coefficients in response to selection equations is still unex-

plored. All of these uses of path analysis provide a richer view of natural selection.

### ACKNOWLEDGMENTS

We thank A. Winn for inspiring this work, sharing her unpublished manuscript with us, and commenting on an earlier version of this manuscript; J. Stinchcombe and coauthors for sharing their unpublished manuscript and comments on our manuscript; R. Mitchell for advice on path analysis; and B. Shipley, J. Kingsolver, and two anonymous reviewers for insightful reviews. We had useful discussions of these issues with D. Houle, T. Mitchell-Olds, J. Schmitt, and A. Winn. The data in this paper were generated by research funded by National Science Foundation grants GER-9023389, DEB-9306637, DEB-9509239, DEB-9708114, and DEB-9976997 to J. H. Schmitt; grant DEB-9157270 to SJM; and by a National Science and Engineering Research Council of Canada postdoctoral fellowship to LMW. The views expressed in this paper do not necessarily reflect those of the National Science Foundation or the U.S. government.

### LITERATURE CITED

- Arbuckle, J. L., and W. Wothke. 1999. Amos 4.0 user's guide. Smallwaters Corporation, Chicago, IL.
- Arnold, S. J. 1983. Morphology, performance and fitness. *Am. Zool.* 23:347–361.
- Crespi, B. J., and F. L. Bookstein. 1989. A path-analytic model for the measurement of selection on morphology. *Evolution* 43: 18–28.
- Crow, J. F., and T. Nagylaki. 1976. The rate of change of a character correlated with fitness. *Am. Nat.* 110:207–213.
- Donohue, K. 1999. Seed dispersal as a maternally influenced character: mechanistic basis of maternal effects and selection on maternal characters in an annual plant. *Am. Nat.* 154:674–689.
- Donohue, K., D. Messiqua, E. H. Pyle, M. S. Heschel, and J. Schmitt. 2000a. Evidence of adaptive divergence in plasticity: density- and site-dependent selection on shade-avoidance response in *Impatiens capensis*. *Evolution* 54:1956–1968.
- Donohue, K., E. H. Pyle, D. Messiqua, M. S. Heschel, and J. Schmitt. 2000b. Density dependence and population differentiation of genetic architecture in *Impatiens capensis* in natural environments. *Evolution* 54:1969–1981.
- . 2001. Adaptive divergence in plasticity in natural populations of *Impatiens capensis* and its consequences for performance in novel habitats. *Evolution* 55:692–702.
- Dorn, L. A., E. H. Pyle, and J. Schmitt. 2000. Plasticity to light cues and resources in *Arabidopsis thaliana*: testing for adaptive value and costs. *Evolution* 54:1982–1994.
- Janzen, F. J., and H. S. Stern. 1998. Logistic regression for empirical studies of multivariate selection. *Evolution* 52:1564–1571.
- Juenger, T., and J. Bergelson. 2000. The evolution of compensation to herbivory in scarlet gilia, *Ipomopsis aggregata*: herbivore-imposed natural selection and the quantitative genetics of tolerance. *Evolution* 54:764–777.
- 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 phenotypic selection in natural populations. *Am. Nat.* 157:245–261.
- Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–1226.
- MacKinnon, D. P., and J. H. Dwyer. 1993. Estimating mediated effects in prevention studies. *Eval. Rev.* 17:144–158.
- 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.
- Mazer, S. J., and L. M. Wolfe. 1998. Density-mediated maternal effects on seed size in wild radish. Pp. 323–343 in T. A. Mousseau and C. W. Fox, eds. *Maternal effects as adaptations*. Oxford Univ. Press, New York.
- Mitchell-Olds, T., and R. G. Shaw. 1987. Regression analysis of natural selection: statistical inference and biological interpretation. *Evolution* 41:1149–1161.
- Pearson, K. 1903. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Philos. Trans. R. Soc. Lond. A* 200: 1–66.
- 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.
- Pugesek, B. H., and A. Tomer. 1996. The Bumpus house sparrow data: a reanalysis using structural equation models. *Evol. Ecol.* 10:387–404.
- Rausher, M. D. 1992. The measurement of selection on quantitative traits: biases due to 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.
- Robertson, A. 1966. A mathematical model of the culling process in dairy cattle. *Anim. Prod.* 8:95–108.
- Scheiner, S. M. 2003. Experiments, observations, and other kinds of evidence. In M. L. Taper and S. R. Lele, eds. *The nature of scientific evidence*. Univ. of Chicago Press, Chicago, IL, *In press*.
- Scheiner, S. M., R. J. Mitchell, and H. Callahan. 2000. Using path analysis to measure natural selection. *J. Evol. Biol.* 13:423–433.
- Schluter, D. 1988. Estimating the form of natural selection on a quantitative trait. *Evolution* 42:849–861.
- Shipley, B. 1997. Exploratory path analysis with applications in ecology and evolution. *Am. Nat.* 149:1113–1138.
- . 2000. *Cause and correlation in biology*. Cambridge Univ. Press, Cambridge, U.K.
- Shonle, I., and J. Bergelson. 2000. Evolutionary ecology of the tropane alkaloids of *Datura stramonium* L. (Solanaceae). *Evolution* 54:778–788.
- 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.
- Tiffin, P., and M. D. Rausher. 1999. Genetic constraints and selection acting on the tolerance to herbivory in the common morning glory *Ipomoea purpurea*. *Am. Nat.* 154:700–716.

Corresponding Editor: J. Willis