

Criteria for Testing Character Displacement

Dolph Schluter (1) purports to demonstrate that competition among populations of the threespine stickleback leads to character divergence, based on an experiment in two divided ponds. Both halves of each pond were stocked with individuals from a benthic stickleback species, and individuals of a second, limnetic species were then added to a randomly chosen side of each pond. Thus, the half-pond is the experimental unit, and there are two replications of the comparison between treated and control halves. The responses of interest are based on the growth rates of individuals of the benthic species, in the presence and absence of the limnetic species.

A key part of Schluter's evidence that "the presence of the limnetic species altered natural selection in the target species" is in table 3 of his report, which shows "growth differentials" (slopes of linear regressions of log growth rate versus a morphological index reflecting the continuum between benthic and limnetic characteristics) in the four half-ponds. If competition from the limnetic species is most severe for individuals of the target species having limnetic characteristics, then the growth differentials in the experimental half-ponds should be more negative than those in the control half-ponds. However, the difference between slopes in the experimental and control halves, accounting for the pairing within pond, is not statistically significant (paired t test on the four slopes in table 3 of the report; $t^* = -2.826$, $df = 1$, two-tailed test, $P = 0.22$). Schluter reports an apparently erroneous P value of 0.016 for a one-tailed test (which implies that with a two-tailed test, $P = 0.032$).

Schluter further reports a significant correlation between growth differentials and the final densities of limnetic fish in the half-ponds. His method, involving regression of three independent contrasts of the four slopes in table 3 of his report against the corresponding contrasts of the four measured densities of limnetic fish, is flawed because one of the response contrasts depends on the magnitude of the pond effect and has a variance different from that of the other two contrasts (2). In any event, the most direct test of the association between the limnetic fish treatment and the growth differentials in table 3 of the report remains the nonsignificant paired t test.

The same inappropriate methodology (2) is used to support an association between mean growth rates and the total numbers of fish in the four half-ponds (table

2 of the report by Schluter). A direct test of the effect of the experimental manipulation (which led to the different fish densities) on mean growth rates shows no evidence for an effect of the treatment (paired t test on the four rates in table 2; $t^* = -1.1951$, $df = 1$, two-tailed test, $P = 0.44$).

Another line of evidence that the presence of limnetic fish affects growth rates is figure 1 in the report by Schluter, which is a plot of transformations of log growth rate versus morphological index. This plot appears to show a decreasing trend for fish in the presence of the limnetic species and no trend for fish in the absence of the limnetic species. My statistical modeling of the points shown in the graph does not indicate different slopes for the experimental and control groups (3), but, even if it did, this would not justify the conclusion that the presence of the limnetic species was driving the difference. Without information on the pond-to-pond variation in the slopes of such regression lines, which is lost in the pooled data of figure 1, we cannot judge whether or not the limnetic fish treatment is causing additional variability.

The trends in Schluter's data are consistent with effects of competition from the limnetic species, but the key statistical comparisons of experimental units do not support his conclusion that "resource competition promotes morphological diversification in a radiating lineage." The lack of statistical significance does not necessarily mean that competition is unimportant, as only large effects would be detectable in an experiment with two replicates.

Paul A. Murtaugh

Department of Statistics,
Oregon State University,
Corvallis, OR 97331, USA

REFERENCES AND NOTES

1. D. Schluter, *Science* **266**, 798 (1994).
2. Let Y represent the response (for example, growth differential, or mean growth rate) and X represent the explanatory variable (for example, density of limnetic fish, or total fish density), with subscripts denoting half-ponds as follows: 1 = experimental side of pond 1; 2 = control side of pond 1; 3 = experimental side of pond 2; and 4 = control side of pond 2. Suppose μ is the mean response in pond 1 with $X = 0$; α is the effect of pond 2; βX is the effect of X units of the explanatory variable; and the ϵ 's are random errors. Then we can model the responses in the four half-ponds as follows:

$$Y_1 = \mu + \beta X_1 + \epsilon_1 \quad Y_3 = \mu + \alpha + \beta X_3 + \epsilon_3$$

$$Y_2 = \mu + \beta X_2 + \epsilon_2 \quad Y_4 = \mu + \alpha + \beta X_4 + \epsilon_4$$

The three independent contrasts used by Schluter (personal communication) have expectations $E(Y_1 - Y_2) = \beta(X_1 - X_2)$; $E(Y_3 - Y_4) = \beta(X_3 - X_4)$; and

$$E\left(\frac{Y_3 + Y_4}{2} - \frac{Y_1 + Y_2}{2}\right) = \alpha + \beta\left(\frac{X_3 + X_4}{2} - \frac{X_1 + X_2}{2}\right)$$

When these response contrasts are regressed against the corresponding contrasts in X (that is, $X_1 - X_2$, $X_3 - X_4$, and $\frac{X_3 + X_4}{2} - \frac{X_1 + X_2}{2}$), the resulting slope will provide a biased estimate of β because of the dependence of the third response contrast on the pond effect, α . Furthermore, the variance of the third response contrast is different from that of the other two contrasts, violating an assumption of the regression model. If $\text{Var } \epsilon_i = \sigma^2$ for all i , and we assume a non-zero correlation, ρ , between error terms from the same pond [that is, $\text{Cor}(\epsilon_1, \epsilon_2) = \text{Cor}(\epsilon_3, \epsilon_4) = \rho$], then $\text{Var}(Y_1 - Y_2) = \text{Var}(Y_3 - Y_4) = 2\sigma^2(1 - \rho)$, and

$$\text{Var}\left(\frac{Y_3 + Y_4}{2} - \frac{Y_1 + Y_2}{2}\right) = \sigma^2(1 + \rho)$$

A possible modification of Schluter's test is to fit a regression line through the origin using just the first two contrasts, which have equal variance, and expectations not involving α . This approach does not find a significant association between Y contrasts and X contrasts for either the analysis of growth differentials by limnetic fish density ($\beta = -0.0001$, $df = 1$, two-tailed $P = 0.13$) or the analysis of mean growth rate by total fish density ($\beta = -0.0054$, $df = 1$, two-tailed $P = 0.20$).

3. Let GROWTH = log growth rate, MI = morphological index, and COMP = indicator for presence (1) or absence (0) of limnetic fish. Then the "full" regression model fit to 88 points read off figure 1 in the report by Schluter is: GROWTH = 3.8887 - 0.0406 COMP - 0.0013 MI - 0.0307 COMP-MI. The coefficient for the interaction term, COMP-MI, is nonsignificant ($P = 0.087$), which indicates that we cannot reject the hypothesis that the slope of GROWTH versus MI is the same in the presence and in the absence of limnetic fish. As each symbol in figure 1 of the report is an "average of three adjacent points," and "growth rates within each treatment were pooled to the same mean," we cannot be sure how an analysis of the full set of untransformed data would come out.

14 December 1994; accepted 25 January 1995

The study by Schluter (1) has problems in its design, conduct, and analysis; at the request of the editors, we limit our comment to the first two areas. The first problem concerns experimental design. The treatment in which the putative competitor was introduced confounds two factors: presence of "heterospecific" individuals and a 1.7-fold increase in total fish density. Both additive and replacement designs are appropriate for competition experiments, but each is used to test distinct hypotheses (2, 3). The additive design used by Schluter would be appropriate were he simply attempting to detect competition between his two "species," that is, to determine whether the two "species" have a joint carrying capacity (or density function). Evidence of character displacement requires a distinct effect: that heterospecific competitors differentially affect some subset of the population relative to intraspecific competition. If addition of heterospecifics produces an identical effect to that of adding the same number of conspecifics, then character displacement, as classically defined, should not be invoked.

Given this criterion, experimental tests for character displacement require treatments in which total densities are held constant to determine whether inter- and intraspecific competition are equivalent from the perspective of the focal "species." Lacking such controls, treatment effects cannot be attributed uniquely and unambiguously to interspecific competition as distinct from increased fish density (4). The appropriate null hypothesis that must be rejected here is that the two "species" are equal competitors with regard to all phenotypes of the focal "species." The appropriate minimal test is a replacement design holding densities in the control (single "species") and competition (two "species") treatments constant, and testing for competitive equivalence (3) between the limnetic and different segments of the phenotypic distribution of the focal "species." Confounding is a serious, common error in ecological experiments (5), and while the specific type of confounding present in this study is appropriate in testing certain hypotheses regarding competition, it constitutes a fatal flaw in a test for character displacement. Thus, Schluter has not unambiguously established interspecific competition as the cause of the purported differential effect on phenotypes of the focal "species" because the appropriate hypothesis-specific control was not included.

We see two difficulties in the conduct of the experiment concerning extrapolation back to nature. The first concerns the use of hybrids. Schluter argues (1, p. 799), "Hybridization is a valid manipulation because all previous crosses between closely related freshwater sticklebacks have not revealed any intrinsic reduction in offspring viability." This may be true, but other problems arise from the use of hybrids. First, the pool of fish on which effects of interspecific competition was assayed was assembled using equal numbers of offspring from three experimental crosses (1), and hybrid fish were created using the limnetic species (the species used as a putative competitor), as one parental type. Because the limnetic species contributed half the genes to the $L \times C$ hybrids in which the greatest effect of competition was expected, the question essentially asked is: Do the limnetics have a greater effect on their own hybrid-genetic progeny than on genetically unrelated individuals? This is not an illuminating question in the context of character displacement: Because of this relatedness, Schluter is not studying an interspecific competitive interaction, which is fundamental to the hypothesis that interspecific competition drives character displacement. Second, there is both morphological and genetic correspondence between these $L \times C$ hy-

brids and the limnetic competitor, that is, genotypes and morphology are confounded. Hence, the particular hybrids ($L \times C$) most likely to compete with the limnetics are also most genetically similar to limnetics. Further, the use of hybrids introduces the potential for heterotic genetic effects (6) in responses of hybrids to the limnetic morph that are not genetically accessible to individuals of a single species in nature. Hence, one cannot argue that any effect observed in the experiment would be evidenced by fish in nature. Although the experiment is internally consistent in that the putative control also consisted of hybrids, the responses of the experimental fish cannot be unambiguously attributed to either their morphological phenotypes or to their artificial hybrid genetic constitution. That is, any resulting selection differential may have little to do with morphological similarity. Experimental expansion of a phenotypic distribution for the purpose of analyzing selection is creative, but is interpretable only if the novel phenotypes so produced are not confounded with other, unmanipulated parts of the phenotype, or with changes in genetic constitution (7).

The second difficulty arises from the relationship of experimental to natural densities and the background conditions of the experimental ponds. Densities in the experiment were not matched to natural densities. Rather, "Densities were set such that growth rate of pond fish would equal that in the wild, as judged from an earlier pilot experiment . . ." (1, p. 800). This seems reasonable, except that experimental ponds were fish-free for 2 years before the experiment, thus accumulating far greater invertebrate biomass and attaining far different species composition (8) than would be expected in natural ponds containing sticklebacks. Experimental densities far above natural densities would be required to achieve natural growth rates, thus increasing encounter rates and the potential for interference competition. Also, resource competition (the presumed mechanism driving character displacement) would have occurred against an artificial resource background.

It is tempting to accept the results of provocative, high-profile experiments at face value (9). Although we strongly support using experiments to validate mechanisms hypothesized from comparative data (10), such experiments must comply with rigorous standards of design, conduct, and analysis. Because of deficiencies on all three points, this is not a landmark demonstration of character displacement.

Joseph Bernardo

Department of Zoology,
University of Texas,
Austin, TX 78712, USA

William J. Resetarits Jr.
Center for Aquatic Ecology,
Illinois Natural History Survey,
Champaign, IL 61820, USA

Arthur E. Dunham
Department of Biology, Leidy Laboratories,
University of Pennsylvania,
Philadelphia, PA 19104, USA

REFERENCES AND NOTES

1. D. Schluter, *Science* **266**, 798 (1994).
2. C. T. de Wit, *Verslagen van Landbouwkundige Onderzoekingen*, no. 66.8 (1960); A. J. Underwood, in *Community Ecology: Pattern and Process*, J. Kikkawa and D. J. Anderson, Eds. (Blackwell, London, 1986).
3. J. E. Fauth, W. J. Resetarits, H. M. Wilbur, *Oikos* **58**, 91 (1990).
4. An equally likely explanation for the purported differential effect on the most limnetic-like hybrid fish is that these fish are poorer competitors. Thus, an increase in the overall competitive environment (increased fish density) has greater impact on the poorer competitor regardless of whether that increase is accomplished by adding limnetics or conspecifics (that is, other individuals of the target form). The design (7) cannot distinguish this possibility from Schluter's interpretation.
5. N. Fowler, *Bull. Ecol. Soc. Am.* **71**, 161 (1990).
6. M. L. Arnolds and S. A. Hodges, *Trends Ecol. Evol.* **10**, 67 (1995).
7. B. Sinervo and R. B. Huey, *Science* **248**, 1106 (1990); J. Bernardo, *Trends Ecol. Evol.* **6**, 1 (1991).
8. D. J. Hall, W. E. Cooper, E. E. Werner, *Limn. Oceanogr.* **15**, 839 (1970); L. B. Crowder and E. E. Cooper, *Ecology* **63**, 1802 (1982); M. A. McPeck, *ibid.* **71**, 83 (1990).
9. P. R. Grant, *Science* **266**, 746 (1994); S. Pimm, quoted in J. Weiner, *ibid.* **267**, 30 (1995).
10. J. Bernardo, *Am. Nat.* **143**, 14 (1994); W. J. Resetarits, *Ecology* **72**, 1782 (1991); A. E. Dunham, *Ecol. Monogr.* **50**, 309 (1980).

24 January 1995; accepted 22 March 1995

Response: Murtaugh mentions three concerns about my statistical analyses (1), which I address in turn.

1) The paired comparison of growth differentials was incorrect.

This criticism is accurate. The growth differentials were properly listed, but the results of the paired t test should have read as follows: $t = -2.826$, $df = 1$, $P = 0.11$; one-tailed test. Complementary analyses suggest that the treatment effect was nevertheless real. Growth differentials in the presence of the competitor were significantly negative in both ponds, but were near zero in the competitor's absence (1). The slopes differed significantly between treatment and control sides [Fisher combined probability test (2); $\chi^2 = 9.49$, $df = 4$, $P = 0.050$]. The steepness of the growth differential increased with increasing competitor density. Spearman rank correlations between growth rate and morphological index also differed between treatments (one-tailed paired t test; $t = -6.708$, $df = 1$, $P = 0.047$). Two ponds (replicates) can withstand only so much statistical analysis; but the important point is that the growth patterns are in the predicted direction and that

despite low power most analyses detect a difference.

2) The regressions based on independent contrasts were flawed.

This concern is unwarranted, as all possible corrections lead to the same answer. The criticism is based on the fact that the variance of the between-pond contrast is not the same as that of the two within-pond contrasts. However, adjusting for heterogeneity in variance by standardizing the contrasts (3) has no effect on the results, which is why I neither standardized nor elaborated on the method in my report. At one extreme is the possibility that ponds do not differ (except in ways caused by the experiment), in which case the between-pond contrast has one-half the variance of the other contrasts. When the contrasts were standardized accordingly, the correlation between growth differential and competitor density ($r = 0.97$, $P = 0.018$; one-tailed test) was nearly identical to that originally reported (1). At the other extreme is the possibility of a large pond effect. For example, if the variance of the between-pond contrast is 10 times that of the other two contrasts, standardizing yields a result not much different from the first ($r = 0.98$; $P = 0.012$; one-tailed test). The same was true when the method of independent contrasts was used to compare growth rate with total fish density ($r = 0.94$, $P = 0.029$ and $r = 0.95$, $P = 0.024$; one-tailed tests). Consequently, the method is justified and the original results hold.

3) The data plotted in figure 1 of my report do not show a significant treatment effect.

The purpose of this figure was solely to depict the shape of the relationship between growth rate and the morphological index. The figure does not contain data that would allow a valid statistical test of treatment effect. The points in the figure were averages of trios of observations, and combine data from both ponds. Murtaugh's analysis is pseudoreplicated because it assumes that individual observations within ponds are independent, an unlikely situation. However, were independence a valid assumption, the actual data would still support a significant treatment effect on growth differentials [one-tailed test of heterogeneous slopes; $F(1,260) = 2.916$, $P = 0.044$].

Bernardo *et al.* list two general design concerns. Both were raised in my report (1) where little space was available for details.

1) A different experimental design should have been used.

My design, hereafter called (A), had two treatments: a control, in which the target species occurred alone; and an experimental, in which the target and a competitor species (the limnetic) were present. The starting density of the target was constant. The advantage of (A) is that any change in natural selection on the target species can be traced to the presence of the competitor species. Its weakness is that (A) cannot rule out the possibility that changes in selection were solely the result of an increased density of fish. A reasonable expectation is that increased density alone would effect all phenotypes of the target equally, which did not happen. Moreover, details of habitat use corroborate the frequency-dependent prediction: limnetics specialized on plankton, and the more planktivorous phenotypes of the target population suffered the brunt of their presence. In contrast, the "density-only" hypothesis predicts that the planktivorous phenotypes would suffer disproportionately even if the competitor species avoids plankton. A mechanism for this is difficult to envision.

Bernardo *et al.* suggest a second design (B) in which the total density of fish in both treatments is constant. This is accomplished by replacing individuals of the target species by the same number of competitors (limnetics) in the experimental treatment. The strength of this design is that any difference between treatments in natural selection can be attributed to a change in the frequency of different phenotypes overall. Its weakness is that the effects may be a result of a lower population density of the target species rather than of the presence of the competitor species. Design (B) therefore does not test whether presence of the competitor species is the cause of a treatment effect. Hence it is an inferior design.

A third two-treatment design (C) uses two competitor species. A plankton specialist (the limnetic) is added to one treatment, and a specialist on benthos (the benthic species) is added to the other. This has the advantage of varying phenotype frequency while keeping constant both the total density of fish and the density of the target species. I used (A) rather than (C) for two reasons. First, I wanted a baseline measurement of selection on the target species alone (that is, a control). Second, I wished to test specifically whether the differences between the modern limnetic and benthic species are the result of ecological character displacement (4). This was best achieved by using the phenotypes most likely to have been those present in lakes when sympatry

between their ancestors was first established (1).

2) The experimental conditions were not natural.

A concern was that the limnetic competitor was genetically related (through hybridization) to one component of the target population. However, genetic similarity between limnetics and part of the target population was intentional (I am interested to hear how similarity in a heritable suite of ecologically relevant traits might be achieved without genes in common). In figure 1 of my report, I looked for visual evidence that effects of the added competitor were limited to the limnetic-hybrid component, but found none. Other possible effects of hybridization are common to both treatments and are therefore controlled for. Additional design concerns are minor, but surround a deeper question: Are ponds identical to natural lakes? They are not. Nevertheless, ponds are similar to lakes in enough ways that have allowed many important ideas in ecology to be tested using them. But it is essential that a hypothesis such as character displacement be tested with a combination of methods that includes comparisons of wild populations (4).

In sum, the experimental results greatly bolster earlier conclusions based on comparative studies (4) that competition was an important force in the diversification of sticklebacks. They do not constitute final proof of the evolutionary significance of competition in general, which awaits further experimental study especially of other systems.

A final correction: In the summary of my report in that issue's "This Week in *Science*" (p. 709), the phrase "and the first generation of offspring clearly showed an increased divergence between the two types" was incorrect. My experiment was conducted wholly within a generation and involved no evolutionary change between generations.

Dolph Schluter

Zoology Department and
Centre for Biodiversity,
University of British Columbia,
Vancouver, BC V6T 1Z4, Canada

REFERENCES

1. D. Schluter, *Science* **266**, 798 (1994).
2. R. R. Sokal and F. J. Rohlf, *Biometry* (Freeman, New York, ed. 2, 1981).
3. J. Felsenstein, *Am. Nat.* **136**, 569 (1985).
4. D. Schluter and J. D. McPhail, *ibid.* **140**, 85 (1992).

10 January 1995; accepted 25 January 1995