

Estimates of the Average Strength of Natural Selection Are Not Inflated by Sampling Error or Publication Bias

Frances N. Knapczyk* and Jeffrey K. Conner†

Kellogg Biological Station and Department of Plant Biology,
Michigan State University, Hickory Corners, Michigan 49060

Submitted January 11, 2007; Accepted May 1, 2007;

Electronically published August 17, 2007

ABSTRACT: Kingsolver et al.'s review of phenotypic selection gradients from natural populations provided a glimpse of the form and strength of selection in nature and how selection on different organisms and traits varies. Because this review's underlying database could be a key tool for answering fundamental questions concerning natural selection, it has spawned discussion of potential biases inherent in the review process. Here, we explicitly test for two commonly discussed sources of bias: sampling error and publication bias. We model the relationship between variance among selection gradients and sample size that sampling error produces by subsampling large empirical data sets containing measurements of traits and fitness. We find that this relationship was not mimicked by the review data set and therefore conclude that sampling error does not bias estimations of the average strength of selection. Using graphical tests, we find evidence for bias against publishing weak estimates of selection only among very small studies ($N < 38$). However, this evidence is counteracted by excess weak estimates in larger studies. Thus, estimates of average strength of selection from the review are less biased than is often assumed. Devising and conducting straightforward tests for different biases allows concern to be focused on the most troublesome factors.

Keywords: natural selection, phenotypic selection, selection gradient, publication bias, sampling error.

Since Lande and Arnold (1983) introduced the methodology for quantifying selection through the use of multiple regression, evolutionary biologists have measured selection gradients to understand the strength and direction of selection on individual quantitative traits. Selection gradients

rose quickly in popularity because they can be used in quantitative genetic equations to predict how individual traits will evolve, and in standardized form, they can be compared across environments, traits, and species. Considered collectively, as in reviews by Endler (1986), Hoekstra et al. (2001), Kingsolver et al. (2001), and Hereford et al. (2004), the multitude of published selection gradients begins to give a picture of the strength of selection in nature. This information can be used to parameterize models of evolutionary change in quantitative traits (e.g., Albert and Otto 2005). Further, summaries of selection in nature may shed new light on old debates such as the relative importance of genetic drift versus selection in causing evolutionary change (e.g., Roff and Mousseau 2005), or whether traits are well or poorly adapted with respect to the environments in which they are currently found.

The impacts of results from meta-analyses of published studies are often broad; for example, a review by Kingsolver et al. (2001) has been cited 269 times. Therefore, it is imperative to critically examine any biases inherent in review data sets. Kingsolver et al. (2001) suspected that sampling error, publication bias, and low statistical power of selection analyses were dominant sources of bias that might inflate estimates of the median magnitude of selection in nature. Hersch and Phillips (2004) confirmed that most estimates of selection in Kingsolver et al.'s (2001) review for which they could perform power analyses had low power to detect weak to moderate selection. They showed by simulation models that low power, in concert with sampling error and bias against publishing nonsignificant selection gradients, could cause as much as a 10-fold inflation of the estimate of median magnitude of selection.

The narrowing of the range of published selection gradients as sample size increased (fig. 1, redrawn from Kingsolver et al.'s [2001] fig. 1) suggests that sampling error variance could have inflated the range of selection gradients at small sample sizes. Sampling error contributes to the total variance among measurements of selection within and among studies. By definition, this sampling

* E-mail: knapczyk@msu.edu.

† E-mail: connerj@msu.edu.

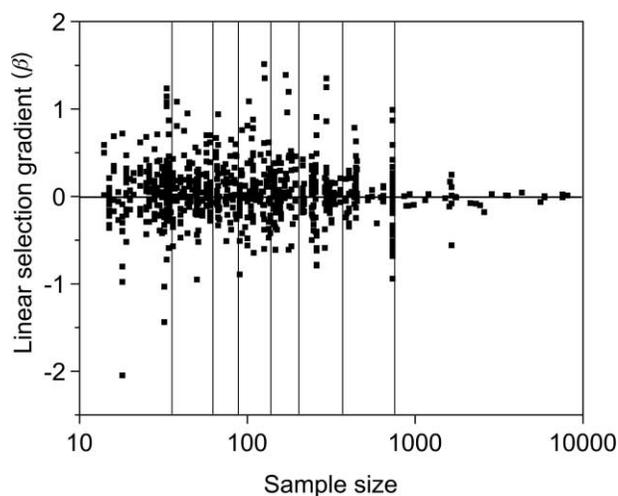


Figure 1: Distribution of 993 linear selection gradients (β) versus sample size from 63 studies in a review by Kingsolver et al. (2001). Bin boundaries used in our analyses are plotted on the graph; the number of gradients in each bin was (in ascending order of bin number): 256, 158, 61, 165, 81, 125, 112, 35. Redrawn from figure 1 of Kingsolver et al. (2001).

error variance among selection gradients declines with increasing sample size:

$$\text{Variance} = \frac{\sum (X_i - \bar{X})^2}{N - 1}. \quad (1)$$

In contrast, as long as a random sample of traits, fitness measures, and study organisms is represented across sample sizes, the true biological variance among selection gradients will be independent of sample size. As a result, the total variance (true biological variance plus sampling variance) among selection gradients in a review should decline as study sample size increases. This pattern is superficially observed in the Kingsolver et al. (2001) review. If sampling error contributed to the pattern in the review, then positive and negative estimates of selection that are stronger than they are in nature as a result of sampling error would cause the average strength of selection to be biased upward (fig. 2; Conner 2001; Kingsolver et al. 2001; Hersch and Phillips 2004).

The inverse relationship between total variance and sample size could be masked in a meta-analysis if studies with different amounts of true variance among estimates had different sample sizes. For example, one might hypothesize that selection on phenological traits is more variable in nature than selection on reproductive traits. If sample sizes were usually smaller for measuring reproductive traits than phenological traits, then a plot of effect size versus sample size including both types of selection

estimates might not show total variance decreasing with increasing sample size. Thus, the expected inverse relationship between total variance and sample size could be exaggerated or obscured by correlations between sample size and true variance across study type.

In the Kingsolver et al. (2001) review, the range of selection gradients narrows most strikingly at sample sizes larger than 1,000, where most of the estimates of selection are between 0.1 and -0.1 . In contrast, at sample sizes less than 1,000, selection gradients range from -2.0 to 1.5. Thus, the few studies that have the most power to measure selection accurately suggest that selection is weak. If sampling error is indeed driving the pattern in the review, many estimates of strong selection in studies with small sample sizes could be artifacts of sampling error. However, it is difficult to make this conclusion with confidence because only 28 of the 993 selection gradients in the data set were estimated with sample sizes greater than 1,000

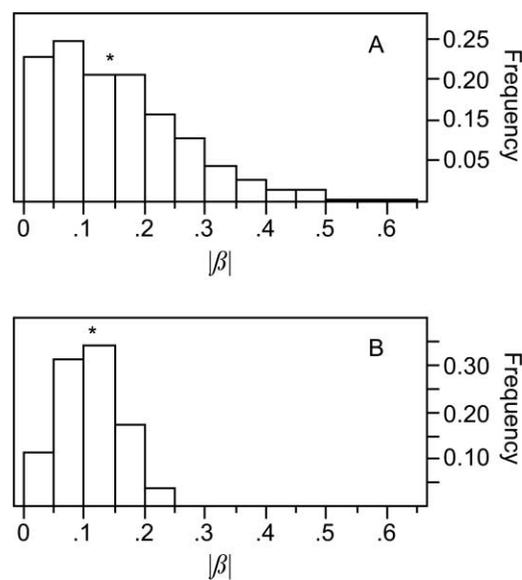


Figure 2: Sampling error variance can inflate the median magnitude of selection ($|\beta|$) from a review data set. Sampling distributions were created by randomly subsampling the J. K. Conner (unpublished) data set to create smaller data sets with sample sizes of 50 and 500 (see “Methods” for details). We performed this bootstrap procedure 1,000 times at each subsample size. We then estimated the selection gradient for ovule number from each of the bootstrapped data sets. The magnitude of selection on ovule number estimated from the full data set ($|\beta| = 0.11$) is similar to the median magnitude of selection from Kingsolver et al. (2001; $|\beta| = 0.16$). A, Sampling distribution of 1,000 magnitudes of selection on ovule number when the sample size for each estimate is 50. The median of the distribution is 0.14 (asterisk). B, Sampling distribution of 1,000 magnitudes of selection when the sample size for each estimate is 500. The median of the distribution is 0.11. The distribution in A had more sampling error variance than the distribution in B; as a result, the median of A is larger.

(representing only three of the 63 different studies). Further, the range of estimates decreases abruptly at $N = 734$, rather than declining smoothly across sample size as we would predict. This effect is primarily due to one study with $N = 734$ that reported 41 selection gradients between -0.94 and 0.99 , but there are too few studies with large sample sizes to determine whether these studies are representative.

Here, we use a novel approach to study the influence of sampling error variance on patterns of selection strength across sample size in the Kingsolver et al. (2001) data set. We compare the relationship between sample size and total variance of selection gradients in the review to that for 15 traits derived from subsampling four large empirical data sets at multiple sample sizes. If sampling error variance created the pattern in the review, we expect the relationship between total variance in selection gradients and sample size to be similar for the review and our subsampled data sets.

Publication bias, or selective reporting, could also obscure our understanding of the range and distribution of selection gradients in nature. The average absolute strength of selection would be inflated if small studies with small, nonsignificant estimates of selection were published less frequently than others. Formal tests for the presence and corrections of publication bias in review data sets are controversial, so in this study, we use Palmer's (2000) graphical test as a simple first approach to the problem.

Palmer (2000) identified three patterns in review data sets that are expected to result from random sampling and therefore are consistent with a lack of publication bias: first, variance of effect sizes decreases with increasing sample size; second, the distribution of effect sizes is similar to the natural distribution (usually the normal distribution) at all sample sizes; and third, the mean effect size is independent of sample size. The inverse relationship between sampling error variance and sample size should produce an inverse relationship between total variance and sample size among selection gradients. However, when publication bias is present, total variance should decline more steeply with sample size than sampling variance. This occurs because total variance of selection gradients from small studies is larger than expected from sampling when small studies with small effect sizes are underrepresented in the literature. Here, we assume that magnitudes of selection are exponentially distributed in nature and that therefore, published magnitudes of selection should be exponentially distributed as well. Our assumption follows from the reviews by Endler (1986) and Kingsolver et al. (2001) that demonstrate that magnitudes of published selection gradients were exponentially distributed. To our knowledge, there are no theoretical expectations for the distribution of selection strengths in nature. In contrast,

publication bias should cause weak, nonsignificant selection gradients from small studies to be underrepresented. Hersch and Phillips (2004) showed that even selection studies with a sample size of 100 have low power (<0.3) to detect moderate selection ($|\beta| = 0.12$) relative to the median selection strength of 0.16 in Kingsolver et al. (2001). Therefore, selection strengths from small to moderately sized studies may not be exponentially distributed with publication bias. Finally, the average selection gradient should be constant across sample size. However, if weak selection gradients at small sample sizes are underrepresented, we do not expect the average selection gradient to change across sample size. This is because selection gradients were roughly centered around 0 (fig. 1), and bias against publishing small positive and small negative selection gradients should be equal. Therefore, we do not test for this pattern in this study.

Using the above approaches, we show that the review data set as a whole has a surprising shape that is the opposite of what sampling error alone would produce. We discuss the likelihood that other biases contribute to the unique distribution of selection gradients versus sample size. Further, we conclude that there is bias against publishing estimates of weak selection, but only for studies with sample sizes of less than 38. Across the whole review, bias in the smallest studies may be counteracted by an overabundance of small estimates among studies with moderate ($N = 88$ – 137) and large ($N \geq 742$) sample sizes. Together, our two main results suggest that our understanding of selection is not strongly biased by these commonly invoked sources of error (Conner 2001; Kingsolver et al. 2001; Hersch and Phillips 2004; Hendry 2005).

Methods

We modeled the decline in total variance of selection gradients across sample size that sampling error variance alone produces. To do this, we subsampled several large empirical data sets and measured the variance of the resulting estimates of selection across multiple traits in the subsampled data sets. The advantage of using real data over using simulated data is that the result will be influenced by real matrices of phenotypic variances and covariances (the P-matrix); this would be difficult to simulate given the essentially infinite number of possible P-matrices.

We obtained the data sets used to calculate the selection gradients of Rausher and Simms (1989), Anholt (1991), Linden et al. (1992), and J. K. Conner (unpublished data). These data sets were chosen for our analyses because of their exceptionally large size and because they cover a wide range of study organisms (plants, vertebrates, and invertebrates), traits (morphological, life history, defense), and

fitness measures (survival and total seed production; table 1). We randomly subsampled each study's data set with replacement to create smaller data sets of $N = 25, 50, 75, 100, 175, 250,$ and 500 . We also resampled each data set with replacement at the same sample size as the original data set. These bootstrap procedures were performed 1,000 times at each sample size. We calculated linear selection gradients (β) on traits in each subsampled data set, using the multiple regression analysis described by Lande and Arnold (1983). The same linear regression model was used for all subsampled data sets derived from the same original data set. In all models, the dependent variable was relative fitness (absolute fitness divided by mean fitness), and the independent variables (the traits) were standardized to a mean of 0 and standard deviation of 1. Selection gradients were calculated for a total of 15 traits across the four studies; note that the number of traits per study is almost identical to the median of four traits per study in the Kingsolver et al. (2001) data set.

To assess the likelihood that sampling variance created a relationship between total variance among selection gradients and sample size in the Kingsolver et al. (2001) review (fig. 1), we compared the effect of sample size on total variance among selection gradients between the subsampled and review data sets. For each of 1,000 bootstrap replicates at each subsample size, we calculated the variance among

the 15 selection gradients, which consisted of sampling and true biological variance among selection gradients. The median of the 1,000 variances at each subsample size was used in further analyses. Selection gradients in the Kingsolver et al. (2001) data set (available at <http://www.bio.unc.edu/faculty/kingsolver/>) were grouped into eight bins circumscribed by the midpoints between the bootstrap subsample sizes: 6–37, 38–62, 63–87, 88–137, 138–212, 213–375, 376–741, 742+; fig. 1). We calculated variance among selection gradients in each bin; these variances also consisted of sampling and true variance.

We performed an ANCOVA of total variance among selection gradients on the inverse of sample size, with selection gradient type (either from subsampling or from the review) as the categorical variable. We reciprocally transformed sample size because sampling variance and sample size are by definition inversely related (eq. [1]). An interaction between selection gradient type and sample size would provide evidence for a difference in the relationship between variance and sample size between the subsampled and review data sets. Because a significant interaction was found, we also performed separate linear regressions of variance on the inverse of sample size for the subsampled and review data sets. To ensure that the effect of sample size on variance among review selection gradients was robust to the location of the bin boundaries,

Table 1: Summary of subsampled data sets

| Study organism and trait | Study | N | Fitness measure | Standardized β (95% confidence interval) |
|--|--------------------------------|-------|--------------------------|--|
| Morning glory (<i>Ipomoea purpurea</i>) | Rausher and Simms 1989 | 1,647 | Lifetime seed production | |
| Leaf area | | | | .68 (.61–.74) |
| Damage by tortoise beetles | | | | -.04 (-.07 to -.01) |
| Damage by flea beetles | | | | .09 (-.06 to .13) |
| Damage by other beetles | | | | -.08 (-.12 to -.05) |
| Damselfly (<i>Enallagma boreale</i>) males | Anholt 1991 | 933 | Survival to adulthood | |
| PC 1 (body dimensions, emergence time) | | | | -.09 (-.59 to .41) |
| PC 2 (body dimensions, emergence time) | | | | .84 (.16–1.56) |
| Damselfly females | Anholt 1991 | 878 | Survival to adulthood | |
| PC 1 | | | | -.25 (-.59 to .11) |
| PC 2 | | | | .48 (.13–.19) |
| Great tit (<i>Parus major</i>) | Linden et al. 1992 | 5,828 | Annual survival | |
| Hatching date | | | | -.30 (-.40 to -.21) |
| Clutch size | | | | -.05 (-.19 to .08) |
| Body mass | | | | .11 (.00–.22) |
| Wild radish (<i>Raphanus raphanistrum</i>) | J. K. Conner, unpublished data | 742 | Lifetime seed production | |
| Ovule number | | | | .11 (.03–.19) |
| Flower number | | | | 1.22 (1.08–1.37) |
| PC 1 of floral dimensions | | | | .02 (-.06 to .10) |
| Stigma exertion | | | | -.10 (-.19 to -.02) |

Note: The 95% confidence intervals of selection gradients (β) are from bootstrapping the full data set 1,000 times (see "Methods" for details). PC = principal component.

we also grouped selection gradients into bins of regularly spaced sample size intervals of $N = 25, 50, 75, 100,$ and 150 (when there were fewer than 25 estimates in a bin, we combined the bin with the adjacent bin with the fewest gradients). Altering bin locations did not affect the ANCOVA results, so we present only results with the original bins.

A correlation between sample size and true variance among selection gradients across different traits or fitness measures could have hidden the expected inverse relationship between total variance and sample size in the review. Positive correlations between sample size and true variance across categories of selection could have led to a lack of relationship between sample size and total variance in the review; negative correlations could have led to a more exaggerated inverse relationship between variance and sample size in the review. To test for such a correlation, we grouped selection gradients by type of trait measured (behavioral, life history, morphological, principal components, other) and calculated median sample size and variance among selection gradients for each trait type. We then calculated the Spearman rank correlation coefficient (ρ) between median sample size and variance across all five trait types. We also grouped selection gradients according to type of fitness measurement (fecundity, mating success, net reproductive rate, survival, total fitness, other) and calculated the Spearman rank correlation coefficient between median sample size and variance among selection gradients across the six fitness measures.

To assess publication bias, we performed the Kolmogorov-Smirnov goodness-of-fit tests on the null hypothesis that selection strengths ($|\beta|$) in each sample size bin were drawn from an exponential distribution. A small test statistic (D) indicates that the sample distribution is close to exponential. If publication bias prevented estimates of weak selection from being included in the review, we would expect fewer weak selection strengths in small sample size bins than predicted from the exponential distribution. To determine whether we could detect publication bias using Kolmogorov-Smirnov tests, we compared the distribution of selection strengths in the three largest sample size bins ($N > 213$) with an exponential distribution before and after introducing a publication bias. The distribution of selection gradients in these large sample size bins should not already be affected by publication bias because of the large power of these estimates to detect selection. We introduced publication bias to the data set by removing estimates that came from studies that had no significant selection gradients. All statistical tests were performed in JMP, version 5.0.1.2 (SAS 2003).

Results

In our simulation of the effect of sample size on total variance among selection gradients, the median total variance among the 15 selection gradients from the subsampled data sets declined curvilinearly as subsample size increased. With increasing subsample size, total variance approached the true variance among the 15 selection gradients, as calculated from the full data sets (fig. 3). As expected, increasing sample size yielded diminishing returns in minimizing sampling error; most of the decline in sampling error occurred between sample sizes 25 and 250. The effect of sample size on variance differed between the subsampled and review data sets; there was a significant interaction between data type and sample size in the ANCOVA ($F = 138.2, df = 1, 15, P < .0001$). The linear relationship between variance and the inverse of sample size was strong and significant in the subsampled data sets ($R^2 = 0.99, P < .0001$), yet there was no such relationship in the Kingsolver et al. (2001) data set above the smallest sample size bin ($R^2 = 0.39, P = .10$; fig. 3).

There was no evidence for the positive relationship between median sample size and variance among selection gradients across different types of traits or fitness measures that would explain this lack of decline in variance with increasing sample size in the review data set. Correlations between sample size and variance among selection gradients were far from significant, and the trends were negative rather than positive (across trait types: $\rho = -0.60$,

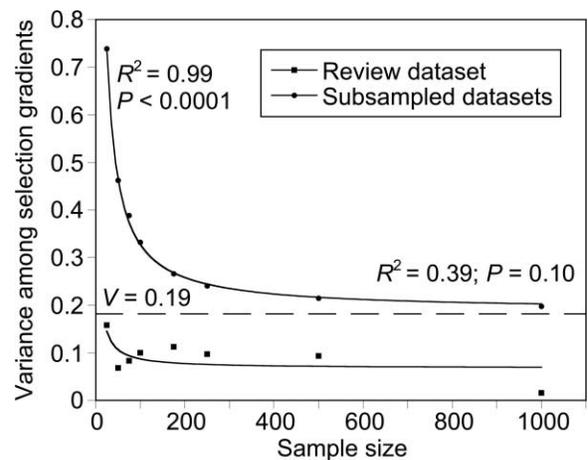


Figure 3: Median variance among the 15 selection gradients from each subsample size are plotted as circles. Variances of selection gradients in the sample size bins from the review data set (Kingsolver et al. 2001) are plotted as squares. The curves represent linear regressions of variance on the inverse of sample size; the R^2 and P values result from the regressions. The variance among the 15 selection gradients calculated from the full data sets is depicted as a dashed horizontal line ($V = 0.19$).

$P = .28$, $N = 5$; across fitness types: $\rho = -0.37$, $P = .47$, $N = 6$).

We found evidence for publication bias in the review among the 256 selection gradients measured with sample sizes of less than 38. The distribution of selection strengths in this category was significantly different from exponential, which we assumed to be the distribution of strengths of selection in nature. Figure 4 shows that this difference is due to a lack of weak selection gradients. However, this figure also shows that bias against publishing weak estimates is not large; 24% of estimates (or 61) have magnitudes of less than 0.1. The distributions of selection strengths were also significantly different than the exponential distribution in the 88–137 and 742+ sample size categories, but here there were actually more weak selection gradients than expected from an exponential distribution. The Kolmogorov-Smirnov test was sensitive to

publication bias in our test. The 263 selection strengths with sample sizes greater than 213 were exponentially distributed before the simulated bias was imposed (the bias removed 20 selection gradients) but not afterward ($D = 0.063$, $P = .083$ vs. $D = 0.075$, $P = .026$).

We also did not observe the trend in total variance across sample size that publication bias is expected to produce (fig. 3). If there are fewer estimates of weak selection from small studies, total variance among selection gradients at small sample sizes should be greater than expected. Therefore, publication bias should result in total variance declining more steeply with increasing sample size than the decline of sampling variance alone. However, as noted above, we observed that the relationship between variance and sample size in the review was more shallow than what sampling variance produced in the subsampled data sets.

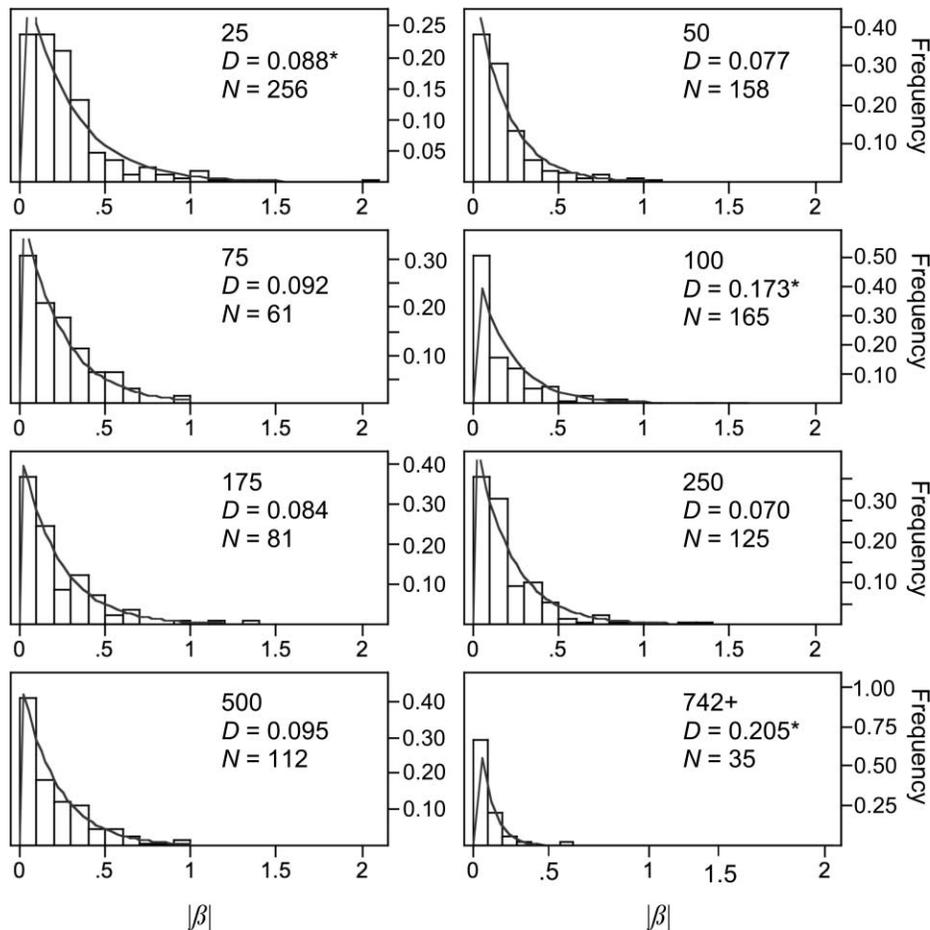


Figure 4: Frequency distributions of selection strengths in each of the sample size bins from Kingsolver et al.'s (2001) data set. The lines represent best-fit approximations of exponential distributions. Kolmogorov-Smirnov (D) tests for deviations from the exponential distribution in each bin are presented. Small values of D indicate that the distribution is close to exponential. Asterisks indicate that the distribution is significantly different from exponential at the $P < .05$ level.

Discussion

Central to evolutionary biology are questions concerning the strength, shape, and frequency of selection in nature. Accordingly, reviews of published selection gradients, the standard measure of selection, are received with intense interest and scrutiny. This study follows others (Conner 2001; Hereford et al. 2004; Hersch and Phillips 2004) in examining the Kingsolver et al. (2001) data set for potential biases. After we explicitly tested for the presence of bias from sampling error or publication practices, our study yielded two results that were surprising: first, total variance of selection gradient estimates did not decrease with sample size and therefore did not mimic the relationship between sampling error variance and sample size, and second, although there is evidence for bias against publishing weak selection gradients from the smallest studies, there are excess weak selection gradients from moderately sized and very large studies.

We expected that total variance in the review data set would decrease curvilinearly with sample size if there were no publication bias and all selection gradients in all sample size bins were randomly or equally drawn from the same distribution of selection in nature (and therefore if true variance were constant across sample size). At first glance, the form of publication bias that Hersch and Phillips (2004) modeled—when studies are published only if they have at least one significant selection gradient—appears to be abundant in the review: 89% of studies in Kingsolver et al.'s (2001) data set presented at least one statistically significant selection gradient. Hersch and Phillips (2004) showed that such bias could produce as much as a 10-fold inflation in the estimate of magnitudes of selection at small sample sizes and a clear pattern of declining magnitudes of selection with increasing sample size.

Using Palmer's (2000) criteria, we found evidence for publication bias in the review only among studies with sample sizes of less than 38 and only among selection gradients with magnitudes of less than 0.1. Given that almost a quarter of the estimates from these small studies have magnitudes of less than 0.1, publication bias may not strongly bias the median estimate of selection upward. Additionally, in studies with sample sizes of between 88 and 137 or greater than 742, there were excess weak selection gradients. This excess may counterbalance the effect of publication bias at the smallest sample sizes in the estimation of overall median magnitude of selection in nature. Publication bias may not be as strong as expected because there were many nonsignificant selection gradients published in studies that included one or more significant ones. These conclusions are contingent on the assumption that selection strengths are exponentially distributed. Theoretical examinations of the expected distribution of se-

lection in nature are needed because it is difficult to explain the overabundance of estimates of weak selection at moderate and large sample sizes in the review and to explain the overall exponential distribution of selection gradients in the Kingsolver et al. (2001) data set.

We did not find the inverse relationship between sample size and variance among estimates of selection in the review that we predicted sampling should have created. This negative result could be explained if estimates of selection for different types of traits or fitness measures were measured with different median sample sizes (for experimental reasons) and had different amounts of true variance in nature (for biological reasons). If types of selection with small amounts of true variance were measured using small sample sizes, then the expected inverse relationship between sample size and variance among estimates could be negated. In contrast, we showed that there was a nonsignificant trend for variance among selection gradients to increase as sample size declined across trait and fitness categories. Therefore, the relationship between sample size and variance across categories of selection does not explain the lack of relationship between variance and sample size in the review.

Certainly, there are other ways of categorizing selection gradients, and the lack of relationship in the review data set merits further exploration. Given that our resampling showed that sampling error variance can add more than 1.7 times the amount of true variance to the total variance at sample sizes of less than 100 (fig. 2), it is extremely surprising that we did not see any evidence of a decline in sampling error variance with increasing sample size in the Kingsolver et al. (2001) data set. Nonetheless, we are left with the conclusion that sampling error variance has not biased estimation of the average magnitude of selection from the review. Further, our results suggest that large magnitudes of selection gradients at small sample sizes should not be dismissed as simple artifacts of sampling error. The lack of estimates of strong selection at the largest sample sizes probably resulted because the few studies that used large sample sizes were not a representative sample of selection. For example, it is possible that researchers with prior information of weak selection in their system used large sample sizes to increase power of detection. Thus, we provide further evidence for Endler's (1986) conclusion that selection in nature can sometimes be very strong.

Significant biases in the review of selection gradients have been broadly accepted in the literature (Conner 2001; Hoekstra et al. 2001; Kingsolver et al. 2001; Hereford et al. 2004; Hersch and Phillips 2004; Hendry 2005). Indeed, it was recently opined that "biologists may be deluding themselves if they think they have a good handle on ... selection" based on reviews published so far (Hendry 2005,

p. 694). This study demonstrates the importance of combining explicit tests for biases with models of potential effects of biases on review data sets. We were able to show that sampling error has not had a strong impact on the data set of published selection gradients. Further, we showed that publication bias is apparent at sample sizes of less than 38 but that an excess of weak selection gradients at larger sample sizes may prevent this bias from altering the estimate of median magnitude of selection. We echo others in calling for evolutionary biologists to broaden the range of traits and modes of selection that they study in the future because perhaps the largest source of bias in the Kingsolver et al. (2001) review, which limits its broad applicability, comes from the predominance of estimates of linear selection on morphological traits.

Acknowledgments

We thank C. Brassil, D. Houle, C. McCulloch, P. Phillips, and D. Schluter for clarifying conversations about the ideas in this article and A. Jarosz, R. La Rosa, A. Malcolm, C. Mills, A. Royer, and D. Schemske for comments on the manuscript. We also thank B. Anholt, M. Linden, and E. Simms for allowing us full access to their data. This work was funded in part by a National Science Foundation predoctoral fellowship and a Michigan State University Distinguished Fellowship to F.N.K. and by the National Science Foundation under grant DEB 0108354 to J.K.C. This is Kellogg Biological Station contribution 1425.

Literature Cited

Albert, A. Y. K., and S. P. Otto. 2005. Sexual selection can resolve sex-linked sexual antagonism. *Science* 310:119–121.

- Anholt, B. R. 1991. Measuring selection on a population of damselflies with a manipulated phenotype. *Evolution* 45:1091–1106.
- Conner, J. K. 2001. How strong is natural selection? *Trends in Ecology & Evolution* 16:215–217.
- Endler, J. A. 1986. *Natural selection in the wild*. Princeton University Press, Princeton, NJ.
- Hendry, A. P. 2005. The power of natural selection. *Nature* 433:694–695.
- Hereford, J., T. F. Hansen, and D. Houle. 2004. Comparing strengths of directional selection: how strong is strong? *Evolution* 58:2133–2143.
- 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. *Proceedings of the National Academy of Sciences of the USA* 98:9157–9160.
- 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. *American Naturalist* 157:245–261.
- Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–1226.
- Linden, M., L. Gustafsson, and T. Part. 1992. Selection on fledgling mass in the collared flycatcher and the great tit. *Ecology* 73:336–343.
- Palmer, A. R. 2000. Quasireplication and the contract of error: lessons from sex ratios, heritabilities, and fluctuating asymmetry. *Annual Review of Ecology and Systematics* 31:441–480.
- Rausher, M. D., and E. L. Simms. 1989. The evolution of resistance to herbivory in *Ipomoea purpurea*. I. Attempts to detect selection. *Evolution* 43:563–572.
- Roff, D. A., and T. Mousseau. 2005. The evolution of the phenotypic covariance matrix: evidence for selection and drift in *Melanoplus*. *Journal of Evolutionary Biology* 18:1104–1114.
- SAS. 2003. JMP. SAS Institute, Cary, NC.

Associate Editor: Edmund D. Brodie III
Editor: Michael C. Whitlock