INTEGR. COMP. BIOL., 45:391–404 (2005)

How and When Selection Experiments Might Actually be Useful[1]

REBECCA C. FULLER,[2,*] CHARLES F. BAER, AND JOSEPH TRAVIS

*School of Computational Science, Florida State University, Tallahassee, Florida 32306-4120
Department of Zoology, University of Florida, Gainesville, Florida 32611-8525
Department of Biological Science, Florida State University, Tallahassee, Florida 32306-4340

SYNOPSIS. Laboratory natural selection and artificial selection are vital tools for addressing specific questions about evolutionary patterns of variation. Laboratory natural selection can illuminate whether a putative selective agent is capable of generating long-term, sustained changes in individual traits and suites of traits. Artificial selection is the essential tool for understanding the general evolvability of traits and the extent to which genetic correlations constrain evolution. We review the contexts in which each type of experiment seems capable of offering key insights into important evolutionary issues. We also discuss theoretical and methodological considerations that play critical roles in designing selection experiments that are relevant to evolutionary patterns of trait variation. In particular, we focus on the critical role of selection intensity and the consequences of experiments with different intensities. While selection experiments are not practical in many cases, sophisticated selection experiments—designed with careful consideration of the theory of selection—should be taken beyond model organisms and used in well-chosen natural systems to understand natural patterns of variation.

INTRODUCTION

Selection experiments are vital tools of evolutionary biology and the results of nearly a century's worth of selection experiments have helped establish the genetic component of evolutionary theory (Provine, 1971; Falconer, 1992). In addition, selection experiments have provided stocks that have been useful for many other topics, from estimating mutation rates to understanding the molecular, biochemical, and physiological foundations of trait variation (Hill and Caballero, 1992; Mackay, 2001*a*; Conner, 2003; Garland, 2003). In this paper we suggest that selection experiments are irreplaceable tools for answering questions about adaptation and the genetic basis of adaptive trait clusters (*i.e.*, repeated evolution of suites of traits in particular environments).

We develop this argument in stages. First, we describe the two types of selection experiments and clarify our definition of ''complex traits.'' Next, we describe how these two types of experiments can be deployed for understanding the ecological agents driving adaptation and for untangling the genetic control of correlated responses to selection. More importantly, we discuss the insights that detailed genetic analyses can offer and why those insights are important for understanding patterns of comparative biology. In the subsequent section, we review the issues of experimental design and execution that play critical roles in focusing selection experiments and discuss some challenging choices among experimental possibilities. We conclude by endorsing selection experiments as vital tools for questions about adaptation and comparative biology. However, because the design of these experiments can affect the results so profoundly, we urge researchers to consider carefully the design and execution of selection experiments to ensure that they will answer the question they were conceived to address.

CLASSIFYING EXPERIMENTS AND TRAITS

Varieties of selection experiments

We distinguish two types of selection experiments. In the first, ''laboratory natural selection'' (LNS), the experimenter divides replicate lines among two or more environmental treatments and examines how the experimental stocks respond over time (Rose *et al.*, 1990; Travis and Reznick, 1998). In effect, the scientist creates varieties of ''nature,'' that is, the combinations of experimental populations and environmental conditions, and allows ''nature'' to choose which traits respond. The experimenter controls which forces can act as selective agents by creating environments that differ in one or more key attributes but does not control the strength of selection or the consistency of selection across generations. The experimenter does not control which traits respond to selection, although there is usually an *a priori* expectation, which, obviously, defines the traits in which the experimenter looks for a response.

Laboratory natural selection experiments are used in two ways. First, they are used to scrutinize fundamental predictions of evolutionary theory with limited reference to observed patterns in nature. For example, Lenski and colleagues have manipulated *Escherichia coli* stocks on various food media and in a variety of thermal regimes to examine the roles of adaptation, chance, and history on evolution (Travisano *et al.*, 1995), the nature of local adaptation and performance trade-offs, (Bennett and Lenski, 1993; Cullum *et al.*, 2001), the long-term dynamics of frequency-dependence (Rozen and Lenski, 2000), the mechanisms

[1] From the Symposium on *Selection Experiments as a Tool in Evolutionary and Comparative Physiology: Insights into Complex Traits* presented at the Annual Meeting of the Society for Integrative and Comparative Biology, 5–9 January 2004, at New Orleans, Louisiana.

[2] E-mail: fuller@csit.fsu.edu

maintaining genetic variation (Elena and Lenski, 1997), and the evolution of mutation rates (Sniegowski *et al.*, 1997) (see Feldgarden *et al.*, 2003 for a recent review of LNS experiments with microbes in ecology). In another example, Mueller and colleagues have used manipulations of larval and adult density regimes in *Drosophila melanogaster* to test a variety of predictions derived from density-dependent selection theory (reviewed in Mueller, 1997).

The second use of laboratory natural selection is to test hypotheses about the selective agents responsible for maintaining particular patterns of variation observed in nature. The comparative approach can identify the possible selective agents but will not often be conclusive because phenotypic variation is usually associated with correlated variation in several environmental factors. Laboratory natural selection experiments can isolate the influence of individual factors and test their efficacy as selective agents. For example, Conover and Munch (2002) manipulated size-specific harvesting regimes to test whether they could account for the long-term changes in life history and biomass observed in many exploited fish.

The second type of selection experiment is called ''artificial selection'' (AS). Here individuals are selected to propagate the next generation if they express particular values of specific traits. The experimenter is the direct agent of selection and chooses the trait in advance. The experimenter also chooses the selection regime: method, strength, and consistency. For example, the ''selected'' individuals could be chosen from a truncation process (*e.g.*, all individuals with trait values greater than X are allowed to reproduce), on the basis of relative value (*e.g.*, individuals in the top 25% of trait values are allowed to reproduce), or from a graduated covariance between trait value and reproductive opportunity. The experimenter can vary the strength of selection (*e.g.*, the truncation value or the percentile that is selected) and its consistency (every generation, every other generation, etc.). These experiments are employed for many purposes, from ascertaining the symmetry of responses in a particular trait (*e.g.*, selecting upward and downward, Pitnick and Miller, 2000) to searching for correlated responses to selection on a focal trait (Schwarzkopf *et al.*, 1999) to ascertaining the reciprocity of responses in two genetically correlated traits (Joshi and Thompson, 1995, 1997).

Some workers recognize additional categories of selection experiments but we believe that those experiments can be embraced by our dichotomy. One such experimental variety is termed ''laboratory culling'' (Rose *et al.*, 1990), in which individuals are exposed to an environmental stress and either the survivors (if the stress is lethal, *e.g.*, Juliano and Gravel, 2002) or the hardiest (if the stress is sublethal, *e.g.*, Baer and Travis, 2000) are allowed to reproduce. This method is a variation of laboratory natural selection—the stress is an environmental effect imposed without regard to specific trait values—with the advantage that the experimenter can control the strength and consistency of selection.

Properties of complex traits

There are many ways to define a ''complex trait,'' (see *e.g.*, Weber, 1992, 1996; Reznick *et al.*, 2002; Garland, 2003), but for the purposes of quantitative genetics an explicitly genetic definition is appropriate, and useful. We define a complex trait as any trait determined by many genes; by this definition, complex traits exhibit three consistent properties (reviewed in Lynch and Walsh, 1998). First, their phenotypic variation is based on allelic variation in many genes. Second, the genetic covariation among traits is based either on pleiotropic effects or co-segregation of alleles produced by linkage disequilibrium (LDE). Third, considerable variation in complex traits is induced by environmental effects, which may vary among genotypes.

Evolutionary patterns in complex traits have two consistent features. First, trait variation in nature is usually associated with correlated variation in several environmental factors. Second, several traits usually vary concordantly. These attributes characterize interas well as intraspecific variation (Reznick and Travis, 1996). Selection experiments are designed to gain insight into intraspecific variation, usually with the hope that those insights can illuminate patterns of variation at higher levels (see Travis and Reznick, 1998).

We argue that selection experiments are essential tools for understanding evolutionary patterns in complex traits. Part of their utility resides in the ability of laboratory natural selection to pinpoint the causal agents of selection and help understand how trait variation comes to be associated with environmental variation. In addition, the response to artificial selection and the genetic analyses of the resultant phenotypes can offer insights into how correlated trait variation is developed and maintained. An important part of this argument is that insights from selection experiments into the genetic control of trait variation are not merely interesting to geneticists but are important in understanding the power and consistency of adaptive evolution.

PUTTING SELECTION EXPERIMENTS TO USE

Using LNS to understand adaptation

Almost every study of trait variation in nature includes a correlative study of environmental variation to identify potential agents of selection (Reznick and Travis, 1996). An experimental investigation usually follows. However, given the difficulties involved in good experimentation, one should be certain that environment and phenotype are associated beyond what one would expect by chance. Migration and drift can generate almost any observable pattern of spatial variation in the absence of selection (Kimura and Maruyama, 1971; Hansen *et al.*, 2000; Felsenstein, 2002). Different approaches can be taken to compare patterns against neutral expectations, ranging from assessing

spatial patterns of variation at neutral markers and focal traits (Reznick and Travis, 2001) to comparing QTL patterns among divergent populations (Orr, 1998).

Once a putative adaptive pattern has been identified, subsequent experimental work falls into two categories. In the first, manipulative experiments are used to determine whether the observed trait associations might be produced by phenotypic plasticity (*e.g.*, Trexler and Travis, 1990; Trexler *et al.*, 1990; Fuller and Travis, 2004). In the second, manipulative experiments are used to identify the ecological agents responsible for maintaining trait variation; these experiments in turn can be divided into two groups. One group consists of single-generation, ecological studies designed to identify the selective agents and demonstrate their ability to select for different phenotypes in different locations (Lande and Arnold, 1983). By their nature, within-generation experiments provide no information about the evolutionary *response* to selection and we will not consider them further (see Kingsolver *et al.*, 2001; Stinchcombe *et al.*, 2002 for recent reviews of the conceptual issues). The other group includes laboratory natural selection experiments.

There are advantages and liabilities to each approach to identifying the agents of selection. The short-term, ecologically oriented studies usually have the practical advantage. For one reason, many organisms are not amenable to laboratory natural selection. For another, there may be so many plausible agents of selection that a factorial experiment in laboratory natural selection is impractical. In these cases, the ecological studies are the better choice. But short-term ecological studies have a clear disadvantage; they identify selection gradients (*sensu* Lande and Arnold, 1983) that point to the immediate trajectory that a selective agent would create but they cannot guarantee how much trait change a given selective agent would generate in the long run. This is because selection gradients are likely to change as the range of expressed trait variation changes. Moreover, as those gradients change, the correlated responses to selection may also change, which could make the short-term experiments misleading as to the long-term trajectories expressed by a suite of traits. Laboratory natural selection offers considerable insight into the evolvability of individual traits and trait clusters, a virtue that balances its practical difficulty for many situations.

Two case studies illustrate laboratory natural selection in this context. We draw our first case study from the scrutiny of how the thermal environment affects evolution in *Drosophila* species. The original observations were that several traits (body size, wing size, development time, timing of reproduction, desiccation tolerance, alcohol utilization, etc.) varied significantly with latitude (Stalker and Carson, 1947; Tantawy and Mallah, 1961; Berry and Kreitman, 1993; Robinson *et al.*, 2000). Some of these patterns were repeated across different continents (James *et al.*, 1995, 1997; Azevedo *et al.*, 1998; van'T Land *et al.*, 1999), which suggested that a widespread, uniform agent of selection could be responsible. The most obvious candidate is the thermal environment. Working from this hypothesis, Partridge and colleagues undertook an LNS experiment on *D. melanogaster* by maintaining lines under two thermal environments, 16.5°C or 25°C. These lines diverged in several traits: development time (Huey *et al.*, 1991), thermal tolerance (Huey *et al.*, 1991), thorax size, and wing size (Partridge *et al.*, 1994), and the divergence matched the latitudinal clines. Furthermore, the lines differed in growth efficiency (Neat *et al.*, 1995) and life history (Partridge *et al.*, 1995).

TABLE 1. *Comparative pattern across population in relation to predation regime in* Daphnia.

	Fish present	Fish absent
Main predator	**fish**	**invertebrates**
Lake depth	**deep**	**shallow**
Resource quality	**low**	**high**
Size at maturity	small	large
Age at maturity	young	old
Clutch size	small	large
Relative offspring size	large	small
Negative phototaxis	present	absent

Traits in **bold** represent correlated environmental factors.

While this work identified the thermal environment as an agent of selection that could be responsible for maintaining variation in many traits, it did not rule out the possibility that other selective agents contribute to the pattern. Other aspects of the environment vary with latitude, including, among others, photoperiod, seasonality, and rainfall (Tantawy and Mallah, 1961; Tantawy, 1964; James and Partridge, 1998). This illustrates the dilemma inherent in the use of LNS, which is that the robustness of the inference drawn from these studies depends critically upon the choice of experimental factors or, alternatively, the ability to study many factors. In the *Drosophila* example, Kennington *et al.* (2003) began another LNS experiment in which flies were raised under high and low humidity with temperature being held constant. They found that larger wings evolved under low relative humidity, a pattern that also matches the latitudinal cline.

A different perspective on laboratory natural selection emerges from the study of *Daphnia*. A suite of traits show consistent inter- and intraspecific variation among environments (Table 1; see Lynch, 1980; deMeester, 1996). Large adult size, late age at maturity, large clutch size, small offspring size, and a lack of negative phototactic behavior are associated with the absence of predatory fish. However, other environmental factors are associated with the presence or absence of fish, including the density and species composition of invertebrate predators, water depth, thermal regimes, and phytoplankton abundance (Desmarais and Tessier, 1999; Tessier and Woodruff, 2002).

Spitze (1991) used LNS to test the hypothesis that trait variation in *D. pulex* could be driven by differences in the intensity of predation on earlier instars by

larvae of the fly *Chaoborus americanus*. Note that both *Chaoborus* and *D. Pulex* are commonly found in lakes and ponds without fish. Spitze established replicate lines with and without *Chaoborus* and found that, over 8–12 generations, the predator caused lines to evolve larger body sizes and larger clutch sizes but no changes in the other traits. At one level, this was not surprising because there was no evidence for strong genetic correlations among traits (Spitze *et al.*, 1991). At another level, it was surprising because it suggests that different ecological agents mold different components of these trait clusters, which evolve independently of one another, with the net result that the same pattern emerges time and again in different localities. Obviously, additional experiments are needed, but if further experiments confirmed this implication, it would be strong testimony to the power of selection to mold integrated phenotypes one trait at a time, time and again. It also suggests that LNS should be complemented by additional studies, focused on specific trait combinations and their genetic control. This is the domain of artificial selection.

Using AS to understand phenotypes

At their simplest, artificial selection (AS) experiments answer the question of whether a trait can respond to selection. But, given that AS experiments are difficult to execute, they are not the best way to answer that question—one can do almost as well in most cases by estimating breeding values. We argue that AS should be used as the primary tool to identify correlated responses to selection and estimate the extent to which those correlations prevent (or at least hasten) individual traits from evolving to their naturally selected (local) optima. In the next sections we describe each of these deployments of AS and why they are important.

Correlated characters and constraints in evolution

Genetic covariances loom large in evolutionary biology, for multiple reasons. First, correlated responses produced by genetic covariances are, potentially, a simple, explanation for trait clusters like those observed in *Daphnia* populations in different habitats (Table 1). In principle, such clusters could be molded trait by trait or through direct selection on one or a few traits and correlated responses by the others (Lande, 1979). This question is at the heart of developing a predictive theory of multivariate trait evolution. If correlated responses are not important or if selection, in the long run, can overcome genetic covariances (see Archer *et al.*, 2003), then predictive theories can be ecological in orientation and based on optimality criteria. If correlated responses are important, then predictive evolutionary theories must include genetics.

The evidence in hand on trait clusters is mixed. On the one hand, many AS experiments have detected correlated responses that suggest substantial, and often non-intuitive, genetic covariances among traits. For example, Pitnick and Miller (2000) selected upward and downward on testis size in *Drosophila hydei* and found that thorax length, sperm length, egg-to-adult development time, and post-eclosion maturation time all had positive correlated responses to selection for increased testis size. On the other hand, Baer and Lynch (2003) tested whether strong genetic correlations among traits could create patterns in *Daphnia* life-history variation. They selected on size at maturity, sorting among clones of *D. pulicaria*, and looked for correlated responses in a variety of other traits. They found little evidence for strong genetic covariances that would produce correlated responses. This result reinforces Spitze's (1991) conclusions and indicates that selection acts directly on individual traits to create the associations among them.

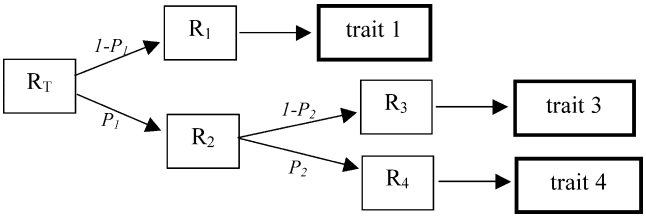

FIG. 1. Redrawn from Worley *et al.*, 2003. Resource allocation between 3 traits (traits 1, 3, and 4) involving two allocation ''decisions'' (P_1 and P_2). R_T refers to total resources. R_{1-4} refers to resource pools resulting from allocation decisions that affect trait values.

The second arena in which genetic covariances play a starring role is the study of constraints on adaptation. For example, life history theory is predicated on unavoidable trade-offs among individual traits, say, between early and late fecundity or between individual offspring size and offspring number, that constrain the possible phenotypes that selection can produce (Stearns, 1992; Roff, 1992, 2002). The intuitive understanding of such trade-offs is that they reflect the division of finite resources between competing functions (*i.e.*, traits 3 and 4 in Fig. 1) and have their origins in negative genetic covariances (Reznick, 1992). For example, in *Drosophila melanogaster*, selection for early development time produced a correlated response in lower adult mass (Nunney, 1996). In mice, Thomson *et al.* (2002) found that selection for activity early in life caused an eventual decrease in liver function.

Yet many experiments designed to identify such constraints have not found them, suggesting that AS is not an infallible tool. In fact, other factors can make existing trade-offs and genetic constraints hard to find. Genetic variation in resource acquisition can easily obscure a trade-off between two traits (van Noordwijk and de Jong, 1986; Houle, 1991). Even when there is no variation in resource acquisition, variation across multiple allocation pathways can easily obscure trade-offs depending on the relative levels of genetic variation at each step (Fig. 1). In a simulation study, Worley *et al.* (2003) examined a simple pathway involving three traits where there were two allocation ''deci-

sions'' (Fig. 1). When the levels of genetic variation at the earlier allocation were greater than those later in the pathway, artificial selection could not detect a trade-off between traits 3 and 4. Trade-offs were only detectable when the levels of variation in early and late allocations were similar or when there was greater variation in the later allocation. Note that while selection experiments do not invariably detect trade-offs, they are far more likely to find trade-offs than are breeding studies because they provide data on the genetic covariance over multiple generations. In a system like the one described by Worley *et al.* (2003), there is a large parameter space where a positive genetic covariance or no genetic covariance between traits later in the hierarchy would be indicated at generation 0, but a negative genetic covariance would emerge over multiple generations as genes affecting earlier stages become fixed.

Understanding the source of genetic covariances is critically important. Life-history theory assumes that constraining covariances are due to strong pleiotropy (Roff, 1992; Stearns, 1992). If this is not the case and covariances are based on linkage disequilibrium, then two traits can become evolutionarily independent because they are the products of different loci. One can formally test whether genetic correlations are the result of LDE or pleiotropy by suspending selection and randomly mating individuals within a population (or line). If genetic correlations are due to LDE, then they should quickly decrease in magnitude as the genes become randomly assorted across individuals. Conner (2002) performed just such an experiment in wild radish (*Raphanus raphanistrum*). After nine generations of random mating and relaxed selection, there was no evidence for a reduction in the genetic correlations, providing strong evidence for pleiotropy. Davies (1971) performed a similar experiment in *Drosophila melanogaster* where he used balancer stocks to investigate the extent to which recombination eroded the correlation between sternopleural and abdominal bristle numbers. In this case, there was little evidence for pleiotropy and a large role of LDE.

The role of pleiotropy as a genetic constraint is often misstated. In a constant environment, pleiotropy will only be an ineluctable constraint on multivariate evolution in the long run when genetic correlations are $+1$ or -1 (Via and Lande, 1985). If the absolute value of a genetic correlation is less than 1, selection will *eventually* bring trait means to their respective individual optima, although it may require an extremely long time. If selection varies temporally, strong genetic correlations can act as constraints even if their absolute values do not equal 1 (Houle, 2001) because populations do not attain equilibrium values and constantly lag behind in their response to a changing environment. When this happens, genetic covariances influence the directions of multivariate evolution and their effects can be lasting ones (Schluter, 1996).

Antagonistic artificial selection is a strong method for examining the constraining power of genetic covariances (Brakefield, 2003). In this design, the experimenter selects for all combinations of two traits, as done by Beldade et al. (2002) for anterior forewing eyespot size (small and large) and posterior forewing eyespot size (small and large) in the butterfly, *Bicyclus anynana* (Fig. 2). There is a positive phenotypic correlation between their size exhibited among populations of this species and among species within this group of butterflies. The question of interest was whether the lack of populations or species with small anterior and large posterior eyespots (or vice-versa) represented the result of an ineluctable constraint or the fact that selection favored positively correlated eyespots. The critical test is whether the response to selection in one direction is contingent on selection on the other trait. Beldade *et al.* (2002) found significant responses in all four directions, although greater responses occurred in the direction of natural variation. The results indicate that the genetic correlation between the traits was not acting as an inescapable constraint; either there were rare genes with pleiotropic effects that ran counter to the original axis of variation or there was substantial independent genetic variation for each trait that could have contributed to the original correlation via LDE.

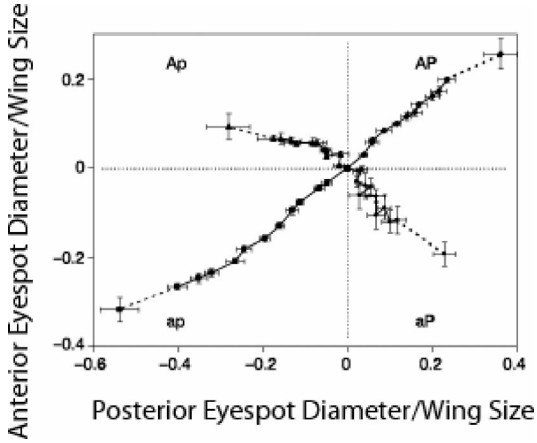

FIG. 2. Results from Beldade *et al.*, (2002) reprinted with permission from Nature. Results from antagonistic selection for increased anterior and posterior (AP), increased anterior/decreased posterior (Ap), decreased anterior and posterior (ap), and decreased anterior/increased posterior (aP) eyespot size. Means and standard errors are shown.

Similar results have been found for forewing area and pupal mass in *Bicyclus anynana* (Frankino *et al.*, 2005) and for abdominal and sternopleural bristle numbers in *Drosophila melanogaster* (Davies and Workman, 1971) where antagonistic selection resulted in responses both with and against the major axis of variation. However, other antagonistic selection experiments indicate that pleiotropy can easily limit responses. For example, antagonistic selection on larval weight and pupal weight in *Tribolium castenata* resulted in limited responses to selection against the natural direction of covariation (Bell and Burris, 1973;

see Roff, 1997 for a further discussion of antagonistic selection and pleiotropy).

Correlated traits and mating systems

Genetic covariances play important roles in three of the five major theories for the evolution of female mating preferences for costly male traits (Fisherian Runaway, good-genes, direct benefits, sexual conflict, and sensory bias; see Mead and Arnold, 2004). The direct benefits and sexual conflict models involve direct sexual selection on female preference to increase the immediate fitness of females (in the case of direct benefits) or to decrease the cost of mating with males (sexual conflict) and neither model requires female preference to covary with any trait other than fitness.

The other three models (Fisher, good-genes, and sensory bias) predict the evolution of preference as a correlated response to selection on another trait; the models diverge in the identity of the ''other'' trait. In the Fisher process, female preference becomes correlated with male ornaments and both evolve to ever-higher levels in a self-reinforcing cycle (Lande, 1981; Kirkpatrick, 1982). In the case of the good-genes mechanism, preference alleles become associated with alleles conferring high viability and preference increases as a by-product of increasing viability (Pomiankowski, 1988). In the case of sensory bias, preference is correlated with other behaviors that share the same sensory system. Hence, natural selection on sensory system properties is driven by selection on other non-mating behaviors and results in correlated changes in female preference (Fuller *et al.*, submitted).

Artificial selection experiments have been used mainly to test the assumption of the Fisher process—that being a genetic correlation between female preference and male ornamentation that arises through LDE (Bakker and Pomiankoski, 1995). In most of these experiments, males have been selected for increased (and decreased) levels of ornamentation, and females have been measured in subsequent generations to ascertain whether or not female preferences have diverged in a similar direction (Houde, 1994; Wilkinson and Reillo, 1994; Brooks and Couldridge, 1998). The problem with these experiments is that if the genetic correlation between preference and ornamentation is due to LDE (as the Fisher model assumes), then we should observe the largest correlated response in the first generation with decreases in subsequent generations (Pomiankowski and Sheridan, 1994). In fact, Houde (1994) observed this general pattern where she found correlated responses in female mating preferences to selection for increased (or decreased) orange coloration in male guppies (*Poecilia reticulata*) after the first generation of selection, but decreased responses in subsequent generations. However, Wilkinson and Reillo (1994) were able to detect strong correlated responses in female mating preference after 13 generations of selection for increased (and decreased) male eye-span in stalk-eyed flies (*Cyrtodipsis dalmanni*). This raises the possibility that either (1) pleiotropy contributes to the correlation between preference and eye-span or (2) sexual selection was occurring *within* each of the lines (Butlin, 1993; Gray and Cade, 1999).

Artificial selection IS a good method to test the assumptions of the sensory bias model, but has not been used as such. Sensory bias assumes that the same underlying neurological pathways that affect non-mating behaviors (*i.e.,* foraging, avoiding predators, finding proper habitat, etc.) also affect female mating preferences; therefore, selection on non-mating behaviors results, via pleiotropy, in correlated responses in female mating preferences (Rodd *et al.,* 2002; Smith *et al.,* 2004). Antagonistic selection experiments as we described them earlier offer a direct means to test this idea.

Analyzing the genetic control of trait variation

We can extract considerable information about the nature of the genetic differences that accumulate between lines selected for opposite values of a complex trait. We can estimate the number of genes at which the lines have diverged in allele frequencies and the distribution of allelic effects among those genes. With some effort, we can examine the extent to which lines diverge through expression of different genes (as opposed to diverging in allele frequencies at the same genes) and evaluate the form and magnitude of any interactions among alleles (dominance) and loci (epistasis) for which the lines differ.

While this information may seem esoteric, it is fundamental for taking our understanding of adaptive evolution beyond the phenomenological. For example, traits controlled by more genes present larger targets for mutation. A larger target can ensure a larger reservoir of genetic variation that, in turn, enhances the ability of a population to respond to novel selection pressures (Lynch and Walsh, 1998). A trait governed by more genes is also capable of more extensive directional evolution before becoming limited by available genetic variation and also capable of generating greater correlated responses through pleiotropy (Falconer and Mackay, 1996; Lynch and Walsh, 1998). The distribution of allelic effects on the phenotype, that is, whether individual genes have large or small effects, can influence the direction, rate, and extent of response to selection (Mackay, 2001*b*).

Other effects, such as whether fitness differences accumulate through diverging allele frequencies or through different loci underlying fitness in different environments, can have more profound consequences. For example, recent work has shown how local adaptation can arise and persist not only by trade-offs between alleles in different environments but also by having different genes contributing to fitness in different environments (Fry, 1996; Orr, 1996; Kawecki, 1997; see Verhoeven *et al.,* 2004 for an example). Understanding whether this genetic detail underlies adaptive differentiation might resolve the paradox of why so much local adaptation occurs without reproductive isolation (see Kawecki, 1997).

Epistatic effects can play a similarly nuanced but profound role, although the assessment of their magnitude and importance remains controversial (Mackay, 2001*b*; Orr, 2001). Epistatic effects can cause selection for identical trait values in different populations to drive different alleles to fixation, creating genetic divergence among populations and potentially creating genetic incompatibilities that can lead to reproductive isolation (Whitlock *et al.*, 1995; Fenster and Galloway, 2000; Wade, 2002). In contrast, if epistasis is negligible, then selection should favor the same set of alleles regardless of genetic background (Ungerer *et al.*, 2003) and populations facing similar selection regimes should be able to exchange migrants freely.

A powerful way to examine the genetic control of trait variation is to use quantitative trait locus (QTL) analysis, which examines, statistically, the phenotypic effects associated with genetic regions that are delimited by molecular markers (*i.e.*, microsatellite loci, RAPD markers, transposons, etc.). The methods of QTL analysis are reviewed elsewhere (Falconer and Mackay, 1996; Mackay, 2001*a*, *b*; Lynch and Walsh, 1998). Divergently selected lines are frequently used to create marker populations with high amounts of genetic variation for QTL analysis. QTL maps are highly effective tools for analyzing complex traits because they allow us to (1) determine the number of loci contributing to variation in a trait and the distribution of those allelic effects, (2) determine whether pleiotropy or linkage disequilibrium accounts for correlated traits, and (3) examine the effects of epistasis and dominance on trait expression.

The insights that QTL analyses offer are amply illustrated in a series of analyses of *D. melanogaster* lines selected upward and downward for sternopleural bristle number and the analyses of the correlated divergence in abdominal bristle number (Long *et al.*, 1995; Gurganus *et al.*, 1998; Nuzhdin *et al.*, 1999; Dilda and Mackay, 2002). For example, different genes contributed to the response to upward and downward selection and more genes of large effect were found in the correlated response for abdominal bristle number than for the direct response in sternopleural bristle number. Lines selected upward were characterized by looser linkage among genes influencing trait variation than lines selected downward. While many epistatic effects were detected, epistatic effects among distant genes were more often found in abdominal bristle number, which is the correlated response, than in sternopleural bristle number, which was the object of selection, and those were more likely manifested in the lines selected for lower bristle numbers.

These results are similar to those from genetic analyses of life history traits (*e.g.*, Leips and Mackay, 2000) and raise some provocative issues. They suggest that genetic responses to divergent selection do not follow the simple paradigm of diverging frequencies of alleles at the same set of genes (see also Verhoeven *et al.*, 2004). They also suggest considerable potential for idiosyncrasy in total response, idiosyncrasy created by epistatic effects and some unpredictability in the magnitude of correlated responses. In that light, the consistency with which adaptive trait clusters emerge in nature is all the more striking.

PERFORMING USEFUL SELECTION EXPERIMENTS

Selection experiments are difficult and risky. To reconcile that statement with our assertions that selection experiments are useful and powerful, we are obliged to offer some guidance on how to maximize usefulness and power.

In contrast to the extensive theory on experimental design for artificial selection, there is little theory specific to the design of laboratory natural selection experiments. However, basic selection theory applies in both arenas (Kimura, 1957; Robertson, 1960, 1961; Hill, 1982). Here we concentrate on two issues, the effects that follow from different intensities of selection and the role of mutations during the experiment. We will review the theoretical results on these issues and then describe what they imply for the design, implementation, and interpretation of selection experiments.

The strength of selection

Of the many factors that govern the outcome of selection experiments, those under the most direct control of the investigator are selection itself, population size, and evolutionary time. Of particular importance is the ''strength'' of selection. Strength of selection can be quantified in terms of the selection differential, S (the difference in average trait values between selected and unselected individuals), and the intensity of selection, $i = S/\sigma_P$, where σ_P is the phenotypic standard deviation. The short-term evolutionary response to selection, R, is provided by the breeder's equation $R = h^2 i \sigma_P$ (Falconer and Mackay, 1996). The long-term evolutionary response in finite populations depends primarily on the joint quantity $N_e i$, where N_e is the effective population size (Robertson, 1960) and secondarily on the magnitude and dominance of allelic effects (Kojima, 1961) such that the magnitude of the selective response is an increasing function of $N_e i$. The analog of $N_e i$ in terms of fitness is $N_e s$ where s is the selection coefficient (Kimura, 1957).

Strength of selection can determine the magnitude of response. For a locus with two alleles with additive effects on fitness, the probability of fixation of the beneficial allele, $u_1(q) = (1 - e^{-2Nesq})/(1 - e^{-2Nes})$ where N_e and s are as above and q is the frequency of the beneficial allele (Kimura, 1957; when $N_e s = 0$, $u_1[q] = q$, as expected for a neutral allele). All else equal, an allele with a given beneficial effect on the trait under selection will be more likely to fix in a large population and/or if it is common in the base population. In a sufficiently large population, at a sufficiently high starting frequency, a beneficial allele is guaranteed to reach fixation. If we imagine many equivalent loci that potentially affect a trait, the total possible advance in fitness is $1 - q$; over 70% of the potential advance

will be achieved when $N_e sq > 1$ and over 90% when $N_e sq > 2$. If an infinitesimal model is assumed (infinitely many loci each with infinitesimal effects), the greatest selective advance will be achieved if 50% of the population is allowed to reproduce (Dempster, 1955; Robertson, 1960).

These arguments can be extended to quantitative traits with a surprising result. Robertson (1960), following Haldane (1931), extended Kimura's theory to quantitative traits by relating the selection coefficient to the intensity of selection such that $s = ia/\sigma_P$, where a is the additive effect of an allele. In the case of artificial selection, the probability of fixation can be written $u_1(q) = f(N_e ia, q)$ and the expected limit of selection is a function of $N_e i$. The crucial result is that the exact form of the function will depend on the distribution of allele frequencies and of the type of gene action (additive or dominant) (Kojima, 1961). If $N_e i$ is small, only alleles with large effects on the phenotype and/or those segregating at high frequencies will be likely to fix. Thus, the genetic architecture of selected traits that will be uncovered in genetic analyses of selected lines depends fundamentally on the nature of selection that produced the lines. There is an important additional consideration about ''effective size'' itself that should influence the design of experiments. Obviously, for a given population size, increasing the intensity of selection necessitates reducing the fraction of individuals that contribute to the next generation, which in itself reduces N_e. More subtly, selection causes a reduction in the effective population size relative to an unselected population with the same number of individuals (Robertson, 1961). The degree to which selection reduces N_e depends on the heritability and the degree of dominance, but the effect can be substantial (see Walsh and Lynch, http://nitro.biosci.arizona.edu/zbook/volume_2/vol2.html, Example 17.2). The reduction in N_e increases with the strength of selection, and this effect accumulates in strength over generations of selection. An increase in $N_e i$ is more surely achieved by increasing the number of individuals than the intensity of selection; more surprisingly, a sufficiently large increase in i can actually reduce $N_e i$ via the resultant disproportionate decrease in N_e.

The reduction in N_e from selection has several implications for laboratory natural selection. Control populations with the same number of individuals as selected populations will be less likely to fix deleterious alleles. If deleterious mutations have consistent effects on the phenotype (*e.g.,* they reduce body size or increase time to maturation), one result of the experiment could be to erroneously infer an average pleiotropic effect of the alleles underlying the trait(s) responsible for adaptation to the selective environment. This problem is compounded by interference between selected loci (the ''Hill-Robertson effect,'' Hill and Robertson, 1966), whereby selection at one locus reduces the efficiency of selection at linked loci. Also, if selection is relatively weak (*i.e.,* if $4N_e s$ is small), replicate selected lines are *more* likely to diverge via drift than are control lines (the ''Cohan effect''; Cohan, 1984), a result that could potentially be misinterpreted as idiosyncratic adaptive divergence among selected lines. In other words, in an LNS experiment, a pattern of drift might be misinterpreted as selection favoring different combinations of traits in different lines as responses to the same selective environment. These problems are not easily overcome, and they suggest that investigators employing laboratory natural selection ought to estimate N_e in both control and selected populations.

While the implications of these considerations are obvious for AS experiments, they also loom large—but in subtle ways—for designing the treatments used in LNS. While the actual selection intensity in these experiments will be unknown, the experiments are usually designed with the implicit assumption that the greater the deviation between the selecting environment and the natural environment from which the stock population is drawn, the greater will be the intensity of selection. For example, intuition leads us to expect that fewer individuals will contribute genes to the next generation when the selecting environment is five standard deviations colder than the mean annual temperature of the control environment than when it is one standard deviation colder because only the most cold-tolerant phenotypes will be able to survive. This intuition is derived from the considerable evidence that the effects of deleterious alleles are magnified in ''stressful'' environments (Jimenez *et al.,* 1994; Ritland, 1996; Shabalina *et al.,* 1997), which suggests that a smaller fraction of the population would contribute to the next generation in an extreme environment than in a moderate one. The design decision that might follow from this intuition is to make the selection scheme efficient by making the environmental differences as large as possible.

This design may not be the best choice. Many LNS experiments impose a selection regime that is *immediately* novel on a stock population that was well-adapted to the control conditions (*e.g.,* Mueller and Ayala, 1981; Huey *et al.,* 1991; Lenski *et al.,* 1991). These experiments *immediately* change the mean environment experienced by the treated population to the extremes of those found in nature. Sudden, drastic changes in the environment certainly occur in nature, and selection may often act in that way (*e.g.,* trematodes that divide their time between aquatic habitats and the guts of endothermic vertebrates). However, in many cases the environmental factors responsible for phenotypic evolution may change gradually over generations. An experiment designed to impose an environment that changes gradually (*presumably* weak selection) will take longer than one in which the selection regime is immediately altered to the desired environmental endpoint (strong selection), so there is an obvious logistical advantage to the latter design. But recall that different strengths of selection create important differences in the evolutionary process; in particular, strong selection will bias the outcome toward

fixation of alleles of large effect. Since most selection experiments employ population sizes much smaller than those observed in nature, the process of adaptation in selection experiments may be biased toward outcomes governed by alleles of large effect and strong pleiotropy, which need not reflect how the process unfolds in nature.

The role of mutations during the experiment

Evolutionary biologists are usually interested in the long-term response to selection, which is usefully considered in terms of fixation probabilities of alleles underlying the trait(s) under selection. Over the short term (a few generations), the response to selection will primarily be a function of the alleles segregating in the base population. In the medium term (a few tens of generations), both pre-existing variants and new mutations will contribute to the evolutionary response. In the long term, the asymptotic response to selection will depend on the input of new alleles by mutation.

There is substantial empirical evidence that new mutations are important in selection experiments, even on fairly short time scales (Mackay *et al.*, 1994; Barton and Keightley, 2002). In most cases, selection experiments are started with outbred stocks, but there are exceptions (*e.g.*, Lenski *et al.*, 1991). Whether to begin a selection experiment with (replicate) highly inbred lines, in which case the response to selection will be entirely due to the effects of newly arising mutations, or with a large outbred stock will depend on several considerations. First, except for organisms whose generation time is very fast, it will take a long time for mutation to introduce sufficient variation to detect a response. Second, the nature of genetic variation produced by new mutation may be qualitatively different from standing variation that has been subjected to the selective sieve in nature. Whether evolution usually proceeds via selection on standing variation or on new mutations is an open question. It is well-established that new mutations are deleterious on average (Drake *et al.*, 1998). Whether the standing genetic variance can be explained by mutation-selection balance is controversial (Houle *et al.*, 1996), but there is evidence that at least some standing variation is maintained by some form of balancing selection (Charlesworth and Hughes, 1999). Third, the mutational covariance may differ substantially from the standing genetic covariance. Mutational covariances are typically large and positive, which suggests that the average new mutation has multiple deleterious pleiotropic effects (Houle *et al.*, 1994).

Putting theory into practice: methodological issues

There is a vast literature on the methodology of artificial selection, and we will not attempt even a cursory summary. The reader is referred in particular to Bruce Walsh's excellent web site where the unpublished manuscript of Walsh and Lynch, Volume 2 *Evolution and Selection of Quantitative Traits* is available in pdf format; the URL is http://nitro.biosci.arizona.edu/zbook/volume_2/vol2.html. We will touch on some issues that may be underappreciated.

Replication and power. While the importance of replication in selection studies is well-appreciated (*e.g.*, Hill and Caballero, 1992; Rose *et al.*, 1996), there will always be constraints on the number of organisms that can be included in any experiment. There will usually be a tradeoff between the amount of replication and the size of the selected population, and a further tradeoff between selection intensity and population size. For short-term experiments, the statistical power will be an increasing function of the heritability of the trait(s) in question and of the intensity of selection (Walsh and Lynch in preparation, chapter 6). In long-term experiments, the power will depend on the mutational heritability. Similar considerations hold for the estimation of correlated responses. Replication is obviously desirable, but individual replicates should be large enough to have reasonable power. The standard error of the expected response to selection is a function only of the number of parents and it does not matter if they are subdivided in replicate lines (Hill, 1971). However, since fixation probabilities of alleles with different properties of dominance and magnitude of effect depend on population size, two small experiments are not equivalent to one large experiment. In certain cases, particularly if the goal is to generate divergent lines for further genetic dissection (say in QTL studies), it may be preferable to do one large but unreplicated experiment rather than two smaller but less powerful experiments in order to more fully capture the relevant genetic variation. Standard power theory can often be applied to artificial selection experiments (Hill, 1971; Lynch and Walsh, 1998 Appendix 5), but in laboratory natural selection, power analysis may require simulation studies (*e.g.*, Baer *et al.*, 2000).

What treatment(s) to use? Selection can usually be categorized into selection for increased or decreased trait value, which we will call ''up'' and ''down'' selection. In a perfect world a selection experiment would have multiple up and down selected lines and multiple unselected control lines. The size of the ideal experiment may be impractical, and a carefully chosen subset of treatment groups may provide a more efficient approach to estimate quantitative genetic parameters. Classical selection theory predicts that if a symmetric response to selection is expected, divergent selection (up and down lines) is more efficient than maintaining the same number of individuals in up, down, and unselected control groups (Hill, 1971). If a symmetric response is not expected, then quantitative genetic parameters (*e.g.*, realized heritability) estimated from the cumulative divergence of up and down selected lines will be biased, and unbiased estimates will require comparison to unselected control lines. Asymmetric responses to selection can arise for a variety of reasons (different allelic effects and/or frequencies; different pleiotropic effects on fitness, different loci underlying trait increase *vs.* decrease), and

in fact may be the rule rather than the exception (Falconer and Mackay, 1996).

A more subtle issue is: What are the appropriate up and down treatments? The answer will often be clear—large and small size at maturity, for example. In other cases the answer will not be obvious. For example, consider the evolution of drought tolerance in a plant. There are several ways to approach the problem. One could mimic a drought regime and see what traits evolved. Or, one could employ laboratory culling, by applying desiccation and picking the last $\times$ % of the survivors. Or, one could select on a trait known to be associated with drought tolerance, such as some measure of water balance. In each case selection is ''up.'' However, the appropriate ''down'' treatment is not so clear. An obvious candidate would be some measure of drought-sensitivity (which may require family-level or index selection; see below). Depending on the details of the genetic architecture, the evolution of drought tolerance may involve fixing rare beneficial alleles that increase tolerance over the wild-type or purging alleles that are neutral in a non-drought environment but are deleterious in drought conditions. In a different context, we might imagine that flood-tolerance may be an appropriate ''down'' treatment, if the question of interest involved potential trade-offs in fitness or performance in different environments.

What is the appropriate control? The control is typically a line or lines maintained under the same environmental and demographic conditions as the selected group, except for the selective factor per se. As noted above, populations under selection will have a smaller N_e/N ratio than unselected populations and will diverge more via drift than will unselected populations (but recall the ''Cohan effect''). In principle, it would be possible to design a control population with approximately the same N_e as the selected population, although this is obviously difficult in laboratory natural selection.

It is often possible to maintain the ancestral stock population in an inert state (*e.g.,* as seeds or in cryostorage) and then revive it to compare to the evolved descendent. The advantage of maintaining the control in this way is that it greatly reduces the necessary resources, thereby allowing larger selected populations and/or more replicate selected lines. However, directional changes in selected lines unrelated to selection itself (*e.g.,* inbreeding depression, accumulation of deleterious mutations) would be unable to be separated from the actual response to selection.

What is the appropriate base population? The short to medium-term response to selection will depend on the properties of the genetic variation segregating in the base population. Considerable theory exists concerning the properties of various base populations from which to begin artificial selection (Lush, 1947*a*, *b*; Robertson, 1960, 1961, 1966; Madalena and Hill, 1972; Hill, 1982), especially with regard to the ultimate selective advance and the speed with which a particular advance is achieved. For example, inbreeding prior to selection increases the frequency of homozygotes, thereby speeding the response to selection if recessive alleles are important (Robertson, 1961). An obvious strategy to maximize the amount of genetic variation in the base population is by pooling stocks, *e.g.,* from multiple populations from across a species' range, thereby converting among-population genetic variance to within-population variance. However, evolutionary biologists are often interested in the relationship between divergence among populations and variation within populations (*e.g.,* Lande, 1979; Schluter, 1996; Begin and Roff, 2001). If this is the question of interest, reconstituting the among-population (co)variance could lead to a misleadingly circular conclusion, because the alleles available for selection would be those currently responsible for the observed differences.

What is the appropriate method of artificial selection? There are situations in which the usual methods of artificial selection, truncation or graduated selection among individuals, are not the best choices. If individual selection is impossible, multiple family members can be measured and individuals from the best families can be saved to propagate the next generation. Family selection can be performed within families (*e.g.,* the best $\times$ % of each family is chosen) or among families. If heritability is low, family selection is in fact more efficient (Lush, 1947*b*; Falconer and Mackay, 1996). In any case, within-family selection is expected to produce a greater ultimate selective advance than individual selection (Dempfle, 1974). The obvious limitation is the requirement of knowing parentage. If pedigree information is available, powerful mixed-model methods can be used to estimate quantitative genetic parameters (Henderson, 1984; Sørensen and Kennedy, 1984).

Finally, no treatment of selection would be complete without mention of group selection, wherein selection acts among populations. Group selection has a long and controversial history in evolutionary biology, beginning with the theoretical work of Sewell Wright (1931). For our purposes, we point out that there are circumstances in which selection among groups for a particular trait will produce an evolutionary response when selection among individuals will not. If the genetic variation for a trait is due to dominance, or dominance by (additive, dominance) epistasis, the trait will not respond to individual selection (except in an inbred population; Robertson, 1952), but will respond to selection among groups. Genetic variation resulting from additive by additive epistasis will initially respond to selection, but the response will diminish as recombination breaks up linkage disequilibrium (Griffing, 1960). Further, genetically-based interactions between individuals will not respond to individual selection but will respond to group selection (Griffing, 1977, 1981*a*, *b*). The theory of artificial group selection has been reviewed lucidly by Goodnight and Stevens (1997), and empirical studies have confirmed that group selec-

tion can be effective in situations when individual selection is not (Craig, 1982; Goodnight, 1985, 1990).

SUMMARY

Selection experiments—LNS and AS—provide an irreplaceable tool for the study of evolution at the within-species level. LNS allows the investigator to vary the environmental context in a controlled way; evolutionary responses must thus necessarily have been in response *to that environmental factor(s)* per se, although the phenotypic target of selection cannot be determined unambiguously. Conversely, AS allows the investigator to directly control selection on a particular trait (or traits). If traits other than the trait under selection evolve, they must necessarily be genetically correlated with the trait in question, either by pleiotropy or by linkage disequilibrium. We have attempted to show the range of important questions for which these methods are especially useful and highlight some issues in designing and interpreting these experiments that are perhaps underappreciated. In particular, we think it imperative to call attention to the fact that differing selection regimes (in terms of N_ei and/or N_es) may produce similar phenotypic responses to selection through qualitatively different underlying genetic architectures. Given that many selection experiments are implemented to uncover the genetic architecture of complex traits, we advise careful consideration of this fact in the design of these experiments, lest investigators obtain just what they paid for and not what they sought.

REFERENCES

Archer, M. A., J. P. Phelan, K. A. Beckman, and M. R. Rose. 2003. Breakdown in correlations during laboratory evolution. II. Selection on stress resistance in *Drosophila* populations. Evolution 57:536–543.

Azevedo, R. B. R., A. C. James, J. McCabe, and L. Partridge. 1998. Latitudinal variation of wing: Thorax size ratio and wing-aspect ratio in *Drosophila melanogaster*. Evolution 52:1353–1362.

Baer, C. F. and M. Lynch. 2003. Correlated evolution of life-history with size at maturity in *Daphnia pulicaria*: Patterns within and between populations. Genet. Res. 81:123–132.

Baer, C. F. and J. Travis. 2000. Direct and correlated responses to artificial selection on acute thermal stress tolerance in a livebearing fish. Evolution 54:238–244.

Baer, C. F., J. Travis, and K. Higgins. 2000. Experimental evolution in *Heterandria formosa*, a livebearing fish: Group selection on population size. Genet. Res. 76:169–178.

Bakker, T. C. M. and A. Pomiankowski. 1995. The genetic-basis of female mate preferences. J. Evol. Biol. 8:129–171.

Barton N. H. and P. D. Keightley. 2002. Understanding quantitative genetic variation. Nat. Rev. Genet. 3:11–21.

Begin, M. and D. A. Roff. 2001. An analysis of G matrix variation in two closely related cricket species, *Gryllus firmus* and *G. pennsylvanicus*. J. Evol. Biol. 14:1–13.

Beldade, P., K. Koops, and P. M. Brakefield. 2002. Developmental constraints versus flexibility in morphological evolution. Nature 416:844–847.

Bell, A. E. and M. J. Burris. 1973. Simultaneous selection for 2 correlated traits in *Tribolium*. Genet. Res. 21:29–46.

Bennett, A. F. and R. E. Lenski. 1993. Evolutionary adaptation to temperature. 2. Thermalniches of experimental lines of *Escherichia coli*. Evolution 47:1–12.

Berry, A. and M. Kreitman. 1993. Molecular analysis of an allozyme cline-alcohol dehydrogenase in *Drosophila melanogaster* on the east cast of North America. Genetics 134:869–893.

Brakefield, P. M. 2003. Artificial selection and the development of ecologically relevant phenotypes. Ecology 84:1661–1671.

Brooks, R. and V. Couldridge. 1999. Multiple sexual ornaments coevolve with multiple mating preferences. Am. Nat. 154:37–45.

Butlin, R. K. 1993. A comment on the rvidence for a genetic correlation between the sexes in *Drosophila melanogaster*. Anim. Behav. 45:403–404.

Charlesworth, B. and K. A. Hughes. 1999. The maintenance of genetic variation in life history traits. *In* R. S. Singh and C. B. Krimbas (eds.), *Evolutionary genetics from molecules to morphology*. Cambridge University Press, Cambridge, UK.

Cohan, F. M. 1984. Can uniform selection retard random genetic divergence between isolated conspecific populations? Evolution 38:495–504.

Conner, J. K. 2002. Genetic mechanisms of floral trait correlations in a natural population. Nature 420:407–410.

Conner, J. K. 2003. Artificial selection: A powerful tool for ecologists. Ecology 84:1650–1660.

Cullum, A. J., A. F. Bennett, and R. E. Lenski. 2001. Evolutionary adaptation to temperature. IX. Preadaptation to novel stressful environments of *Escherichia coli* adapted to high temperature. Evolution 55:2194–2202.

Conover, D. O. and S. B. Munch. 2002. Sustaining fisheries yields over evolutionary time scales. Science 297:94–96.

Craig, D. M. 1982. Group selection versus individual selection: An experimental analysis. Evolution 36:271–282.

Davies, R. W. 1971. Genetic relationship of 2 quantitative characters in *Drosophila melanogaster*. 2. Location of effects. Genetics 69:363–375.

Davies, R. W. and P. L. Workman. 1971. Genetic relationship of 2 quantitative characters in *Drosophila melanogaster*. 1. Responses to selection and whole chromosome analysis. Genetics 69: 353–361.

deMeester, L. 1996. Evolutionary potential and local genetic differentiation in a phenotypically plastic trait of a cyclical parthenogen, *Daphnia magna*. Evolution 50:1293–1298.

Dempfle, L. 1974. Note on increasing limit of selection through selection within families. Genet. Res. 24:127–135.

Dempster, E. R. 1955. Genetic models in relation to animal breeding. Biometrics 11:535–536.

Desmarais, K. H. and A. J. Tessier. 1999. Performance trade-off across a natural resource gradient. Oecologia 120:137–146.

Dilda, C. L. and T. F. C. Mackay. 2002. The genetic architecture of *Drosophila* sensory bristle number. Genetics 162:1655–1674.

Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow. 1998. Rates of spontaneous mutation. Genetics 148:1667–1686.

Elena, S. F. and R. E. Lenski. 1997. Long-term experimental evolution in *Escherichia coli*. 7. Mechanisms maintaining genetic variability within populations. Evolution 51:1058–1067.

Falconer, D. S. 1992. Early selection experiments. Annual Review of Genetics 26:1–14.

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

Feldgarden, M., D. M. Stoebel, D. Brisson, and D. E. Dykhuizen. 2003. Size doesn't matter: Microbial selection experiments address ecological phenomena. Ecology 84:1679–1687.

Felsenstein, J. 2002. Contrasts for a within-species comparative method. *In* M. Slatkin and M. Veuille (eds.), *Modern developments in theoretical population genetics*, pp. 118–129. Oxford University Press, Oxford.

Fenster, C. B. and L. F. Galloway. 2000. Population differentiation in an annual legume: Genetic architecture. Evolution 54:1157–1172.

Frankino, W. A., B. J. Zwaan, D. L. Stern, and P. M. Brakefield. 2005. Natural selection and developmental constraints in the evolution of allometries. Science 307:718–720.

Fry, J. D. 1996. The evolution of host specialization: Are trade-offs overrated? Am. Nat. 148:S84–S107.

Fuller, R. C. and J. Travis. 2004. Genetics, lighting environment, and heritable responses to lighting environment affect male col-

or morph expression in bluefin killifish, *Lucania goodei*. Evolution 58:1086–1098.
Garland, T. 2003. Selection experiments: An underutilized tool in biomechanics and organismal biology. *In* V. L. Bels, J.-P. Gasc, and A. Casinos (eds.), *Vertebrate biomechanics and evolution*, pp. 23–56. Scientific Publishers Ltd, Oxford.
Goodnight, C. J. 1985. The influence of environmental variation on group and individual selection in a cress. Evolution 39:545–558.
Goodnight, C. J. 1990. The response to selection at the community level. Evolution 44:1614–1624.
Goodnight, C. J. and L. Stevens. 1997. Experimental studies of group selection: What do they tell us about group selection in nature? Am. Nat. 150:S59–S79.
Gray, D. A. and W. H. Cade. 1999. Correlated-response-to-selection experiments designed to test for a genetic correlation between female preferences and male traits yield biased results. Anim. Behav. 58:1325–1327.
Griffing, B. 1960. Theoretical consequences of truncation selection based on the individual phenotype. Aust. J. Biol. Sci. 13:307–343.
Griffing, B. 1977. Selection for populations of interacting genotypes. *In* E. Pollak, O. Kempthorne, and T. B. Baily (eds.), *Proceedings of the International Congress on Quantitative Genetics*, pp. 413–434. Iowa State University, Ames, Iowa.
Griffing, B. 1981*a*. A theory of natural selection incorporating interaction among individuals. I. The modeling process. J. Theor. Biol. 89:635–658.
Griffing, B. 1981*b*. A theory of natural selection incorporating interaction among individuals. II. Use of random groups of inbred individuals. J. Theor. Biol. 89:679–690.
Gurganus, M. C., J. D. Fry, S. V. Nuzhdin, E. G. Pasyukova, R. F. Lyman, and T. F. C. Mackay. 1998. Genotype-environment interaction at quantitative trait loci affecting sensory bristle number in *Drosophila melanogaster*. Genetics 149:1883–1898.
Haldane, J. B. S. 1931. A mathematical theory of natural and artificial selection. VII. Selection intensity as a function of mortality rate. Proc. Camb. Phil. Soc. 27:131–136.
Hansen, T. F., W. S. Armbruster, and L. Antonsen. 2000. Comparative analysis of character displacement and spatial adaptations as illustrated by the evolution of *Dalechampia* blossoms. Am. Nat. 156:S17–S34.
Henderson, C. R. 1984. *Applications of linear models in animal breeding*. University of Guelph, Guelph, Ontario.
Hill, W. G. 1971. Design and efficiency of selection experiments for estimating genetic parameters. Biometrics 27:293–311.
Hill, W. G. 1982. Predictions of response to artificial selection from new mutations. Genet. Res. 40:255–278.
Hill, W. G. and A. Caballero. 1992. Artificial selection experiments. Ann. Rev. Ecol. Syst. 23:287–310.
Hill, W. G. and A. Robertson. 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8:269–294.
Houde, A. E. 1994. Effect of artificial selection on male color patterns on mating preference of female guppies. Proc. R. Soc. London Ser. B 256:125–130.
Houle, D. 1991. Genetic covariance of fitness correlates—what genetic correlations are made of and why it matters. Evolution 45: 630–648.
Houle, D. 2001. Characters as the units of evolutionary change. *In* G. P. Wagner (ed.), *The character concept in evolutionary biology*, pp. 109–140. Academic Press, San Diego.
Houle, D., K. A. Hughes, D. K. Hoffmaster, J. Ihara, S. Assimacopoulos, D. Canada, and B. Charlesworth. 1994. The effects of spontaneous mutation on quantitative traits. 1. Variances and covariances of life-history traits. Genetics 138:773–785.
Houle, D., B. Morikawa, and M. Lynch. 1996. Comparing mutational variabilities. Genetics 143:1467–1483.
Huey, R. B., L. Partridge, and K. Fowler. 1991. Thermal sensitivity of *Drosophila melanogaster* responds rapidly to laboratory natural selection. Evolution 45:751–756.
James, A. C., R. B. R. Azevedo, and L. Partridge. 1995. Cellular basis and developmental timing in a size cline of *Drosophla melanogster*. Genetics 140:659–666.
James, A. C., R. B. R. Azevedo, and L. Partridge. 1997. Genetic and environmental responses to temperature of *Drosophila melanogaster* from a latitudinal cline. Genetics 146:881–90.
James, A. C. and L. Partridge. 1998. Geographic variation in competitive ability in *Drosophila melanogaster*. Am. Nat. 151:530–537.
Jimenez, J. A., K. A. Hughes, G. Alaks, L. Graham, and R. C. Lacy. 1994. An experimental study of inbreeding depression in a natural habitat. Science 266:271–273.
Joshi, A. and J. N. Thompson. 1995. Trade-offs and the evolution of host specialization. Evol. Ecol. 9:82–92.
Joshi, A. and J. N. Thompson. 1997. Adaptation and specialization in a two-resource environment in *Drosophila* species. Evolution 51:846–855.
Juliano, S. A. and M. E. Gravel. 2002. Predation and the evolution of prey behavior: An experiment with tree hold mosquitoes. Behav. Ecol. 13:301–311.
Kawecki, T. J. 1997. Sympatric speciation via habitat specialization driven by deleterious mutations. Evolution 51:1751–1763.
Kennington, W. J., J. R. Killeen, D. B. Goldstein, and L. Partridge. 2003. Rapid laboratory evolution of adult wing area in *Drosophila melanogaster* in response to humidity. Evolution 57: 932–936.
Kimura, M. 1957. Some problems of stochastic processes in genetics. Ann. Math. Stat. 28:882–901.
Kimura, M. and T. Maruyama. 1971. Pattern of neutral polymorphism in a geographically structured population. Genet. Res. 18: 125–131.
Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. Am. Nat. 157:245–261.
Kirkpatrick, M. 1982. Sexual selection and the evolution of female choice. Evolution 36:1–12.
Kojima, K-i. 1961. Effects of dominance and size of population on response to mass selection. Genet. Res. 2:177–188.
Lande, R. 1979. Quantitative genetic-analysis of multivariate evolution, applied to brain—body size allometry. Evolution 33: 402–416.
Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. Proc. Nat. Acad. Sci. U.S.A. 78:3721–3725.
Lande, R. and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution 37:1210–1226.
Leips, J. and T. F. C. Mackay. 2000. Quantitative trait loci for life span in *Drosophila melanogaster*: Interactions with genetic background and larval density. Genetics 155:1773–1788.
Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2000 generations. Am. Nat. 138: 1315–1341.
Long, A. D., S. L. Mullaney, L. A. Reid, J. D. Fry, C. H. Langley, and T. F. C. Mackay. 1995. High-resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. Genetics 139:1273–1291.
Lush, J. L. 1947*a*. Family merit and individual merit as bases for selection. 1. Am. Nat. 81:241–261.
Lush, J. L. 1947*b*. Family merit and individual merit as bases for selection. 2. Am. Nat. 81:362–379.
Lynch, M. 1980. The evolution of cladoceran life histories. Quart. Rev. Biol. 55:23–42.
Lynch, M. and B. Walsh. 1998. *Genetics and analysis of quantitative traits*. Sinauer Associates, Sunderland, Massachusetts.
Mackay, T. F. C. 2001*a*. Quantitative trait loci in *Drosophila*. Nature Reviews Genetics 2:11–20.
Mackay, T. F. C. 2001*b*. The genetic architecture of quantitative traits. Ann. Rev. Genet. 35:303–339.
Mackay, T. F. C., J. D. Fry, R. F. Lyman, and S. V. Nuzhdin. 1994. Polygenic mutation in *Drosophila melanogaster*: Estimates from response to selection of inbred strains. Genetics 136:937–951.
Madalena, F. E. and W. G. Hill. 1972. Population structure in artificial selection programs—simulation studies. Genet. Res. 20: 75–99.

Mead, L. S. and S. J. Arnold. 2004. Quantitative genetic models of sexual selection. TREE 19:264–271.
Mueller, L. D. 1997. Theoretical and empirical examination of density-dependent selection. Ann. Rev. Ecol. Syst. 28:269–288.
Mueller, L. D. and F. J. Ayala. 1981. Trade-off between r-selection and K-selection in *Drosophila* populations. Proc. Natl. Acad. Sci. U.S.A. 78:1303–1305.
Neat, F., K. Fowler, V. French, and L. Partridge. 1995. Thermal evolution of growth efficiency in *Drosophila melanogaster*. Proc. R. Soc. London Ser. B. 260:73–78.
Nunney, L. 1996. The response to selection for fast larval development in *Drosophila melanogaster* and its effect on adult weight: An example of a fitness trade-off. Evolution 50:1193–1204.
Nuzhdin, S. V., C. L. Dilda, and T. F. C. Mackay. 1999. The genetic architecture of selection response: Inferences from fine-scale mapping of bristle number quantitative trait loci in *Drosophila melanogaster*. Genetics 153:1317–1331.
Orr, H. A. 1996. Dobzhansky, Bateson, and the genetics of speciation. Genetics 144:1331–1335.
Orr, H. A. 1998. Testing natural selection vs. genetic drift in phenotypic selection using quantatitive trait locus data. Genetics 149:2099–2104.
Orr, H. A. 2001. The genetics of species differences. TREE 16:343–350.
Partridge, L., B. Barrie, K. Fowler, and V. French. 1994. Evolution and development of body size and cell size in *Drosophila melanogaster* in response to temperature. Evolution 48:1269–1276.
Partridge, L., B. Barrie, N. H. Barton, K. Fowler, and V. French. 1995. Rapid laboratory evolution of adult life-history traits in *Drosophila melanogaster* in response to temperature. Evolution 49:538–544.
Pitnick, S. and G. T. Miller. 2000. Correlated response in reproductive and life history traits to selection on testis length in *Drosophila hydei*. Heredity 84:416–426.
Pomiankowski, A. 1988. The evolution of female mate preferences for male genetic quality. Oxford Surv. Evol. Biol. 5:136–184.
Pomiankowski, A. and L. Sheridan. 1994. Linked sexiness and choosiness. TREE 9:242–244.
Provine, W. B. 1971. *The origins of theoretical population genetics*. Univ. of Chicago Press, Chicago.
Reznick, D. 1992. Measuring the costs of reproduction. TREE 7: 42–45.
Reznick, D., and J. Travis. 1996. Empirical studies of adaptation. *In* M. Rose and G. Lauder (eds.), *Adaptation: Perspectives and new approaches*, pp. 243–289. Academic Press, New York.
Reznick, D. and J. Travis. 2001. Adaptation. *In* C. W. Fox, D. A. Roff, and D. J. Fairbairn (eds.), *Evolutionary ecology*, pp. 44–57. Oxford Univ. Press, New York.
Reznick, D., N. Mateos, and M. S. Springer. 2002. Independent origins and rapid evolution of the placenta in the fish genus *Poeciliopsis*. Science 298:1018–1020.
Ritland, K. 1996. Inferring the genetic basis of inbreeding depression in plants. Genome 39:1–8.
Robertson, A. 1952. The effect of inbreeding on the variation due to recessive genes. Genetics 37:189–207.
Robertson, A. 1960. A theory of limits in artificial selection. Proc. Roy. Soc. London B 153:234–249.
Robertson, A. 1961. Inbreeding in artificial selection programmes. Genet. Res. 2:189–194.
Robertson, A. 1966. Artificial selection in plants and animals. Proc. Roy. Soc. London B 164:341–349.
Robinson, S. J. W., B. Zwaan, and L. Partridge. 2000. Starvation resistance and adult body composition in a latitudinal cline of *Drosophila melanogaster*. Evolution 54:1819–1824.
Rodd, F. H., K. A. Hughes, G. F. Grether, and C. T. Baril. 2002. A possible non-sexual origin of mate preference: Are male guppies mimicking fruit? Proc. Roy. Soc. London Ser. B 269:475–481.
Roff, D. A. 1992. *The evolution of life histories: Theory and analysis*. Chapman and Hall, New York.
Roff, D. A. 1997. *Evolutionary quantitative genetics*. Chapman and Hall, New York.
Roff, D. A. 2002. *Life history evolution*. Sinauer, Sunderland, Massachusetts.
Rose, M. R., J. L. Graves, and E. W. Hutchinson. 1990. The use of selection to probe patterns of pleiotropy in fitness characters. *In* F. Gilbert (ed.), *Insect life cycles*, pp. 29–42. Springer-Verlag, New York.
Rose, M. R., T. J. Nusbaum, and A. K. Chippindale. 1996. Laboratory evolution: The experimental wonderland and the Cheshire cat syndrome. In M. R. Rose and G. V. Lauder (eds.), *Adaptation*, pp. 221–241. Academic Press, San Diego, California.
Rozen, D. E. and R. E. Lenski. 2000. Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. Am. Nat. 155:24–35.
Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50:1766–1774.
Schwarzkopf, L., M. W. Blows, and M. J. Caley. 1999. Life-history consequences of divergent selection on egg size in *Drosophila melanogaster*. Am. Nat. 154:333–340.
Shabalina, S. A., L. Y. Yampolsky, and A. S. Kondrashov. 1997. Rapid decline of fitness in panmictic populations of *Drosophila melanogaster* maintained under relaxed natural selection. Proc. Natl. Acad. Sci. U.S.A. 94:13034–13039.
Smith, C., L. Barber, R. J. Wootton, and L. Chittka. 2004. A receiver bias in the origin of three-spined stickleback mate choice. Proc. Roy. Soc. London Ser. B 271:949–955.
Sørensen, D. A. and B. W. Kennedy. 1984. Estimation of genetic variances from unselected and selected populations. J. Anim. Sci. 59:1213–1233.
Sniegowski, P. D., P. J. Gerrish, and R. E. Lenski. 1997. Evolution of high mutation rates in experimental populations of *E. coli*. Nature 387:703–705.
Spitze, K. 1991. Chaoborus predation and life-history evolution in *Daphnia pulex*—temporal pattern of population diversity, fitness, and mean life history. Evolution 45:82–92.
Spitze, K., J. Burnson, and M. Lynch. 1991. The covariance structure of life-history characters in *Daphnia pulex*. Evolution 45: 1081–1090.
Stalker, H. D. and H. L. Carson. 1947. Morphological variation in natural populations of *Drosophila robusta* Sturtevant. Evolution 1:237–248.
Stearns, S. C. 1992. *The evolution of life histories*. Oxford University Press, Oxford.
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.
Tantawy, A. O. 1964. Studies on natural populations of *Drosophila*. III. Morphological and genetic differences in wing length in *Drosophila melanogaster* and *D. simulans* in relation to season. Evolution 18:560–570.
Tantawy, A. O. and G. S. Mallah. 1961. Studies on natural populations of *Drosophila*. I. Heat resistance and geographical variation in *Drosophila melanogaster* and *D. simulans*. Evolution 15:1–14.
Tessier, A. J. and P. Woodruff. 2002. Trading off the ability to exploit rich versus poor food quality. Ecol. Lett. 5:685–692.
Thomson, S. L., T. Garland, J. G. Swallow, and P. A. Carter. 2002. Response of Sod-2 enzyme activity to selection for high voluntary wheel running. Heredity 88:52–61.
Travis, J. and D. Reznick. 1998. Experimental approaches to the study of evolution. *In* W. J. Resitarits and J. Bernardo (eds.), *Issues and perspectives in experimental ecology*, pp. 437–459. Oxford University Press, New York.
Travisano, M., J. A. Mongold, A. F. Bennett, and R. E. Lenski. 1995. Experimental tests of the roles of adaptation, chance, and history in evolution. Science 267:87–90.
Trexler, J. C. and J. Travis. 1990. Phenotypic plasticity in the sailfin molly, *Poecilia latipinna* (Pisces, Poeciliidae). 1. Field experiments. Evolution 44:1143–156.
Trexler, J. C., J. Travis, and M. Trexler. 1990. Phenotypic plasticity

in the sailfin molly, *Poecilia latipinna* (Pisces, Poeciliidae). 2. Laboratory experiment. Evolution 44:157–167.

Ungerer, M. C., C. R. Linder, and L. H. Rieseberg. 2003. Effects of genetic background on response to selection in experimental populations of *Arabidopsis thaliana*. Genetics 163:277–286.

van Noordwijk, A. J. and G. de Jong. 1986. Acquisition and allocation of resources—their influence on variation in life-history tactics. Am. Nat. 128:137–142.

van'T Land, J., J. van Putten, B. Zwaan, A. Kamping, and W. van Delden. 1999. Latitudinal variation in wild populations of *Drosophila melanogaster*: Heritabilities and reaction norms. J. Evol. Biol. 12:222–232.

Verhoeven, K. J. F., T. K. Vanhala, A. Biere, E. Nevo, and J. M. M. van Damme. 2004. The genetic basis of adaptive population differentiation: A quantitative trait locus analysis of fitness traits in two wild barley populations from contrasting habitats. Evolution 58:270–283.

Via, S. and R. Lande. 1985. Genotype-environment interaction and the evolution of phenotypic plasticity. Evolution 39:505–522.

Wade, M. J. 2002. A gene's eye view of epistasis, selection and speciation. J. Evol. Biol. 15:337–346.

Weber, K. E. 1992. How small are the smallest selectable domains of form? Genetics 130:345–353.

Weber, K. E. 1996. Large genetic change at small fitness cost in large populations of *Drosophila melanogaster* selected for wind tunnel flight: Rethinking fitness surfaces. Genetics 144:205–213.

Whitlock, M. C., P. C. Phillips, F. B.-G. Moore, and S. J. Tonsor. 1995. Multiple fitness peaks and epistasis. Ann. Rev. Ecol. Syst. 26:601–629.

Wilkinson, G. S. and P. R. Reillo. 1994. Female choice response to artificial selection on an exaggerated male trait in a stalk-eyed fly. Proc. R. Soc. London Ser. B 255:1–6.

Worley, A. C., D. Houle, and S. C. H. Barrett. 2003. Consequences of hierarchical allocation for the evolution of life-history traits. Am. Nat. 161:153–167.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16: 603–619.