Selection Studies in Ecology: Concepts, Methods, and Directions

Ledyard Stebbins is said to have once blurted out, at the end an ecologist’s seminar, “you seem to be ignoring evolution.” An unnamed, eminent living ecologist once plainly stated “evolution does not interest me and is irrelevant to the questions I am asking.” Personalities aside, you cannot argue with the facts of what scientists are publishing in the pages of Ecology. Most every issue contains papers involving selection analyses and artificial selection studies aimed at addressing ecologically relevant questions. More and more graduate students are interested in integrating selection studies into ecology. Yet, some of the design issues, nuances of interpretation, and fruitful directions are not always clear. I have asked six of the top evolutionary biologists, proficient in asking ecological questions and known for their clarity of thought and elegant designs, to present concepts, methods, and directions for using selection studies in ecology.

The result is a surprising array of questions that can be addressed. Classically, the study of generalism and specialism in ecology has been addressed by attempting to detect trade-offs in fitness when organisms use alternate resources. However, selection studies have many other applications in ecology. For example, a growing interest in genotype-by-environment interactions is leading to the development of novel techniques to study the causes and consequences of phenotypic plasticity. Selection studies can also aid in understanding niche and species range boundaries and character displacement. Perhaps less well-known are studies of the role of frequency-dependent selection in the maintenance of species diversity. Finally, a subtle, yet general theme emerges, urging us to consider selection studies to understand organismal responses to global change.

This Special Feature illustrates the rich and timely role for selection studies in ecology. Although there is an obvious underrepresentation of studies on vertebrates, I hope that this bias is based more in the reality of the difficulties of working with vertebrates than in our own taxonomic biases. The authors have, however, emphasized methods for working with model organisms such as microbes and Arabidopsis, while continuing to highlight novel plant and arthropod systems. Use of artificial selection, hybridization, quantitative trait locus (QTL) analysis, and mutagenesis figure prominently in the suggested toolbox for ecologists interested in employing selection studies.

—ANURAG A. AGRAWAL

Special Features Editor

Key words: artificial selection; evolutionary ecology, hybridization; model systems; phenotypic plasticity; quantitative trait loci analysis; trade-offs.

© 2003 by the Ecological Society of America

1 Feature accepted 22 September 2002. Reprints of this 64-page Special Feature are available for $9.50 each, either digitally (as pdf files) or as hard copy. Prepayment is required. Order reprints from the Ecological Society of America, Attention: Reprint Department, 1707 H Street, N.W., Suite 400, Washington D.C. 20006.
ARTIFICIAL SELECTION: A POWERFUL TOOL FOR ECOLOGISTS

JEFFREY K. CONNER

Kellogg Biological Station and Department of Plant Biology, Michigan State University, 3700 East Gull Lake Drive, Hickory Corners, Michigan 49060 USA

Abstract. Artificial selection has been practiced by humans since the dawn of agriculture, but only recently have evolutionary ecologists turned to this tool to understand nature. To perform artificial selection, the phenotypic trait of interest is measured on a population, and the individuals with the most extreme phenotypic values are bred to produce the next generation. The change in the mean of the selected trait across each generation is the response to selection, and other traits can also evolve due to genetic correlations with the selected trait. Artificial selection can directly answer the question of how quickly a trait will evolve under a given strength of selection. This kind of result can help ecologists determine whether range or niche boundaries are determined by a lack of variation for a key phenotypic trait or trade-offs due to genetic correlations with other fitness-related traits. In a related approach, controlled natural selection, the organisms are not selected according to their values for a given trait, but rather are allowed to evolve for one to several generations under experimentally imposed environmental treatments such as temperature, light, nutrients, presence or absence of predators or competitors, etc. The results of this kind of study can tell us how quickly a population can adapt to a given environmental change, either natural or anthropogenic. Finally, artificial selection can create more variation for measurements of natural selection or can be coupled with QTL mapping; both these combinations provide new insights into adaptation. I discuss advantages and disadvantages of these approaches relative to other kinds of studies and highlight case studies showing how these tools can answer a wide range of basic and applied questions in ecology, ranging from niche and range boundaries and character displacement to climate change and invasive species.

Key words: adaptation; artificial selection; character displacement; constraints; controlled natural selection; environmental change; genetic variation and correlation; niche and range boundaries.

INTRODUCTION

All of our domesticated plants and animals, from crops to pets, from Chihuahuas to Great Danes, are the products of artificial selection. Darwin devoted the entire first chapter of On the Origin of Species to the products of artificial selection, in particular the morphologically diverse breeds of domesticated pigeons. Artificial selection is simply human-directed evolution, and is important to biologists because it tells us what a given strength and form of selection can accomplish in terms of phenotypic change and how quickly this change can occur. In other words, artificial selection is the best technique to determine the nature and strength of genetic constraints on evolutionary change.

Why are evolutionary change and genetic constraints important to ecology? Antonovics (1976) made a strong case for why ecologists should care about genetics, and two of his “tenets” are directly relevant to artificial selection. One states that, because the range of environments that an organism can inhabit may be determined at least in part by genetic constraints, explaining the distribution and abundance of species is partly a genetic issue. These constraints include lack of genetic variance for a trait, or a genetic correlation between traits that are both under natural selection. Through artificial selection the ecologist can test the ability of a species to adapt to conditions outside the boundaries of its niche or geographic range.

Antonovics also claimed that the distinction between ecological and evolutionary time was artificial and misleading, and supported his view mainly with examples of rapid evolution in response to anthropogenic change. Therefore, an understanding of how quickly organisms can adapt to environmental change is fundamental to trying to predict and explain alterations in community composition that may result from climate change, invasive species, or habitat destruction and fragmentation (e.g., Etterson and Shaw 2001). There are also accumulating examples of rapid evolution in response to natural changes in the environment, especially community shifts such as introductions or local extinctions of predators or competitors (e.g., Seeley 1986, Reznick et al. 1997, Thompson 1998).

Artificial selection is a particularly useful tool in the study of adaptation to the environment, which is an ecological as well as an evolutionary process. If a trait is important for fitness, then the variants that are less
well adapted have already been removed by selection, so that adaptation in current populations is hard to demonstrate (Grafen 1988). Artificial selection can be used to increase the variation, recreating some of the putatively ancestral phenotypes that have been eliminated by selection. The results of this selection can then be placed in the field to test the fitness effects of these “new” variants.

A new and particularly exciting approach is to combine artificial selection with QTL mapping to understand the genetic basis of character displacement and other selective mechanisms of population and species divergence (see Lexer et al. [2003] for a related approach). Here, artificial selection is used to simulate divergent or disruptive selection in nature, and then the newly divergent lines are crossed to create a mapping population. By mapping the genetic changes that have occurred during artificial selection, one can determine the number of genes causing the crucial early stages of divergence and begin to examine their effects. If crosses between replicated selection lines are mapped separately, then the repeatability of divergent evolution at the genetic level can be assessed. These kinds of data are critical for understanding the patterns and rates of rapid evolutionary responses to environmental change, both natural and anthropogenic.

Here, I review each of these applications of artificial selection as a tool for ecological research. I highlight selected examples in each case and suggest the kinds of questions and studies for which this approach can provide new insights.

**Definitions**

Artificial selection provides information about heritability and additive genetic variances, covariances, and correlations. The total phenotypic variance ($V_P$) in a population can be broken down into environmental variance ($V_E$), additive genetic variance ($V_A$), and two components of nonadditive genetic variance due to dominance and epistasis ($V_D$ and $V_E$, respectively). Of these variance components, only additive variance is directly available for natural or artificial selection to act upon and cause evolutionary change in an outcrossing species. Additive variance is often represented as the heritability, defined as the proportion of phenotypic variance that is due to additive genetic causes ($V_A/V_P$).

I will use the terms additive genetic covariance and correlation ($r_A$) interchangeably in this review, because a correlation is just a standardized covariance. A genetic correlation between two phenotypic traits can be caused either by pleiotropy, in which one locus affects both traits, or by linkage disequilibrium, in which the two traits are affected by distinct gene loci, but some evolutionary force creates and maintains a nonrandom association between the alleles present at these loci (Falconer and Mackay 1996). Genetic correlations are important because natural or artificial selection on one character causes an evolutionary change in a correlated neutral character, or alters the response to selection in a correlated character that is itself under direct selection. In other words, nonadaptive evolution of the second character can occur due to genes shared with the selected trait.

**Approaches**

There are two approaches to artificial selection that need to be distinguished (see Fry [2003] for additional discussion). In the traditional breeder’s approach, the experimenter applies a known amount of selection to a single phenotypic trait by measuring the trait and breeding only those individuals with extreme values for the trait (Fig. 1). The difference between the mean of the entire measured population and that of the subset selected for breeding is called the selection differential ($S$). After one or several generations of this selection, the per generation change in the mean of this trait (the response to selection, $R$) is measured. The ratio of these gives the realized heritability: $h^2 = R/S$.

The higher the heritability, the more rapid the response to a given strength of selection (compare Fig. 1A and B). If selected lines differ in traits other than the one selected after selection, then this correlated response to selection is clear evidence of an additive genetic covariance between the selected trait and these other traits (Fig. 1C). I use the term artificial selection only for this traditional breeder’s approach, and give a number of examples of its application to ecological questions below. This type of experiment provides direct estimates of the magnitude of genetic variation for a trait and the genetic covariance between the selected trait and other traits. Therefore, artificial selection is the best way to determine how fast a single trait will evolve with a given strength of selection.

The second approach can be called controlled natural selection, or more accurately, natural selection in a controlled environment. I call it natural selection because the experimenter does not decide which individuals survive and reproduce, as is done in artificial selection, but rather imposes some environmental treatment (e.g., several temperature, food, or light regimens; presence or absence of a predator or competitor) and lets the organisms reproduce in this environment for one to several generations. The organisms from the different treatments are then all placed in the same environment (a ‘‘common garden’’) or samples from all treatments are placed in each of the treatment environments (a ‘‘reciprocal transplant’’) and the traits of interest are measured. Differences in trait means among treatment groups measured in the same environment means that there is additive variance for the trait(s), and is a direct measure of the speed of the evolutionary response to the different environmental treatments. This is also a direct test of whether genetic constraints limit an organism’s niche along the environmental dimension being manipulated. Examples of this approach
Fig. 1. Hypothetical data depicting heritability, response to selection, genetic correlation, and correlated response to selection. Each point represents mean phenotypic values for one full-sibling family. The lines drawn to the axes represent the means of the entire population before and after selection. (A, B) Regression of trait values (flower number in this example) in offspring on values of the same trait in the parents. The slope of this offspring–parent regression line is the heritability ($h^2$). The filled circles are the individuals selected for breeding; the difference in the mean of all parents and the selected parents is the selection differential, $S$. The difference in the mean of the entire parental generation (both selected and unselected individuals) and the mean of the offspring is the response to selection, $R$, which is the product of $h^2$ and $S$. The selection differential is the same in panels (A) and (B), so the greater response in panel (B) is due to the higher heritability (steep slope). Note that, in a real artificial selection experiment, the offspring in the unselected families (open circles) would never be produced, because those parents would not be selected for mating, so the filled circles represent the entire offspring generation. (C) The relationship between one trait in the offspring and a different trait in the parents (flower number and size in this example) reflects the genetic correlation between these two traits. The genetic correlation is negative in this case, indicating a possible resource trade-off (cf. Worley and Barrett 2000). Selection to increase flower number in the parents produces a reduction in flower size (the correlated response or CR) in the offspring; the magnitude of the CR depends on the strength of selection and the magnitude of the genetic correlation between the two traits. Modified from Arnold (1987).

are also given in Uses of Artificial Selection in Ecology: Controlled natural selection.

Another approach is measuring the strength of natural selection in the wild, sometimes called “selection experiments.” This term is both vague and misleading, because these studies are usually observational, not experimental, measuring the relationship between fitness and phenotypic variation in undisturbed populations. Unlike artificial selection, studies of natural selection in the wild are useful for understanding present-day adaptations, but they do not address questions concerning genetic variation or genetic constraints. Artificial selection imposes selection ($S$) to measure genetic variance ($h^2$), while studies of natural selection in the wild measure selection ($S$) without producing any information about genetic variance ($h^2$). Studies of natural selection in the wild have been reviewed elsewhere (Endler 1986, Brodie et al. 1995, Kingsolver et al. 2001), and another paper in this issue (Lexer et al. 2003) discusses an extension to natural selection studies. Here, I address studies that combine artificial selection with natural selection, an integrated approach to studying adaptation and constraint.

USES OF ARTIFICIAL SELECTION IN ECOLOGY

Detecting genetic variation and covariation

Artificial selection is an excellent tool to determine whether a trait can respond to future selection due to natural or anthropogenic changes in the environment. It also allows testing of hypotheses about range or niche limits, especially when these hypotheses predict that
the range or niche of a species is limited by the values of a key phenotypic trait, such as the bill length of a hummingbird determining the species of plants the bird can feed on. By artificially selecting on that trait, the ecologist can determine if evolution outside the current range or niche is genetically possible.

Artificial selection has several advantages for determining genetic variation and covariation over the alternative methods of offspring–parent regression and sibling analysis (hereafter referred to as single-generation methods). First and foremost, artificial selection directly answers whether the trait can evolve in response to selection, whereas single-generation methods are indirect. Artificial selection is simpler conceptually and practically than sibling analysis, and has greater statistical power for the same number of individuals measured. This is primarily because artificial selection tests differences between line means, whereas sib analysis relies on variance and covariance components (Falconer and Mackay 1996). This means that artificial selection is a good choice for traits that are difficult to measure (e.g., physiology, behavior), because generally fewer individuals need to be measured at one time compared to single-generation methods.

Artificial selection also has some disadvantages compared to single-generation methods. Each artificial-selection experiment can measure additive variance only for the selected trait and can only estimate genetic covariances between this trait and other measured traits. Therefore, a single artificial selection experiment cannot be used to measure the entire matrix of genetic variances and covariances among a group of traits (G); a separate artificial selection experiment for each trait is necessary. In other words, artificial selection provides no information on the genetic covariance or correlation between two unselected traits, and does not provide a quantitative estimate of genetic variance or heritability for any unselected trait. This information is necessary for making quantitative predictions of the speed of evolutionary change for several traits (e.g., Campbell 1996).

Artificial-selection experiments are not practical for many organisms, because they typically require maintaining the organisms in the laboratory or greenhouse, sometimes for long periods, and controlled matings need to be performed. In some organisms, single-generation experiments can be done relatively easily in the field. In monogamous birds, for example, offspring–parent regression has often been carried out with natural matings in natural populations (e.g., Schluter and Smith 1986; reviewed in Boag and van Noordwijk 1987). Since the magnitude of genetic variance and covariance can be strongly affected by the environment (genotype by environment interaction), confining the experiments to unnatural environments can be a serious shortcoming if quantitative estimates of variance are of interest. A final practical problem is that artificial selection is most efficient if the individuals that are measured can then be mated, so that traits that cannot be measured on live individuals pose more difficulty. This difficulty can be overcome by measuring some individuals and using their clones (where possible) or full siblings to mate, but the latter will slow progress because the measured and mated individuals only share half of their genes.

Several studies have used artificial selection to demonstrate genetic variation for traits for which such variation might not be expected. These are traits that are maintained across higher taxonomic levels (families and above) and are diagnostic for these taxa. This means that ecologists should be wary of assuming a trait cannot evolve to adapt to new conditions simply because it varies little across species. Perhaps most remarkably, Holthorp (1944) selected for increased cotyledon number in Brassica, and was able to produce high frequencies of seedlings with three and four cotyledons after only two generations of selection. Therefore, a character that distinguishes dicots, and has been stable in some groups almost since the origin of flowering plants, is genetically variable. Two studies have found similar results at the family level. Huether (1968) was able to create populations of Linanthus with high frequencies of plants that deviated from the five petal lobes characteristic of the phlox family (Polemoniaceae). Dimorphic stamens within flowers (four long and two short) are diagnostic for the family Brassicaceae (mustards). Karoly and Conner (2000) reported significant decreases in dimorphism after one generation of artificial selection for decreased dimorphism in Brassica rapa, and after three generations a number of plants in the selected lines (but none in the control lines) had nearly lost the dimorphism (Fig. 2). Taken together, the results of these three studies suggest that a lack of additive genetic variation may rarely be the explanation for the evolutionary stasis of traits diagnostic at the family or higher taxonomic level. These traits may be maintained by stabilizing selection or constrained by genetic correlations with other traits under selection.

Artificial selection has also been used to test for the presence of genetic correlations, which are at the heart of theoretical work in a number of areas important to ecology, such as sexual selection, evolutionary constraints, and trade-offs among fitness-related traits. The eye stalks of stalk-eyed flies have been enormously exaggerated by sexual selection, so that the span between the eyes is greater than the overall body length in some species. Wilkinson (1993) reported a significant response to artificial selection for increased and decreased eye span in Cyrtodiopsis dalmanni, indicating that further increases in eye span are not constrained by a lack of variation. Females in these lines showed a correlated response to this selection; females in control and increased eye-span lines preferred males with larger eye spans, while females in decreased eye-span lines preferred males with smaller spans (Wilkin-
son and Reillo 1994). This result supports a key prediction of some sexual selection theories—that sexually selected traits in males and the female preferences for them should be genetically correlated.

Negative genetic correlations have been sought as evidence for trade-offs between fitness-related traits using artificial selection. Siemens and Mitchell-Olds (1998) selected for increased and decreased constitutive levels of defensive chemicals (glucosinolates) and an enzyme (myrosinase) involved in the hydrolysis of glucosinolates into compounds that are more toxic. They found little evidence for the negative genetic correlations between constitutive and induced levels of defensive compounds that is predicted by allocation theory. They did report a negative genetic correlation between high constitutive levels of myrosinase and field seed production, evidence for a cost of defense. Using the same myrosinase selection lines, Strauss et al. (1999) reported one positive and one negative genetic correlation between myrosinase and floral traits affecting pollination, but for the most part pollination was uncorrelated with myrosinase. Worley and Barrett (2000) selected for increased flower number and both increased and decreased flower size (two replicates of each), and found significant correlated responses in three of the six total replicates; these were all negative, supporting the existence of trade-offs between flower size and number.

The studies described above performed directional selection on one or two traits to test for the presence of a genetic correlation between them. Another type of study selects on variation in two traits jointly, an approach called index selection by breeders. At least three studies have selected in a direction perpendicular to the major axis of a correlation between two traits, which means selecting in the direction of least variation in bivariate space (Fig. 3A). J. Conner and K. Karoly (unpublished data) selected on the ratio of filament and corolla tube lengths (anther exsertion), two traits that are highly genetically correlated in wild radish (Conner and Via 1993; see Uses of Artificial Selection in Ecology: Creating novel phenotypic variation). The high and low ratio lines diverged rapidly, but the correlation within lines did not change (Fig. 3B). A very similar result occurred when Emlen (1996) selected on the residual of the sigmoidal allometric relationship between horn and body size in a dung beetle (Onthophagus acuminatus). This selection caused the position of the allometric curve to shift to the left and the right in the divergent lines, but the shape of the allometric curves within each line did not change. Stanton and Young (1994) selected on the ratio of petal area and pollen production in wild radish, also highly genetically correlated traits, and also obtained a rapid response. They did not report the correlations within each selected line after selection, so it is not known if they were altered.

These three studies clearly demonstrate that the lower amount of variation perpendicular to a strong bivariate relationship between two traits is not a strong constraint on evolution of the relative magnitudes of the two traits. While the relative magnitudes of the two traits responded to selection, the correlations between the traits within selected lines did not. This latter fact is not surprising theoretically, because these experiments imposed directional selection perpendicular to the correlation, rather than selection directly on the correlation (correlational selection; Brodie 1992). Thus, relationships among traits (correlations, allometries) may be quite robust in the face of directional selection that changes the means of these traits, at least over a few generations. This is relevant to ecology because these correlations may be adaptations to the environment (see next paragraph), and because correlations may cause trade-offs that constrain niche or range expansion. However, combining the divergently selected lines does result in a lower correlation (more
variation perpendicular to the major axis) in the composite population in all three cases, a fact that can be useful for testing the adaptiveness of the correlation experimentally (see Uses of Artificial Selection in Ecology: Creating novel phenotypic variation).

Another feature that these three examples share is that the relationship between the two traits may be adaptive. The sigmoidal allometry between horn and body size in many horned beetles essentially gives rise to two morphs with differing mating strategies. Smaller males have little or no horns, and use stealth or other nonaggressive tactics to obtain mates, while large males use their horns in intermale combat (e.g., Eberhard 1982, Emlen 2000). Emlen (1996) suggests that divergence among species in the genus Onthophagus occurs by shifting the allometric relationship along the body size axis without major alterations in the relationship itself, similar to the response to artificial selection. In wild radish, the ratio of filament and corolla tube determines the position of the anthers relative to the opening of the tube (anther exsertion), and this affects pollination success and male fitness in this and other species (Conner et al. 1995, Morgan and Conner 2001 and references therein). Petal area in wild radish affects pollinator visitation frequency and thus pollen removal rates (Young and Stanton 1990, Stanton et al. 1991, Conner and Rush 1996), so that a correlation between petal area and pollen production could be adaptive. For example, a plant with large petals and low pollen production would have rapid pollen depletion, making further pollinator attraction a waste of the resources devoted to the large petals. Conversely, a plant with small petals and high pollen production might not be visited enough to have most of its pollen removed. Therefore, in all three cases the bivariate relationship between the traits found in natural populations may be adaptive.

**Creating novel phenotypic variation**

While artificial selection by itself sheds little light on adaptation, when used to generate increased variation for studies of natural selection, artificial selection can be an extremely useful tool for understanding the often difficult topic of adaptation. Ecologists commonly use observational methods of measuring natural selection in the wild (selection differentials and gradients; Lande and Arnold 1983) to attempt to understand which traits are adaptations to given biotic or abiotic factors in the environment. By using artificial selection to create additional variation for subsequent measurements of natural selection, one can avoid several problems associated with natural selection studies. The first problem is that the regression estimates of selection differentials and gradients have little power in the phenotypic extremes, because there are typically few individuals in the population expressing these extremes (Schluter 1988). This is particularly likely to be true if the trait is indeed an adaptation, because past

---

**Fig. 3.** Artificial selection perpendicular to the major axis of the correlation between filament and corolla tube lengths in wild radish (*Raphanus raphanistrum*). (A) The genetic correlation in the original population. The arrows show the direction of selection in the high and low anther exsertion lines; note that these are in the direction of least variation in bivariate space. Each point is the mean of all offspring of one sire from a nested half-sibling design (Conner and Via 1993). The correlation of these sire family means is an estimate of the additive genetic correlation. (B) Results after five or six generations of selection. Note that selection has moved the elliptical cloud of points in the directions of the arrows in A without changing the shape of the ellipse, that is, the correlation within each group. Each point is a full-sibling family mean; the resulting correlations are “broad-sense” genetic correlations that include covariance due to dominance and maternal effects. The difference in estimation methods between panels (A) and (B) is responsible for the greater range of values in panel (B).
selection has eliminated extreme individuals. Artificial selection for extreme phenotypes can be very effective in increasing the representation of these extremes in the population.

A fundamental shortcoming of all studies of evolution in present-day populations is that the forces that created the trait of interest occurred in the past. While this can never be completely overcome, artificial selection combined with natural selection can help. As noted above, past selection may have eliminated unfit variants, so that a lack of selection in the present could be due to the current phenotypic distribution of the population occurring at an adaptive peak (Fig. 4A). Artificial selection cannot only increase the representation of existing extreme phenotypes, it can also increase the range of phenotypic variation; subsequent studies of natural selection on these artificially selected populations can determine if these new variants are less fit. If they are (Fig. 4B), then this is good evidence that the trait is adaptive. If the new variants are not less fit (Fig. 4C), then this suggests that the trait is neutral and thus not an adaptation. These kinds of data are even more interesting if combined with comparative or fossil information that indicates that these novel phenotypic variants were present in a recent ancestor.

A final problem with the observational techniques for measuring natural selection is that they can never prove a causal link between a trait and fitness, because there may be correlations between the trait of interest and unmeasured traits or environmental variables that are causing the appearance of selection on the trait (Mitchell-Olids and Shaw 1987, Rausher 1992). For example, soil that is high in nitrates may cause plants growing there to have higher levels of alkaloid chemicals and higher fitness than plants growing in low-nitrate soils. An observational study would show positive selection on alkaloid levels, even if there was no causal link between alkaloids and fitness. These problems can be at least partly overcome by conducting divergent artificial selection on the trait of interest with replicate lines and random-mated controls. If selection in the field is then estimated in populations made up of representatives of all these lines, then a significant difference among the mean fitnesses of these lines is powerful evidence of a causal relationship with fitness. Properly designed and conducted, this technique eliminates artifactual selection due to environmental and phenotypic correlations. However, traits that are genetically correlated with the selected trait will evolve under the artificial selection and thus could be the cause of fitness differences.

An alternative method to artificial selection for overcoming these same problems is direct manipulation of the phenotype, including artificial structures. For example, Andersson (1982) manipulated tail length in Widowbirds and measured the effect of this manipulation on pairing success, and Stone and Thompson (1994) and Schemske and Ágren (1995) created artificial flowers to test hypotheses of selection on anther position and flower size respectively. Direct manipulation can be much simpler and quicker than artificial selection, and can more easily create very extreme phenotypes. It also does not suffer from the problem of genetically correlated traits also evolving in response to artificial selection, which can cause interpretation errors. However, direct manipulation can have a similar problem, in that it can be very difficult to manipulate a trait without harmful or unknown effects on other traits. This can make it difficult to design appropriate controls, and may be an advantage for artificial selection because developmental constraints make it difficult for selection to produce large phenotypic changes, desired or not. In the extreme, there are some traits that cannot be directly manipulated without seriously harming the organism; similarly, there are some traits that are difficult or impossible to select on, or that will not respond to selection, although the latter seem to be rare. Therefore, often the choice between artificial selection and direct manipulation will depend on the organism and the specific trait of interest.

While artificial selection is more time consuming, in many cases it can produce a larger final sample size. For example, it is difficult to directly manipulate all the flowers in a population of a plant species that pro-

Fig. 4. A hypothetical example of the use of expanded phenotypic variation from artificial selection to test for adaptation. The histograms represent the phenotypic frequency distribution, and the curves above are the fitness functions showing the fitnesses of each phenotypic value. The fitness functions can be estimated as selection gradients. (A) The natural population; there is no significant selection, i.e., the slope and curvature of the fitness function are not statistically different from zero. (B) After artificial selection has increased phenotypic variance; there is now significant stabilizing selection, indicating that the trait is an adaptation. (C) Also after artificial selection has expanded variation. There is still no significant selection in this case, indicating that the trait is neutral (not an adaptation) over this expanded phenotypic range.
duces many flowers per individual. By artificially selecting on the floral trait and producing many seeds in the final generation of selection, very large sample sizes can be obtained at one time. Similar arguments could be made for insect species that can be reared in large numbers. On the other hand, artificial selection is more difficult than direct manipulation when the organisms are difficult to rear and breed in captivity. One final advantage of artificial selection over direct manipulation is that it produces information on genetic variation and constraint, topics that manipulation does not address.

J. Conner, K. Karoly, and colleagues are currently using artificial selection to test whether two anther position traits that lack variation at different taxonomic levels are adaptive. As noted in *Uses of Artificial Selection in Ecology: Detecting genetic variation and covariation*, dimorphic anther positions (resulting from stamen dimorphism) are present in most of the over 3000 species of Brassicaceae. Anther exsertion, the degree to which the long stamen anthers protrude beyond the opening of the corolla tube, has low variation in wild radish and some other mustards due to an extremely high correlation between the lengths of the stamens and corolla tube (Conner and Sterling 1995). Five generations of artificial selection for decreased dimorphism (analogous to the *Brassica* experiment in Karoly and Conner 2000), and for increased and decreased exsertion, produced increased variation for these traits in wild radish (e.g., Fig. 5). Note that the population with increased variation in exsertion is not identical to the high and low selected lines together with the random-mated controls. As discussed in *Uses of Artificial Selection in Ecology: Detecting genetic variation and covariation*, this selection did not alter the correlation between filament and corolla tube lengths within lines, and therefore did not alter the variance of exsertion within lines either.

These composite anther exsertion populations were placed in the field for natural pollination, and male fitness (seed siring success) will be measured when molecular genetic paternity analysis is completed. Two sets of complementary analyses will then be conducted. Analysis of variance with male fitness as the response variable and the selection treatment as the predictor variable will test the overall fitness effects of the treatments, correcting for phenotypic and environmental correlations (including correlations with unmeasured traits) but not for any correlated responses to the selection due to genetic correlations with the anther traits. Multivariate selection gradient analysis (Lande and Arnold 1983) will test for the effects of the selected traits on fitness, correcting for correlated responses in other measured traits but not unmeasured traits.

**Controlled natural selection**

As noted in *Approaches*, a closely related alternative to artificial selection is natural selection under controlled environments, or “controlled natural selection.” Here, the experimenter manipulates an environmental variable and measures the evolutionary response. In this approach, selection is not directly imposed by the investigator as it is in artificial selection, but an environmental factor that may cause selection is manipulated. The results of a controlled natural-selection experiment show whether an organism can adapt to a specific environmental change, or move outside the boundaries of its current niche. Since a known amount of selection is not being applied to a specific trait, the heritability cannot be quantitatively estimated, but this is often not the primary concern to an ecologist. In addition, little or no information can be gained about genetic correlations among traits with controlled natural selection, because the traits that are the targets of direct selection are unknown. Fry (2003) gives examples of this approach applied to trade-offs.

Controlled natural selection can be conducted in the field with some species, if key environmental variables can be manipulated. An excellent example of this approach comes from work on guppy evolution in Trinidad (Reznick et al. 1997). In two river drainages, guppies were moved from high-predation sites with cichlid fish to low-predation sites above rapids or waterfalls.
where there were no guppies or cichlids. After 7, 13, and 18 generations, samples of the guppies from the source (control) population and from the transplanted populations were brought to the lab. After one generation in the lab to reduce maternal effects, the age and size at maturity was measured. As predicted by life-history theory, in almost all cases the fish transplanted to the lower predation sites matured later and at a larger size than the high-predation control populations. This is a clear example of rapid evolution in response to a change in the environment.

Controlled natural selection can also be valuable to understand and predict evolution in response to anthropogenic changes in the environment, but this approach has rarely been used. In one of the best examples, Ward et al. (2000) selected for increased seed production for five generations in *Arabidopsis thaliana* at two concentrations of CO₂. Because seed production is an excellent measure of fitness for these selfing annuals, this is essentially a controlled natural selection experiment. The low CO₂ concentration mimicked the low levels that occurred in the Pleistocene, whereas the high concentration represented predicted levels by the end of this century. They found that plants flowered later and attained a higher biomass relative to random mates at low CO₂, and showed the opposite response in these traits at high CO₂. Ward et al. then performed a reciprocal transplant experiment with their selected lines, planting the last generation of each at both CO₂ concentrations, after first passing them all through a generation at intermediate CO₂ to minimize maternal effects (a critical consideration for all common-garden and reciprocal transplant studies). They found that both lines produced the most seed at the concentration they had been selected in, but the difference was only statistically significant when the plants were tested at the lower concentration. This latter result indicates that *A. thaliana* can rapidly adapt to decreases in CO₂, and perhaps to increases as well.

Controlled natural selection is therefore a valuable tool that could be applied to many more questions in ecology, particularly with short-lived organisms. Areas that seem particularly ripe for this approach are niche dimensions, range limits, and character displacement. By manipulating one or more dimensions of an organism’s niche for several generations, one could test the environmental limits of adaptation of an organism, which could shed light on both niche and range limits. Placing individuals from allopatric populations of two species with broadly overlapping niches together for several generations might well reproduce the early stages of character displacement. This type of experiment would be an extension of a single-generation experiment demonstrating selection for character displacement in stickleback fish (Schluter 1994). The controlled natural selection approach could be used to predict the results of anthropogenic changes in the environment other than CO₂, such as habitat fragmentation, invasive species, and biocontrol agents. For example, biocontrol agents have sometimes attacked native species related to the invasive they were imported to attack (e.g., Louda et al. 1997). Even if the biocontrol agent cannot survive on these native hosts when it is introduced, it is possible that rapid evolution of host range could occur after the introduction. This could be tested by rearing the biocontrol agent on different mixtures of the invasive and native species for several generations, to see if adaptation to, and use of, the native increases.

**Future direction: combine with QTL mapping**

I have discussed how artificial selection can be combined with studies of natural selection to produce new insights into adaptation. Another powerful combination is artificial selection (or controlled natural selection) and quantitative trait locus (QTL) mapping, which can be used to examine the genetic basis of adaptive divergence, including character displacement, local adaptation, and incipient speciation. QTL mapping uses genetic markers scattered throughout the genome to find gene regions that affect quantitative traits (Tank- sley 1993, Mitchell-Olds 1995). Quantitative traits are those that are affected by several to many gene loci, as well as the environment, leading to a continuous phenotypic distribution. The vast majority of traits that ecologists are interested in are quantitative.

If QTL mapping is applied to a cross between divergently selected lines, then it can answer questions such as: How many gene loci were responsible for the observed response to selection? What is the distribution of the magnitude of effects of these loci on the selected trait, i.e., are there a few loci of major effect, many loci of small effect, or a continuum of magnitudes of effect? These genetic details matter to evolutionary ecologists, because they affect the direction and rate of adaptive evolution, and therefore affect interactions of organisms with their biotic and abiotic environments and the rate of adaptation to global change (Orr and Coyne 1992, Mitchell-Olds 1995, Orr 1998). This approach can also be used to probe the genetic basis of trade-offs, a central theme in evolutionary ecology (Fox et al. 2001). Artificial selection has already been used to study trade-offs (e.g., Service et al. 1988), and the addition of QTL mapping can greatly improve our understanding. For example, artificial selection combined with QTL mapping can tell us whether the trade-offs are due to pleiotropic loci, and provide a first step in determining what these loci code for. This knowledge would give us a mechanistic understanding of ecological trade-offs. Finally, if separate crosses between replicate divergent lines are mapped separately, then one can also ask whether evolution is repeatable; that is, are the same loci responsible for divergence in each case? This question may be relevant to the generation of biodiversity, as less diversity would be cre-
ated if evolution is highly repeatable than if evolution is more idiosyncratic.

Mackay and colleagues (Long et al. 1995, Gurganus et al. 1999, Nuzhdin et al. 1999) have conducted what is probably the best study combining artificial selection and QTL mapping for a natural population. They selected for increased and decreased abdominal and sterno-pleural bristle number in two separate sets of Drosophila melanogaster lines for 25 generations. They then produced detailed genetic maps from these selection lines, using several powerful genetic techniques that have been developed for this species. They found a total of 26 loci that were responsible for the response to their selection, and 20 of these mapped to locations of known bristle number or nervous system gene loci (the bristles are sensory). These results strongly suggest that these approaches can be successful in identifying the genes responsible for adaptation to specific factors in the environment, although this goal is more distant for nonmodel organisms.

A similar approach is being taken by J. Conner, L. A. Prather, and J. Hancock, using the wild radish anther exsertion selection lines. From each replicate pair of high- and low-exsertion lines, the plants with the highest and lowest exsertion were crossed, and then the F1 offspring of these matings were crossed separately to create two separate F2 mapping populations, one for each pair of high and low exsertion lines. QTL mapping will be used to identify the gene regions responsible for the rapid divergence of exsertion observed during the artificial selection. By mapping the crosses between the two replicates separately, we can ask how repeatable short-term evolution is genetically. Therefore, this kind of study should help ecologists understand the details of how adaptation to changes in the environment occurs, help to explain current species limits, as well as predict the speed of adaptation in response to natural and anthropogenic environmental change. Because wild radish is a serious weed, it may improve our knowledge of how weedy and invasive species can rapidly adapt to new environments.

**Conclusions**

In this review, I have tried to illustrate several ways in which artificial selection and controlled natural selection are useful tools for ecologists. These tools can provide unique insights into a number of fundamental and difficult issues, such as constraints on adaptation, trade-offs among fitness-related traits, niche and range limits, and character displacement. These multigeneration approaches have rarely, if ever, been applied to the latter two questions. When used to generate increased variation for studies of natural selection, or when combined with QTL mapping, artificial selection can shed light on the adaptiveness of traits in nature and the genetic mechanisms of rapid adaptation. Artificial selection and controlled natural selection can also be used to directly test the adaptive responses of populations to a variety of anthropogenic changes in the environment. Knowledge gained from these kinds of studies will improve our ability to predict the effects of such changes on natural communities and agroecosystems. For these reasons, I argue that artificial selection and controlled natural selection should become standard approaches in ecological science.

**Acknowledgments**

I thank Keith Karoly for many discussions of artificial selection; Peter Curtis, Doug Schemske, Dolph Schluter, Allison Snow, and Joy Ward for help with literature; Anurag Agrawal for encouraging me to write this; and Jeremie Fant, Frances Knapczyk, Angela Roles, Heather Sahli, Doug Schemske, Buffy Silverman, and Christy Stewart for comments on the manuscript. My research is supported by the National Science Foundation under grants DEB 9903880 and 0108354 and by the Cooperative State Research, Education, and Extension Service, U.S. Department of Agriculture, under Agreement No. 2002*35320-11538. This is Kellogg Biological Station contribution no. 978.

**Literature Cited**


ARTIFICIAL SELECTION AND THE DEVELOPMENT OF ECOLOGICALLY RELEVANT PHENOTYPES

PAUL M. BRAKEFIELD

Institute of Biology, Leiden University, P.O. Box 9516, 2300 RA Leiden, The Netherlands

Abstract. Artificial selection is used by quantitative geneticists to examine the genetic variances and covariances that underlie the evolution of traits that show continuous phenotypic variation. Such traits are crucial for adaptive evolution in natural populations. Here, I discuss the extent to which artificial selection experiments can provide insights about both the proximate and ultimate causes of adaptive evolution. In particular, such experiments can explore the basis of interactions among traits, and the potential consequences of genetic coupling of traits for the paths taken during evolution. They can also yield phenotypes of relevance to examining the reasons why differences in reproductive success and fitness occur. I focus here on the methodology, with particular reference to examples from investigations of the evolution of the development of morphologies and of phenotypic plasticity.

Key words: adaptive evolution; artificial selection; development; evolutionary constraints; genetic variances; life history; morphology; natural selection; phenotypic plasticity.

INTRODUCTION

This paper explores the extent to which artificial selection can provide a useful tool to ecologists. Artificial selection is generally used to examine the genetics of quantitative traits in the wider context of either describing the genetic variances involved in microevolution or of producing useful plant cultivars and animal breeds. Almost all traits of concern to ecologists are of this type, which show continuous phenotypic variation and whose study requires measurement. They may involve any feature of life history, physiology, behavior, or morphology. While some textbook cases of natural selection involve genetic polymorphisms and discrete phenotypes, adaptive evolution in response to environmental change will almost always involve quantitative traits each specified by alleles at more than one gene. Their evolution may have included a series of selective sweeps, each involving a particular allele increasing in frequency through to fixation under the influence of natural selection (cf. experimental evolution in laboratory populations of bacteria; Travisano et al. 1995). The genetic basis of phenotypic variation in the trait in extant populations is likely to involve polymorphism at a number of genes.

The evolution of adaptive traits in response to a new or changed environment is, in essence, a straightforward process. Mutation is the ultimate source of the genetic variation that is translated through developmental and physiological processes to generate variation in the phenotype among individuals. Such phenotypic variation provides the material that can be sorted or sieved by natural selection to evolve more effective adaptive traits of organisms. Individuals of a particular phenotypic class which have higher reproductive success or fitness because of enhanced survival, mating success, fertility, or fecundity will be favored by natural selection. If such a phenotype is at least partly inherited, then those alleles of genes with additive beneficial effects on the phenotype will increase in frequency within a population in a particular environment. Whether such a local adaptation will eventually spread through the whole species distribution will depend on the patterns and processes of population structure and environmental variation (Barton and Partridge 2000). The focus for evolutionary ecologists in the process of adaptive evolution has been on how patterns of phenotypic variation are influenced by natural selection (see Endler 1986, Mousseau et al. 2000). Recent analyses of the developmental processes which link phenotypes to genotypes are leading to more detailed insights about the generation of phenotypic variation which is relevant to evolutionary and ecological change (Stern 2000, Carroll et al. 2001, Beldade and Brakefield 2002). The primary issue in this paper is whether artificial selection in which divergent phenotypes are produced in the laboratory can help to unravel the process of adaptive evolution in more detail.

Artificial selection experiments may also be able to provide a useful tool for exploring the extent to which organisms will be able to ecologically adapt to the effects of climate change and global warming (see, e.g., de Jong and Brakefield 1998, Etterson and Shaw 2001, Thomas et al. 2001, Walther et al. 2002). For example, ecologists examining potential effects of (rapid) climate change have explored the responses of plant species or simple communities to an elevated level of carbon dioxide, such as is predicted to occur in a century.

1 E-mail: brakefield@ruisfb.leidenuniv.nl
or so from now. One problem with such “step-like” experiments is that they take no account of the potential for adaptive changes in gene frequencies and ecologically relevant traits in response to progressive environmental change. They may thus yield misleading results. This type of problem could be overcome, at least in part, by applying artificial selection to a short-lived species in the laboratory in an experiment in which the environment is shifted in a more gradual and realistic manner over generations.

**The Basic Principles of Artificial Selection**

The essence of artificial selection is captured in a key equation in evolutionary biology:

$$R = h^2 S.$$ 

The way in which this equation can be applied and the underlying concept of heritability are illustrated in Fig. 1 (from Boag and van Noordwijk 1987). More details of the methods of quantitative genetics can be found in numerous textbooks (e.g., Falconer and Mackay 1996, Roff 1997, Lynch and Walsh 1998). Here, I will concentrate on the basic principles needed to understand the application of artificial selection without attempting to cover all details.

The heritability in the narrow sense, or simply the heritability ($h^2$), of a trait is an estimate of the proportion of the phenotypic variation ($V_P$) in the trait which is due to additive genetic variance ($V_A$) such that

$$h^2 = \frac{V_A}{V_P}.$$

A broad-sense heritability, or the degree of genetic determination, includes nonadditive components of genetic variance in the numerator. The variation not accounted for by genetic variance ($V_G$) is due to environmental effects during development. This generates the environmental variance ($V_E$), such that

$$V_E = V_G + V_P.$$

Heritability can range from zero (no genetic variance) to unity (no environmental variance). Additive genetic variance is a statistical description of the additive contributions to development of the phenotype made by allelic variation at all the genes that contribute to the trait. This contrasts with nonadditive effects, which occur either through dominance relationships
among the alleles at individual polymorphic genes or the epistatic interactions among genes. The genetic variance \( (V_0) \) is the sum of the additive and nonadditive effects. The phenotypic response to artificial selection is, however, proportional to the additive genetic variance, explaining the importance of estimates of narrow sense heritability \( (h^2) \).

The heritability of a trait measures the extent to which individuals who share a proportion of their genes due to common descent resemble each other with respect to the phenotype of the trait. One method frequently used to quantify such a resemblance is a family breeding experiment in which offspring can be compared to their parents. This is illustrated in the top row of Fig. 1, which shows two hypothetical populations: one with high, and the other with low, heritability. The slope of the linear regression of the mean family values for offspring plotted against the mid-parent values provides an estimate of heritability. The significance of the estimate can be examined using the standard error of the slope.

Unfortunately, as stated by Roff (1997:390), the standard errors on heritability are often depressingly large. This is one potential advantage of an artificial selection experiment. At least for linear responses to selection maintained over a number of generations, the slope of the change in mean phenotype plotted against the cumulative amount of selection (cumulated selection differential) applied to the trait provides a comparatively robust estimate of heritability, a so-called realized heritability. The lower part of Fig. 1 illustrates the predicted phenotypic consequences in a generation of \( F_1 \) offspring given the high or low heritability coupled with a single generation of selection in a parental generation. The selection is imposed as in most artificial selection experiments through application of a threshold phenotypic value: only those individuals in the parental population above this threshold provide parents of the next generation. Thus, this represents a form of directional selection in an upwards direction (in contrast to downward selection in which individuals below a threshold phenotype are used as parents). The selection differential \( (S) \) is straightforwardly assessed as the difference in mean phenotypic value of the whole population at generation 0 from the corresponding mean for the selected parents. In a similar way, the response to selection \( (R) \) is given by the change in mean phenotypic value between generation 0 and that of all \( F_1 \) offspring in the following generation.

The equation, \( R = h^2S \), can be applied whenever values for two of the parameters are available to estimate the third. For example, knowing the amount of selection applied and having an estimate of heritability for the trait under directional selection enables a prediction of the response to selection, effectively the rate of evolutionary change. Fig. 1 illustrates how a higher heritability yields a more rapid phenotypic change than a low heritability. A trait with zero heritability is expected to show no phenotypic response to selection.

The above description only covers the basic principles of artificial selection, and, in particular, only those pertaining directly to a single target trait of selection. There are many factors that need to be taken into account when applying the theory in practice. An estimate of heritability, strictly speaking, only applies to the environment in which it was obtained and to the particular stock used in the breeding experiment. Clearly, if an environment induces more variability in development of the phenotype during ontogeny (and thus a higher \( V_0 \)), such as might be predicted when conditions are more stressful or less favorable, a lower heritability and a slower response to selection are expected. In parent–offspring breeding experiments, it is assumed that the environments of both generations are closely similar. Similarly, a breeding stock with less genetic variance due to some history of founder events, bottlenecks, or inbreeding is expected to yield lower estimates for heritability. Natural populations, as well as captive stocks, of various species have been found to vary in heritabilities for a variety of phenotypic traits (e.g., Falconer and Mackay 1996, Hoffman 2000). In summary, caution must be exercised in extrapolating from single estimates of heritability to any general predictions about responses to selection and trait evolvability (Houle 1992).

Such caution is especially wise when interpreting the results of single-generation breeding experiments, particularly when these are on a small scale with nonmodel organisms for which rearing conditions may be difficult to standardize. One important advantage of artificial selection experiments, especially when continued over many generations, is that they can provide more robust estimates of the genetic parameters for the trait targeted by the selection. Furthermore, in comparison to analyses of parent–offspring relationships or of a limited series of full-sib or half-sib families, such experiments which involve repeated episodes of expression of the consequences of any interactions with other traits tend to yield more convincing descriptions of the potential roles of genetic coupling among traits. Any genetic correlations of the target trait with other traits are likely to manifest themselves in correlated responses to the selection applied to the target trait itself (Lande 1979, Cheverud 1984). A complete description of the genetic variances of traits and of genetic covariances among traits, which can be used to make predictions about the evolution of multiple traits, is given in the so-called \( G \) matrix. The associated body of theory is beyond the scope of the present essay but is readily assessable in the recent review by Steppan et al. (2002).

Inbreeding depression resulting from small effective-population sizes and matings between relatives during artificial selection could represent a problem in later studies of the relationship between the phenotypic value and fitness. Indeed, wider application of the quan-
titative genetics theory assumes no coupling of the evolutionary and ecological dynamics during the process of artificial selection. There are at least two ways to avoid or minimize such problems. First, the artificial selection can use very large populations of selected individuals in each generation, or groups of known effective population size (breeding individuals) larger than those expected to introduce any (strong) effects of inbreeding or genetic drift (see Brakefield et al. 2001). Second, an experiment can include replicate, independent, selected lines in each direction. F₁ hybrids between these lines are expected to suffer much less strongly from any inbreeding depression that may be present in the lines themselves. The presence of replicate lines also has major advantages for the statistical robustness of estimates of genetic variances.

Finally, if one is particularly interested in exploring the fitness consequences of a change in a target trait, one or more periods of relaxed selection can be incorporated into the design of an artificial selection experiment. Monitoring of any change in phenotype of the target trait or of potentially coupled traits during the period(s) of relaxed selection can reveal insights about fitness relationships (see description in Is Natural Selection All Powerful? of the experiment performed by Beldade et al. [2002b]).

**Is Natural Selection All Powerful?**

Patterns of adaptive radiation described by ecologists are frequently spectacular, with everyone having their own favorite example (e.g., Lack 1947, Fryer and Iles 1972, Williamson 1981). If any of the classic case studies are viewed from the perspective of a population ecologist, it is not difficult to uncritically assume that the patterns of evolution have been shaped solely by natural selection. There is, indeed, a dramatic divergence in morphologies, behaviors, and life histories among the species involved in radiations such as those in the Galapagos Finches, the cichlids of the African rift-valley lakes, Hawaiian fruit flies, or South American passion vines. A possible scenario in each case is that whenever a vacant ecological environment has been colonized, evolution by natural selection has taken over to yield, more or less inevitably, a phenotype or design which opportunistically fits the new population or species to the novel environment it inhabits. This is indeed a feasible hypothesis that follows from the notion that natural selection is all powerful in driving or determining the evolution of the phenotype.

There is, however, an alternative hypothesis; namely, that at least some aspects of a pattern of adaptive radiation are also shaped or channeled by the internal properties of the organisms, and in particular the available genetic variances and the developmental mechanisms. Evolutionary paths may tend to follow genetic lines of least resistance (Schluter 1996). The existence of evolutionary constraints leading to bias in the patterns of divergence has been the subject of extensive conjecture and discussion (e.g., Maynard-Smith et al. 1985, Antonovics and van Tienderen 1991, Pigliucci and Kaplan 2000), but few empirical data exist to examine them directly (Travisano et al. 1995, Teotonio and Rose 2000). Few believe that a straightforward lack of genetic variance for a target trait is likely to strongly constrain a response to natural selection. The majority of traits respond to artificial selection even when there has been some prior indication of a potential constraint (e.g., Karoly and Conner 2000). Artificial selection can be a powerful way of exploring the space of possible phenotypes for more complex patterns of multiple traits (Scharloo 1983, Maynard-Smith et al. 1985, Weber 1992). It can then represent a tool for exploring the existence of potential constraints on adaptive evolution in response to a changed environment, at least in terms of short-term evolutionary responses of populations. Is the standing genetic variation in a stock population sufficient to be able to uncouple two or more traits that are genetically coupled? The modular organization of many organisms provides one framework in which to examine potential consequences of genetic correlations for evolutionary change.

Plants, as has been recognized for a long time, have a highly modular organization that promotes morphological and functional diversification during evolution (Schlichting and Pigliucci 1998). Modularity in animal development has more recently begun to receive much attention in evolutionary developmental biology (Raff 1996, Wagner and Altenberg 1996, Bolker 2000, Wagner 2001). The modularity of developing animals is also considered to facilitate the independent evolution of groups of traits belonging to different modules. On the other hand, such an organization may have led to the concerted evolution of traits within each such module. Positive genetic correlations have been widely documented for many morphological and life history traits (e.g., Falconer and Mackay 1996, Roff 1997, Donohue et al. 2000), and it is generally accepted that the patterns of genetic covariances they entail describe developmental constraints among traits that can bias or limit the independent evolution of coupled traits (Lande 1979, Cheverud 1984, 1996, Schluter 1996, Steppan et al. 2002).

The possibility that genetic and developmental coupling among traits can channel the paths taken by evolutionary change in a morphological pattern has been analyzed recently by the use of artificial selection on the forewing eyespots of the tropical butterfly, *Bicyclus anynana* (Beldade et al. 2002b). The two eyespots on the forewing of this species (and in other butterflies) are formed by the same developmental process. An eyespot consists of concentric rings of numerous epithelial scale cells, with those of a particular ring all containing the same color pigment. Each eyespot is formed in larval and pupal wing primordia around an organizing group of cells known as a focus; a series of developmental genes known from wing development...
FIG. 2. The result of artificial selection on the relative size of the anterior and posterior eyespots on the dorsal forewing of the butterfly *Bicyclus anynana*. Crosses at center show over 2000 measured females used to establish the selected lines at generation 0. The four different symbols show phenotypes of individuals taken from the 25th generation in each of the four directions of selection as indicated by arrows (replicates pooled), together with representative wings. Along the leading axis, both traits were selected in the same direction, that is, "reinforcing" the genetic and developmental coupling of the eyespots, while in the other they were selected in opposing antagonistic directions (redrawn with data from Beldade et al. [2002b]).

in *Drosophila* are expressed at different times in the focus and in the developing scale cells of each putative color ring (Brakefield et al. 1996, Brunetti et al. 2001). Furthermore, up or down artificial selection on the size of one of the eyespots yields a highly correlated response in the other, indicating that alleles influencing one eyespot tend to have comparable effects on the other (Monteiro et al. 1994). They can thus be viewed as serial units within the same morphological module.

The strong genetical and developmental coupling of eyespots in *B. anynana* led to the prediction that concerted changes will be more readily produced than opposing ones (Brakefield 1998). Beldade et al. (2002b), however, found that artificial selection over 25 generations could successfully uncouple the two dorsal forewing eyespots to produce highly divergent phenotypes in both the coupled directions and the opposing ones (Fig. 2). This flexibility of morphological evolution in the laboratory is consistent with the distribution of eyespot size phenotypes in the speciose genus, *Bicyclus*. Indeed not only was standing genetic variation within a single outcrossed laboratory stock sufficient to account for production of all phenotypes found in the genus, but also one not explored in any extant species. This suggests that the absence of this latter phenotype is explained by no history of natural selection in its favor rather than by any inability to build the phenotype. The selection experiment also included a period of relaxed selection when parents of each of a series of generations were drawn at random rather than being selected on the basis of the eyespot phenotype. Each of the lines reverted gradually during the period of relaxed selection towards control values of unselected lines with no clear distinction between the coupled and uncoupled directions of phenotypic change. This suggests that there is no strong fitness disadvantage to the uncoupled phenotypes at least in our laboratory conditions. This experiment thus illustrates the potential of artificial selection in certain species amenable to mass rearing to explore the internal processes of forming the phenotype which may, together with natural selection, play a role in shaping the patterns of ecological and evolutionary diversity found in the wild.

A related issue is whether a strong prolonged response to environmental change and selection in a particular direction then precludes a reverse response should the environment revert. In some sense, this has
occurred with respect to industrial melanism in the peppered moth in Europe and North America as melanic frequencies are now declining, apparently in response to reductions in certain air pollutants including sulfur dioxide (Grant et al. 1998, Cook et al. 1999, Brakefield and Liebert 2000). However, in that case, the phenotypic variation is primarily specified by an allelic series at a single gene and polymorphism was maintained within populations before the environmental reversal occurred. Artificial selection has recently been used with numerous replicate populations to explore this issue for life history traits in Drosophila melanogaster (Teotonio and Rose 2000). All populations studied showed reverse evolution in response to a switch in the direction of selection but the extent of this return was dependent both on the trait and on the particular history of the population.

**Exploring Trade-offs**

The evolution of traits is modulated by the interrelationships among them, and especially when these result in a fitness trade-off (e.g., Charlesworth 1990, Stearns 1992, Roff 1997, Roff and DeRose 2001, Strauss et al. 2002). Trade-offs are reviewed in detail by Fry (2003). Here, I will focus on one example that illustrates the analytical power of artificial selection. Not only can an artificial selection experiment provide robust estimates of heritability but also correlated responses observed in other traits can yield useful insights about how these traits are genetically coupled, and the consequences of such interrelationships for fitness. The idea of a trade-off has been an especially attractive notion in life history evolution where a theoretically optimal phenotype for a key trait such as size at maturity may only be produced in combination with a suboptimal phenotype in a genetically coupled trait such as developmental time; the theory predicts that successful phenotypes in natural populations reflect some compromise among various functional demands. Trade-offs can also arise through resource allocation since demands on acquired energetic and nutrient resources have to be packaged and cannot be allocated in an unlimited manner over all traits (de Jong and van Noordwijk 1992).

Artificial selection can explore the justification for specific predictions about trade-offs. One of the clearest demonstrations of a genetic trade-off involving life history traits was obtained from an artificial selection experiment in Drosophila melanogaster performed by Kraaijeveld and Godfray (1997). They selected for improved resistance against Asobara tabida. Host insects, including fruit flies, can sometimes survive the attack of an endoparasitoid by mounting a cellular immune response. Experiments on the final phenotypes obtained by the artificial selection showed that reduced larval competitive ability in unparasitized D. melanogaster is a correlated response to artificial selection for improved resistance against A. tabida. There was thus a clear trade-off among life history traits, which must be profoundly important for the population ecology of these insects in nature (Fellowes et al. 1999).

Selected lines with divergent phenotypes for a target trait may provide insights about potential patterns of genetic and phenotypic variation within a species, which can then be compared to those observed across different species. For example, although there is a clear positive relationship between adult (female) body size and egg size among species of butterflies (Garcia-Barros 2000), this is not reflected either within or across selected lines of the butterfly, B. anynana, which differ in body size (Fischer et al. 2002). The correlation observed across species may be an emergent property reflecting a wider range in phenotypes at the higher taxonomic level.

**The Evolution of Phenotypic Plasticity**

Phenotypic plasticity is an important means by which organisms may adapt to environmental heterogeneity either in space or time (Schlichting and Pigliucci 1998, Pigliucci 2001). Evolutionary ecologists have performed elegant experiments on many different plants and animals to explore how the evolution of this phenomenon is influenced by natural selection. Various types of transplant experiments have, for example, examined whether particular phenotypes have their highest relative fitness in the environment in which they are normally found rather than in a foreign environment. Examples of such a relationship are consistent with adaptive phenotypic plasticity.

Laboratory studies, again on a variety of different organisms, have also examined the genetics of phenotypic plasticity (see Scheiner 1993, Pigliucci 2001). Some of these studies have used the concept of the reaction norm as a framework for analyzing phenotypic plasticity. A norm of reaction is a plot of the relationship between the phenotype produced by individuals of a given genotype across a range of environments. If genetic variation for the slope of the relationship between phenotype and environment is present in a population then the bundle of reaction norms for a range of genotypes (or of families) from the population will show examples of crossing of the lines or, in other words, the rank order of phenotypes of the different genotypes will vary among environments. This genotype–environment interaction will enable the population to respond to natural selection.

Artificial selection experiments have yielded lines with divergent modes of phenotypic plasticity in a variety of plants and animals (Scheiner 1993, Pigliucci 2001). For example, the butterfly, Bicyclus anynana, shows phenotypic plasticity in the form of seasonal polyphenism. Two alternative seasonal phenotypes, the dry season form (DSF) and wet season form (WSF), are produced at low and high larval rearing temperatures, respectively. The phenotypes differ in having...
either no (DSF) or conspicuous (WSF) marginal eyespots on the ventral wings as well as in a suite of other morphological and life history traits (Brakefield 1997). The difference in eyespot patterns reflects natural selection in favor of crypsis in the dry season, in contrast to an active antipredator function of eyespots as deflective targets for predator attacks in the wet season. Although the phenotypic plasticity in the wild is in the form of classical seasonal polyphenism with rather discrete phenotypes (Windig et al. 1994), laboratory experiments which use several environments which span the extremes in the field show that the underlying reaction norms are continuous (the dorsal eyespots are not plastic). All eggs laid by butterflies of an unselected stock produce these alternative forms at the respective cool or warm rearing temperatures. However, artificial selection on ventral eyespot size in upward and downward directions over many generations has produced a HIGH line, which generates only the WSF across all temperatures, while a LOW line yields only the DSF (Brakefield et al. 1996).

These lines have proved useful in describing the proximate mechanisms of regulation of the plasticity via the ecdysteroid hormones (Koch et al. 1996), and of the patterns of gene expression involved in eyespot formation (Brakefield et al. 1996). The loss of the ability in these lines to produce both alternative seasonal forms appears to be due to changes in the dynamics of hormone secretion following pupation across rearing temperatures (Brakefield et al. 1998). In addition, an analysis of phenotypic variation in crosses between these lines estimates the number of effective genetic factors that are fixed for alternative alleles across them as between about five and ten (Wijngaarden and Brakefield 2000). Some of these genes are likely to be involved in specifying the timing of secretion of the ecdysteroid hormone peak that follows the pupal molt.

The pair of HIGH and LOW lines of B. anynana differ markedly in the elevation or height of the bundle of reaction norms. However, phenotypic plasticity is retained at least in the HIGH line where the WSF butterflies reared at higher temperatures have larger ventral eyespots. Two artificial selection experiments have now targeted more specifically the slope of the reaction norm of eyespot size on rearing temperature. The first experiment used truncation selection to try to obtain steeper or shallower reaction norms. Generations were reared in alternating temperatures and either selected in opposing or convergent phenotypic directions, respectively (Wijngaarden and Brakefield 2001). This did not yield a response, so the more sensitive method of sib selection was employed in combination with a more variable base population which included F2 material from a cross of the original HIGH and LOW lines (Wijngaarden et al. 2002; sib selection is a form of family selection, see Falconer and Mackay 1996). Each family in each generation was split over three rearing temperatures (high, low, and intermediate) to estimate its reaction norm. A fourth group of eggs was reared slowly at a very low temperature to provide offspring (full sibs) for use as breeding material if the family reaction norm was of a desired shape. Selection was applied for steeper reaction norms at either the lower or higher portion of the temperature range, or for a horizontal reaction norm at intermediate eyespot size (Fig. 3).

Again, though, little response to selection occurred with no novel shapes of reaction norms obtained. Thus, while extreme changes in elevation can evolve rapidly, the same is apparently not true for changes in shape. This appears to be due to positive genetic covariances or correlations across environments, which means that a response to selection in one environment is accompanied by a parallel response in other environments. Perhaps changes in reaction norm shape cannot be readily generated by development because the underlying hormones that modulate ventral eyespot size are involved in long range and multimodal signaling within the organism (Wijngaarden et al. 2002).

Some experiments using artificial selection in other species have demonstrated genetic variation in the amount of phenotypic plasticity. Scheiner and Lyman (1991) used a family selection design to study the genetics of plasticity of thorax size in Drosophila melanogaster. They selected directly for increased and decreased plasticity by rearing split families in both a high and a low temperature. They observed a response to selection but mean estimates of heritability were rather low, at around 0.06. The ecological significance of the phenotypic plasticity in this trait is unclear.

The artificial selection experiments in B. anynana suggest that, in this species, there may be genetic or physiological constraints on the response to natural selection. Selection for changes in plasticity in the wild could arise because of climate change or through extensions of species ranges into regions with a differing relationship between seasonal climatic and ecological environments. However, surveys of variation in phenotypic plasticity across species of Bicyclus, especially those to the north or south of the equator in Africa, suggest that, given sufficient time, such constraints can be broken since evolution has yielded parallel phenotypic changes in response to very different environmental gradients (Roskam and Brakefield 1996, 1999). This example illustrates how responses obtained by artificial selection in the laboratory can provide useful information for posing or analyzing questions about patterns of adaptive evolution to environmental heterogeneity.

**USE OF ARTIFICIALLY SELECTED PHENOTYPES IN FIELD EXPERIMENTS**

It could also be revealing to use material of divergent phenotypes obtained after artificial selection in field experiments to explore in more detail how natural selection influences the trait and why differences in reproductive success occur among phenotypes. For ex-
Fig. 3. An artificial selection experiment using sib selection on the slope of reaction norms for phenotypic plasticity in the size of the ventral wing eyespots in the butterfly *Bicyclus anynana*. (a) Shapes of the three population-level reaction norms that were the targets of selection (solid lines) together with that for the unselected stock (dashed line). H, SH, and HS are labels for target shape with H and S indicating horizontal or steep segments, respectively. (b) Bundles of reaction norms obtained from rearing split-families from four subpopulations. Sibs from families with reaction norms closest to the target shapes were used to set up the selected lines. Eyespot size is measured as the diameter of the fifth hindwing ventral eyespot relative to forewing length. (c) Population-level reaction norms from the four successive generations of selection illustrating the lack of any substantial response in shape; compare for same symbols with part (a) for targeted shape. Unselected stock is shown as a dashed line without symbols in generation 4. Only data for females are shown (redrawn from Figs. 1–3 in Wijngaarden et al. [2002]).

Example, the HIGH and LOW lines produced for the alternative seasonal forms in *Bicyclus anynana* could provide valuable material to explore the adaptive nature of this example of phenotypic plasticity (Brakefield et al. 1996, Brakefield and French 1999), and also such issues of whether there are various types of costs associated with the phenomenon (DeWitt et al. 1998). I am not aware that selected lines have, as yet, been used in this way but the potential is clear providing careful attention is given to fitness consequences arising as a by-product of the dynamics of the artificial selection rather than the change in the trait itself.

Experiments on natural selection in insects have successfully used releases of adults with other types of direct manipulation of phenotype (e.g., Grether 1996, Kingsolver 1996, 1999). Another approach is to use $F_2$
hybrids of populations or races that are divergent for morphological and life history traits to effectively generate a continuum of phenotypes between two extremes to set up an experiment to measure relative fitness in the wild. This powerful approach has been used successfully by Schemske and Bradshaw (1999) in an analysis of pollinator preference for different floral traits in monkey flowers (Mimulus).

One way of using divergent phenotypes yielded by artificial selection would be cohort analyses in which the survival and reproductive success of groups of individuals is monitored by mark–release–recapture techniques. This would be particularly apt where the artificial selection experiment had been performed in both upward and downward directions. An unselected stock with an intermediate phenotype could provide useful control material, perhaps in addition to native individuals characteristic of the release environment. Furthermore, F1 hybrids of replicate lines would be expected to have reduced inbreeding depression associated with any small population sizes during artificial selection. A refinement of the cohort analysis design would be to use groups of F1 hybrids between upward and downward lines in a parallel to the experiment of Schemske and Bradshaw (1999).

PERSPECTIVE

I have shown how artificial selection can provide insights about (1) the genetic variances and covariances underlying traits of ecological significance, (2) the proximate mechanisms of development—including physiology—which translate genetic variation into ecologically relevant phenotypes, and produce (3) novel phenotypes of value in examining why differences in fitness occur. For traits with high or moderate heritability, strong artificial selection can yield phenotypes that are well separated from the range within the base population within a comparatively small number of generations. A well-developed theoretical underpinning exists for the application of artificial selection. The phenotypes yielded by this technique can be useful to explore how the trait is influenced by natural selection and also to examine whether genetic and developmental mechanisms may constrain in any way the evolution of such phenotypes. Further, artificial selection can provide a powerful means of analyzing trade-offs, both from the perspective of why they arise, and of their potential consequences in the ecology of the organisms.

Biologists can be optimistic that the future integration of insights from evolutionary developmental biology (Raff 1996, Carroll et al. 2001, Arthur 2002, Wilkins 2002) with the gene mapping approaches of quantitative genetics (Mackay 2001, Mauricio 2001) will lead to a highly sophisticated understanding of how phenotypic variation, which is relevant in both evolutionary and ecological terms, is generated (Stern 2000). This process has already begun for specific morphologies in certain groups of organisms such as bristle patterns in Drosophila flies (Stern 1999, Sucena and Stern 2000), eyespots on butterfly wings (Beldade et al. 2002a, b), lateral plates and spines in stickleback fish (Peichel et al. 2001), the growth form of maize (Westerbergh and Doebley 2002), and flower morphology (Meyerowitz 1994). Although morphological traits are currently the focus of such attempts, research on how genetic and developmental mechanisms translate to ecologically relevant phenotypes will increasingly extend to include life histories and behavior (e.g., Partridge and Gems 2002). In the future, we will gain a profound knowledge of how even rather subtle changes in phenotype can be generated, for example, in the course of artificial selection. This success will mean, even more than today, that we need to study precisely how such quantitative variation is influenced by natural selection in the wild. Understanding the mechanisms of adaptive evolution at the level of interactions between ecological environments and variation in fitness among evolutionarily relevant phenotypes will become the paramount challenge.

ACKNOWLEDGMENTS

I am especially grateful to Wim Scharloo who first fired my enthusiasm for artificial selection as a powerful tool for exploring important issues in evolution and development. I thank Anurag Agrawal for his invitation to prepare this essay. I also thank him, together with Joel Kingsolver and an anonymous reviewer, for their constructive comments on the manuscript. I am most grateful to Arie van Noordwijk for making Fig. 1 available to me, to Steve Thurston for drawing Fig. 2, and to Martin Brittijn for producing Figs 1 and 3. My great debt to all those who have performed selection experiments on Bicyclus butterflies will be obvious to all readers.

LITERATURE CITED


Peichel, C. L., K. S. Nereng, K. A. Ohgl, B. L. E. Cole, P. F. Colosimo, C. A. Buerkle, D. Schluter, and D. M. Kings-
DETECTING ECOLOGICAL TRADE-OFFS USING SELECTION EXPERIMENTS

JAMES D. FRY

Department of Biology, University of Rochester, Rochester, New York 14627 USA

Abstract. Theories of the evolution of niche breadth usually depend on the assumption that genotypes that are well adapted to exploit one habitat or resource are not well adapted to others. Such “trade-offs” are often apparent in interspecific comparisons, but have been harder to document at the within-population level. Selection experiments provide a promising means for detecting within-population trade-offs: if selection for adaptation to one environment (e.g., diet, host, temperature) reproducibly lowers fitness in another, trade-offs are the likeliest explanation. Here, I describe strategies for using selection experiments to detect ecological trade-offs and discuss some of the ambiguities that can arise in interpreting the results of such experiments. I also review two sets of studies that found evidence for trade-offs using selection experiments.

Key words: adaptation; artificial selection; ecological specialization; genotype–environment interaction; negative genetic correlation; niche breadth; trade-offs.

INTRODUCTION

Theories of the evolution of niche breadth have traditionally depended on the assumption that a “jack of all trades is a master of none” (Levins 1968, MacArthur 1972, Futuyma and Moreno 1988, Rausher 1988). According to this view, no genotype can have maximal fitness in each of a set of environments (e.g., habitats or hosts), so that a population’s improvement in fitness in one environment comes at the expense of its fitness in others. I refer to the situation where no genotype has maximal fitness in each of two environments as a “trade-off” (see Fry 1996 for a somewhat more subtle definition). Trade-offs in fitness between environments can both favor restriction of niche breadth in generalist species, and prevent expansion of niche breadth in specialist species (Holt and Gaines 1992). They can also maintain genetic variation in ecologically important traits within and among populations (Levene 1953, Hedrick 1986, Gillespie and Turelli 1989).

Negative correlations between fitness in one environment and fitness in another are often observed in comparisons among species or populations, but these do not give definitive evidence for trade-offs (e.g., Kassen 2002). The traditional approach to detecting trade-offs has been to perform a “split-family” experiment, in which individuals from each of a set of full-sib, half-sib, or clonal broods are reared in two or more environments (reviewed in Fry 1996, Roff 1997). If a negative correlation is observed between a family’s fitness in one environment and its fitness in another, a trade-off is concluded to be present. Although relatively easy to perform, split-family experiments have the serious drawback that several factors can cause positive correlations to be observed, even when a trade-off is present (Rausher 1988, Jaenike 1990, Fry 1993, 1996, Shaw et al. 1995). This may help explain why trade-offs have not frequently been detected using split-family designs (see Via 1991, Shaw et al. 1995 for exceptions).

Selection experiments are a powerful but underutilized method that can potentially detect trade-offs not revealed by split-family designs. The purpose of this paper is to describe the uses and limitations of selection experiments for detecting ecological trade-offs. I will discuss practical considerations in the design and execution of selection experiments, and point out some commonly encountered pitfalls and how they can be avoided. I will also discuss two examples of sets of selection experiments that revealed ecological trade-offs. The emphasis will be on selection experiments with sexually reproducing outbreeding species. Experiments with clonal or self-fertilizing organisms impose a different set of challenges and will not be considered.

DESIGN OF SELECTION EXPERIMENTS

The goal of the experimental designs considered here is to determine whether allowing a population to adapt to one habitat, host, or environment reproducibly lowers its fitness in another habitat, host, or environment. If this is the case, then it will usually be safe to conclude that a trade-off is present, i.e., that there is no (homozygous) genotype with maximal fitness in both environments.

It is important to distinguish between two types of selection experiments. In traditional artificial selection experiments, individuals are measured for one or more traits each generation, and the investigator uses the trait
values to determine which individuals are allowed to contribute to the next generation. There is a considerable body of literature on the design and interpretation of such experiments (Hill 1971, 1972a, b, Hill and Caballero 1992, Falconer and Mackay 1996, Roff 1997). The type of selection experiment most useful for ecologists, however, is a “quasinatural” (Kassen 2002) selection experiment. In this type of experiment, differences in contribution to the next generation result from inherent differences in adaptedness to the environment; there is no need for phenotypic scoring by the investigator. An example of such an experiment would be to maintain a population of a phytophagous arthropod on a plant species on which its survival and reproduction are initially low, in order to select for genotypes with higher fitness on the host. After some number of generations, individuals descended from the selected population are compared to individuals from a control population in fitness on the initially unfavorable host. In this type of experiment, unless lifetime fitness of each individual is measured each generation, it is not possible to calculate selection intensities, realized heritabilities, and realized genetic correlations—quantities that are standard in the artificial selection literature (Falconer and Mackay 1996). Nonetheless, quasinatural selection experiments can give important qualitative evidence for or against the existence of trade-offs. This paper will emphasize experiments of the quasinatural type.

The usual quasinatural selection experiment to detect trade-offs will consist of at least two selection regimes, and two assay regimes. “Selection regime” refers to the environment or set of conditions under which a population is maintained. One of the selection regimes should be a control treatment designed to preserve the characteristics of the base population as much as possible. The experiment begins by subdividing the base population into a set of populations; one or more replicate populations is then maintained in each of the different selection regimes. “Assay regime” refers to the environment or conditions under which fitness, or one or more surrogates for fitness, is measured. At intervals (or perhaps just once), a sample of individuals descended from each replicate population is measured in each assay regime. Several decisions therefore need to be made by the investigator. These include (1) how to maintain the control populations; (2) the base population to use; (3) the number of replicate populations to maintain in each selection regime; (4) the size of the populations and the intensity of selection; (5) the fitness traits to assay and the details of the assay procedure; (6) the frequency with which to perform assays; (7) how to analyze the data; and (8) how to interpret the results. These issues will be discussed in turn.

**Control populations**

Control populations should be derived from the same base population as the selected populations, and maintained in a manner that minimizes evolutionary change. How this can best be achieved will depend in part on the nature of the base population (see Design of Selection Experiments: The base population), so only a few general comments will be made here. Ideally, a set of individuals from the base population can be cryopreserved or maintained as seeds, or in another dormant stage. If this option is not available, the best option is to maintain the control population(s) in an environment to which it is already well adapted. For example, for an experiment on the effects of adaptation to cold in *Drosophila*, the control population might be a laboratory population that has been maintained at 25°C for many generations. Alternatively, for selection experiments initiated with individuals directly collected from the wild, one could use one or more collections made from the same site at a later date as controls, but unfortunately there is no guarantee that such a repeat collection would have the same genetic characteristics as the base population.

**The base population**

One of the first choices faced by the investigator will be to decide on a base population to use for the selection experiment. For the sake of illustration, I will return to the example of an investigator who wishes to maintain a phytophagous arthropod population on a toxic or “challenging” host species (call this host B). The investigator has two choices. He or she could make a fresh collection of individuals from a wild population, perhaps from a diversity of host species, to form the base population. Alternatively, he or she could initiate the selection experiment with individuals from a long-established laboratory population, for example, one maintained on host A for 50 generations. These approaches both have advantages and disadvantages.

The newly collected base population would best represent the genetic diversity in a wild population. Unfortunately, there are two disadvantages to using such a population. First, if it is not possible to maintain a dormant or cryopreserved control, the control population is likely to adapt to its environment, and hence depart from the characteristics of the base population. Adaptation of the control can be minimized, but not eliminated, by maintaining it under conditions that are as “benign” as possible (e.g., on a nontoxic host), and by restricting the experiment to a small number of generations. Second, even if it is possible to use a cryopreserved or dormant control, laboratory adaptation in the selected population(s) could confound interpretation of the results. After multiple generations evolving in the laboratory, individuals derived from the populations maintained on host B might outperform individuals from the control populations in laboratory assays of fitness on any host, simply because they are better adapted to laboratory conditions. Such a result could mask underlying trade-offs.
A long-established laboratory population has the advantage of stability; if the population has been kept at a large size and in a constant environment for many generations, its rate of further evolution should be low. (In practice, of course, no environment is completely constant; but the environment can be held relatively constant with respect to the attribute of interest, such as temperature.) Therefore one or more replicate populations continued in the same regime can serve as controls. One might expect long-established laboratory populations to be depauperate in genetic variation, but as long as they are started with multiple founders and maintained at a large size, high levels of quantitative genetic variation should be maintained (Briscoe et al. 1992, Falconer and Mackay 1996). Many selection experiments started from large long-established laboratory populations have given strong responses (e.g., Gould 1979, Chippindale et al. 1997, Borash et al. 2000).

There are, nonetheless, plausible scenarios under which using a base population that has been maintained in a constant environment for many generations could either bias against detecting trade-offs that exist in wild populations, or create the spurious appearance of trade-offs where none exist. To understand the first possibility, imagine that most natural populations of the hypothetical arthropod segregate for two alleles, \( M^A \) and \( M^B \); on host A, the fitness ranking of these alleles is \( M^A > M^B \), while on host B the inequality is reversed. A population that has been maintained on host A for many generations would be expected to be fixed for the \( M^A \) allele; thus if a derivative of this population were transferred on host B, the \( M \) locus could not cause reduction in fitness on host A. To understand the second possibility, consider another biallelic locus (\( N \)) that shows overdominance (heterozygote superiority) on host A, but is neutral on host B. In a population maintained on host A, locus \( N \) will reach a stable polymorphic equilibrium, with allele frequencies that maximize mean fitness on A. If a derivative of the population were then transferred to host B, the \( M \) locus would no longer be under selection, and changes of allele frequencies would be free to occur due to random drift or (especially) hitchhiking with linked alleles favored on host B. Because any change in allele frequencies at the \( N \) locus would lower mean fitness on A, a spurious appearance of a trade-off would be created. A quantitative trait under stabilizing selection on host A but neutral on host B could similarly give rise to the spurious appearance of a trade-off. In addition to being prone to changes in the mean due to drift and hitchhiking after selection is relaxed, such a trait would be expected to increase in variance as linkage disequilibrium built up by stabilizing selection (Bulmer 1985, Hill and Caballero 1992) decays. This would be expected to deterministically reduce fitness on host A. (Mutations that are deleterious on host B but neutral on A could also deterministically reduce fitness on B in a population maintained on A, but would probably not do so significantly in under 100 generations.)

A solution that could minimize the above problems would be to use a base population that had been maintained on a combination of the two hosts for many generations (e.g., by alternating hosts every generation or two). Such a population would be more likely to remain segregating at the \( M \) locus than a population maintained on either host alone. If a derivative of this population were established on host B alone, allele frequencies at the \( N \) locus could still change due to drift, but hitchhiking would be much less important, because the population would already be relatively well adapted to host B.

**Number of replicate populations**

If at all possible, each selection regime (including the controls) should be replicated. Replication allows the effects of selection to be separated from those of random drift. Furthermore, crossing replicate populations within a selection regime permits one to check for inbreeding depression (see below). If replication is not possible, each population should be assayed at frequent intervals. Changes due to drift in successive time intervals are expected to be independent; therefore changes that are consistent in direction over time are likely to be due to selection.

**Population size**

The smaller the population size of each replicate population, the more limited the response to selection is likely to be, and the greater the chance that inbreeding depression will occur (Weber and Diggins 1990, Falconer and Mackay 1996, Chippindale et al. 1997). Investigators should strive to maintain population sizes of at least 50 breeding adults (cf. Weber and Diggins 1990). The need to maintain a reasonable population size will often affect the decision of how harsh the selection regime should be. Harsher regimes (e.g., host plants that are more toxic) result in stronger selection, but also increase inbreeding and genetic drift by reducing the number of individuals that successfully reproduce.

**Fitness assays**

Because the goal is to determine whether populations from different selection regimes differ in total fitness in a given assay regime, as many components of fitness as possible should be assayed. In addition, to exclude nongenetic causes of fitness differences, samples from each population should be reared in a common environment for at least one full generation (preferably two) before the assays. Even with this precaution, microbial infection (including infection of the food medium) could mimic genetic differences. The likelihood of this possibility and the precautions needed to exclude it will depend on the study organism.
Assays of populations from different selection regimes should be carried out simultaneously with proper randomization. If there are multiple replicate populations within each selection regime, comparing pairs of populations at different times may make the fitness assays more manageable (e.g., Joshi and Mueller 1996).

If populations within a given selection regime (say B) are found to have evolved lower fitness in another regime (say A), then it is a good idea to repeat the fitness assays after crossing replicate populations within regimes. If the same difference is found, inbreeding depression can be ruled out as a cause of the fitness declines (e.g., Chippindale et al. 1997).

Frequency of assays

Long intervals between fitness assays allow greater differences between populations to accumulate, but have a potential disadvantage. It is possible that populations maintained in selection regime B will initially decline in fitness in assay regime A, but that this decline will be erased in subsequent generations by the accumulation of modifier alleles that mitigate the trade-off (e.g., McKenzie and Game 1987). Relatively frequent fitness assays (say every 10–15 generations) allow a check for this possibility.

Data analysis

If all populations are assayed simultaneously, the appropriate analysis will be an analysis of variance with replicate populations (a random effect) nested within selection regime (a fixed effect). If a paired design is used, the appropriate analysis is a two-way ANOVA with selection regime crossed to population pair (a random effect); alternatively, a paired t test can be used.

Interpreting the results

If adaptation to one regime is not accompanied by a detectable reduction in fitness in a second regime, there are at least three possible interpretations. First, it is possible that a reduction in fitness occurred, but was not detected due to insufficient statistical power. Calculating a 95% confidence limit for the difference between regimes (e.g., Fry 1990, 1992) can help clarify whether a biologically important fitness difference might still be present. Second, it is possible that trade-offs exist in nature, but are not detectable under the laboratory assay regimes used. For example, a phytophagous insect population selected to survive and grow well on host B in the laboratory may suffer no reduction in survival and growth on host A. In nature, however, the two host species might have different predator faunas that select for different attributes in the herbivore. Detailed knowledge of the biology of one’s study organism might suggest appropriate hypotheses to test regarding the nature of possible trade-offs. A third possibility is that trade-offs are truly absent. It is important to consider whether alternative hypotheses for niche limitation and specialization could apply to one’s study organism (Colwell 1986, Kawecki 1994, Holt 1995, Fry 1996, Whitlock 1996, Kawecki et al. 1997, Bernays 2001).

Alternatively, adaptation to one regime may result in reduced fitness in a second regime. If the reduction in fitness is rapid and reproducible across replicates, it is not likely to be due to drift, hitchhiking, or decay of linkage disequilibrium. In this case, and if inbreeding depression can be ruled out, one can safely conclude that trade-offs are present.

Examples

I discuss here two sets of experiments whose goal was to determine whether trade-offs were present between pairs of ecologically relevant environments. The studies discussed took a diversity of approaches; although they were relatively successful, they illustrate some of the ambiguities that can arise in interpreting the results of selection experiments. I make no attempt to review the literature on ecologically motivated selection experiments; in particular, I ignore the many experiments on costs of resistance to pesticides (Roush and McKenzie 1987), on selection for early or late reproduction (Partridge and Barton 1993), and on density-dependent selection (Mueller 1997).

Trade-offs in fitness on different hosts in the phytophagous mite Tetranychus urticae

The two-spotted spider mite T. urticae Koch is a ubiquitous pest of house plants, greenhouse plants, and field crops in warmer regions. The species is an extreme generalist, having been recorded from hundreds of host species (Jeppson et al. 1975). Nonetheless, there are many hosts on which T. urticae has low reproductive success. This raises the questions of whether T. urticae populations could adapt to such hosts, and if so, whether the adaptation would involve a loss of fitness on initially more favorable hosts. A “yes” answer to both questions would give evidence that the host range of T. urticae is ultimately hindered from expanding by trade-offs.

The short generation time of the mites and the ease with which they can be reared make T. urticae a good candidate for the selection experiment approach. Gould (1979), Fry (1990, 1999), and Agrawal (2000) each established a base population of T. urticae in the laboratory on an initially favorable host, lima bean (Gould 1979, Fry 1990, 1999) or cotton (Agrawal 2000); egg-to-adult survival on these hosts was high. A second (“selected”) population was later established from the base population on a host that initially caused high juvenile mortality, cultivars of cucumber (Gould 1979, Agrawal 2000) or tomato (Fry 1990, 1999). The other population continued to be maintained on the favorable host and was used as a control. Each experimenter periodically compared viability or reproductive rate of
the selected and control populations on the host of the selected population. Adaptation of the selected population to its host occurred rapidly in each case. To test the possibility that a trade-off was present, each investigator created one or two “reversion” lines by re-establishing mites from the selected population on the control host. If alleles favored on the host of the selected population were selected against on the control host, reversion lines would be expected to decline in fitness on the host of the selected population. This was precisely what was observed. The reversion was especially rapid in the two experiments using cucumber; in fact, in Agrawal’s experiment, the reversion line had slightly lower fitness on cucumber than the control population after only five generations of reversion.

A puzzling feature of these experiments is that, when the selected and control populations were compared in oviposition rate (Gould 1979, Fry 1990) or a measure of reproductive rate (Agrawal 2000) on the host of the control population, no differences were observed. This would seem to contradict the evidence for trade-offs provided by the reversion lines. Recent results of Yano et al. (2001) suggest a possible resolution to this paradox. These authors selected a T. urticae population for higher fecundity on Commelina communis, a poor host. After four generations, the selected line had significantly higher fecundity than the control line on C. communis and several other host species. In tests of male ability to compete for mates, however, the selected line was inferior. This suggests that the reversion in the previous experiments may have been caused by selection for higher mating success, rather than by selection on female components of fitness.

In contrast to the rapid reversion observed by Gould (1979) and Agrawal (2000), the reversion lines in Fry’s (1990) experiment declined relatively slowly in fitness on tomato, and never returned to the survival level of the base population, even though one was maintained for over 30 generations. Fry (1999) considered two explanations for this result; both are based on the observation that the selected line had reached an apparent plateau or equilibrium by the time the reversion lines were established (Fry 1990). The first possibility is that the reversion lines did not contain the necessary genetic variation to revert to the survival level of the control population. Alternatively, the declines in fitness on tomato in the reversion lines may have been caused by random drift or hitchhiking rather than by trade-offs. To distinguish between these hypotheses, Fry (1999) established a “hybrid reversion line” by crossing mites from the selected and control populations. This population was necessarily polymorphic for alleles favored on tomato, and therefore should have reverted rapidly under the first hypothesis. Contrary to this prediction, after ~15 generations on bean, the hybrid reversion line had similar survival on tomato as a fresh set of hybrids between the selected and control populations. This led Fry (1999) to conclude that drift or hitchhiking might have been responsible for the original reversion.

All of these experiments had shortcomings. The selection and control populations, although large, were unreplicated. Although genetic drift or hitchhiking seem less likely than selection as an explanation for the rapid reversion of adaptation to cucumber observed by Gould (1979) and Agrawal (2000), the type of selection responsible was not demonstrated. A new set of experiments with replicate populations and measurements of more components of fitness, including male mating success, are necessary to confirm the presence of trade-offs in fitness on different hosts in T. urticae and to elucidate their nature.

Trade-offs between parasitoid resistance and competitive ability in Drosophila melanogaster

Drosophila melanogaster, like many insects, is attacked by several species of parasitic Hymenoptera. Larvae can defend themselves against parasitoids by an encapsulation response, in which cells in the haemocoel form a melanic capsule around a parasitoid egg, eventually killing it. Natural populations of D. melanogaster vary in their ability to encapsulate eggs of two common parasitoids, Asobara tabida and Leptopilina boulardi. Kraaijeveld and Godfray (1997) and Fellowes et al. (1998) used selection experiments to investigate whether trade-offs exist between encapsulation ability and fitness in the absence of parasitoids. Such trade-offs could explain why encapsulation ability in natural populations does not evolve to 100%.

Kraaijeveld and Godfray (1997) selected a susceptible D. melanogaster population for resistance to A. tabida, while Fellowes et al. (1998) selected a different population for resistance to L. boulardi. Selection was carried out by exposing larvae to heavy parasitoid attack and breeding the survivors. Four selection and control lines were established in each experiment, and selection was carried out for eight to nine generations. To minimize inbreeding, population sizes of at least 100 were used. In both experiments, encapsulation ability increased rapidly, from less than 10% to over 40%. Furthermore, in each experiment, larvae from the selected lines showed lower survival under conditions of intraspecific competition for food than larvae of the control lines. In both cases, this lower competitive ability was accompanied, and possibly caused, by a lower feeding rate (Fellowes et al. 1999). Fitness under non-competitive conditions did not differ between the selected and control lines. Kraaijeveld et al. (2001) investigated the mechanistic basis of the response in the lines selected for resistance to A. tabida, and found that larvae of the selected lines had approximately twice the density of haemocytes, a type of cell involved in encapsulation, than the controls. They hypothesized that the increased haemocyte levels may have diverted resources away from trophic functions, accounting for the lower feeding rate.
These studies provide an excellent demonstration of the power of selection experiment in documenting ecologically important trade-offs. Because D. melanogaster larvae often develop under competitive competition in the wild (Atkinson 1979), the trade-off between parasitoid resistance and competitive ability might explain why some D. melanogaster populations that are regularly attacked by parasitoids have low encapsulation ability (Kraaijeveld and Godfray 1999). This hypothesis could be explored further by documenting levels of larval competition and parasitoid attack in wild populations.

**CONCLUSION**

Selection experiments provide a powerful method for testing for trade-offs that could be important in the evolution of niche breadth and in the maintenance of genetic variation in ecologically important traits. This paper has explored only a subset of the ways in which selection experiments can be used to answer questions in evolutionary ecology; selection experiments can also be used to investigate the evolution of phenotypic plasticity (Scheiner 2002), the roles of intraspecific competition (Bolnick 2001) and environmental heterogeneity (Kassen 2002) in the evolution of niche breadth, and mechanisms of speciation (Rice and Hostert 1993). Although the selection experiment literature has a heavy bias towards Drosophila, this bias exists more for historical than for practical reasons. Selection experiments are feasible with any species that can be reared in the laboratory or greenhouse, and which has a generation time of a few weeks to a few months. Therefore, evolutionary ecologists should more often consider using selection experiments as part of their research program.

**ACKNOWLEDGMENTS**

This work was supported by NSF grant DEB-0108730. I thank A. Agrawal and two anonymous reviewers for helpful comments.

**LITERATURE CITED**


Levene, H. 1953. Genetic equilibrium when more than one ecological niche is available. American Naturalist 87:331–333.


SIZE DOESN’T MATTER: MICROBIAL SELECTION EXPERIMENTS ADDRESS ECOLOGICAL PHENOMENA

MICHAEL FELDGARDEN, DANIEL M. STOEBEL, DUSTIN BRISSON, AND DANIEL E. DYKHUIZEN

Department of Ecology and Evolution, Stony Brook University, Stony Brook, New York 11794-5245 USA

Abstract. Experimental evolution is relevant to ecology because it can connect physiology, and in particular metabolism, to questions in ecology. The investigation of the linkage between the environment and the evolution of metabolism is tractable because these experiments manipulate a very simple environment to produce predictable evolutionary outcomes. In doing so, microbial selection experiments can examine the causal elements of natural selection: how specific traits in varying environments will yield different fitnesses. Here, we review the methodology of microbial evolution experiments and address three issues that are relevant to ecologists: genotype-by-environment interactions, ecological diversification due to specialization, and negative frequency-dependent selection. First, we expect that genotype-by-environment interactions will be ubiquitous in biological systems. Second, while antagonistic pleiotropy is implicated in some cases of ecological specialization, other mechanisms also seem to be at work. Third, while negative frequency-dependent selection can maintain ecological diversity in laboratory systems, a mechanistic (biochemical) analysis of these systems suggests that negative frequency dependence may only apply within a narrow range of environments if resources are substitutable. Finally, we conclude that microbial experimental evolution needs to avail itself of molecular techniques that could enable a mechanistic understanding of ecological diversification in these simple systems.

Key words: bacteria; ecological diversity; experimental evolution; genotype by environment; negative frequency-dependent selection; trade-offs.

INTRODUCTION

A fundamental constraint of ecological patterns is organismal physiology, of which cellular metabolism is a key component. Many ecological and evolutionary phenomena are a direct consequence of variation in metabolism, including plant physiology, marine algal herbivory, the distribution and population genetics of marine invertebrates, flower phenology, the existence of geographic clines, the evolution of pathogenesis, local adaptation to abiotic stress, and range limits (Koehn et al. 1976, Place and Powers 1979, Miller and Hay 1996, Mitton 1997, Rausher et al. 1999, Day et al. 2001, Schmidt and Rand 2001, Canterbury 2002). Because changes in metabolism can alter the ecological patterns of a given species, one would like to know how the environment, a causal force of natural selection, can select changes in the metabolism of an organism. The relevance of changes in metabolism is less clear when higher levels of biological organization such as behavior, development, and structure, are the primary targets of selection.

Microbial selection experiments can investigate the linkage between the environment and the evolution of metabolism. In these experiments, a very simple environment can be manipulated to produce predictable evolutionary outcomes. Unlike other systems, selection experiments in microbes can incorporate both short-term ecological dynamics (changes in the abundance of genotypes) and longer-term evolutionary change (the rise of new genotypes by mutation) due to the short generation times and large population sizes of bacteria. Short-term experiments can examine how environmental complexity maintains diversity; in other words, what is the effect of defined variants in a specific environment? These experiments compete two variants over a short enough period of time so that no new mutations sweep through the populations (e.g., Dykhuizen and Davies 1980). Long-term experiments typically begin with no genetic variation; evolution comes from selection on de novo variation (e.g., Lenski et al. 1991). In natural populations, evolutionary change can happen on similar time scales as ecological dynamics and affect ecological phenomena (Endler 1986, Grant 1986). To ignore this change is to assume that species are static and do not evolve in response to the environment (Bohannan and Lenski 2000).

Most selection experiments can be divided into two categories: evolution-of-fit experiments and evolution-of-trait experiments. Evolution-of-fit experiments ask how fitness will change relative to an ancestor under a specific selective regime, without reference to traits or genetic differences (Travisano et al.
These experiments have yielded insights into the trajectory of fitness increases, as well as fitness tradeoffs (Travisano et al. 1995b, Travisano and Lenski 1996). One drawback of this approach is that it is difficult to determine the role that particular environmental parameters play in selecting various genetic changes, since the underlying genetic and physiological changes are not predicted a priori.

Evolution-of-trait experiments address the causes of natural selection: genetic variation and its relationship to the environment. These experiments examine which traits are selected in specific environments, and are highly amenable to hypothesis testing. Many experiments examining the evolution of traits studied selection on metabolic traits, owing to our understanding of the genetics of metabolism and quantitative theories of the interactions of metabolic enzymes (e.g., Dykhuizen and Dean 1994). These experiments can either be short term (Dykhuizen 1978, Dykhuizen and Davies 1980, Dykhuizen and Hartl 1980, Hartl and Dykhuizen 1985) or long term (Helling et al. 1987). Short-term evolution-of-trait experiments require assumptions about the evolutionary process and do not necessarily test what novel innovations might arise if the bacteria were allowed to evolve in that environment. Recent technological advances such as microarrays should allow one to determine which loci may change over time in a particular environment (Cho and Tiedje 2002), so more realistic experiments can be designed.

Here, we will review the methodology of evolution experiments using bacteria, and then address three issues on which a physiological approach to microbial experimental evolution can shed light: genotype by environment interactions, the evolution of ecological specialization, and negative frequency-dependent selection.

**METHODS**

**Batch culture**

In batch culture, nongrowing cells are added to fresh growth medium. The growth medium is a chemically defined salt solution supplemented with a source of carbon and energy. After an initial lag period, where the cells are adapting physiologically to the new conditions, the cells begin to grow exponentially at the maximum growth rate until a resource becomes limiting. Once resources become limited, the growth rate slows until growth stops and the cells enter stationary phase. The cycle is then repeated by transferring a portion of the cells to fresh medium. This changing environment potentially selects for a decreased lag period, a higher maximum growth rate, higher growth when nutrients are limited, and survival during stationary phase; however, when the environment is replenished regularly, selection for higher maximum growth rate dominates (Vasi et al. 1994). Batch culture experiments are seasonal and r-selected (Dykhuizen 1990). The multiple physiological states that each cell experiences during every cycle of batch culture make this technique undesirable for examining cellular metabolism.

**Chemostats**

Chemostats are devices that enable microbial cultures to be maintained at a predetermined growth rate in a constant, homogeneous environment (Dykhuizen 1993). A peristaltic pump feeds fresh sterile medium into the growth chamber. An overflow siphon removes spent medium and cells to maintain a constant volume. Sterile air is forced into the bottom of the culture to maintain adequate oxygen and thorough mixing in the growth chamber. The growth medium is a chemically defined salt solution supplemented with a source of carbon and energy. Concentrations of the components of the fresh medium are such that only one is exhausted by the growing culture. The limiting nutrient is usually a sugar, such as glucose or lactose. When two carbon sources are used, the complexity of the environment has effectively increased. Resources in chemostats are limiting and constant, mutations that increase growth under limited resource availability are typically favored. The constancy of the resource also means that the bacteria will not enter multiple physiological states, which can confound an understanding of metabolic processes. Chemostat experiments are not seasonal and are K-selected (Dykhuizen 1990).

**Genotype-by-Environment Interactions and Community Diversity**

Genotype-by-environment (G × E) interactions have been typically thought of in terms of complex phenotypic characters, such as growth rate, seed set, or milk production. Generally, both the genetic differences, such as different breeds of animals or different ecotypes, and the environmental differences, such as different farms, soils, or elevations, are very complex. In experimental evolution, we can study single genetic changes in simple, well-defined environments, so that the basic properties of G × E interactions can be understood. However, the “character” used in these experiments is typically fitness. Unless the translation from phenotype to fitness is a monotonic function, extrapolation back to phenotype is impossible. This may not be as much of a problem as it first seems, since in the one case which has been examined, fitness was linearly proportional to metabolic flux (Dykhuizen and Dean 1990). This system involves the competition of various alleles of the lactose operon in sugar limited chemostats and was used by both Silva (1992) and Dean (1995) to investigate the molecular basis of G × E interactions.

Dean (1995), using data from Silva (1992), compared the fitnesses of seven different lactose operons when grown on five different β-galactoside sugars and found extensive G × E interactions. Typically, the interaction
effect was as large or larger than the main genetic effect. To understand how this G × E effect arises, we need to attend to the details of the system. The lactose operon contains two genes important in the metabolism of lactose and the other four β-galactosides: (1) the lactose permease which brings the β-galactosides across the membrane and (2) the β-galactosidase which cleaves the β linkage to produce galactose and another compound. If different substrates change the rates of steps in a pathway, G × E interaction can be generated. G × E interaction can also be generated from the change of the kinetic properties of single enzymes. Some G × E interaction was due to variation in the lactose permease, but none in the β-galactosidase. Most of the G × E interaction is caused by changes in the distribution of control of flow of metabolites through the steps of the pathway; i.e., the different β-galactosides tend to bottleneck at different steps in the pathway. Thus, G × E interaction is an emergent property of the metabolic system. G × E interaction could arise either because this is simply the nature of enzymes and pathways or because different lactose operons are adapted to different β-galactosides. We can rule out adaptation as, except for lactose, these β-galactosides are creations of chemists, not seen in nature (Dean 1995). Consequently, the G × E interaction cannot have evolved and must be an inherent part of this metabolic system. We postulate that G × E interactions frequently arise as a consequence of the properties of metabolic pathways, and that G × E interactions are expected for any set of metabolic genotypes in any set of environments. Further work needs to be done to test this statement.

G × E interactions are not only created by metabolism: evolution can increase G × E interactions if selection in one environment does not increase fitness evenly across a range of environments. In particular, if selection increases fitness in one environment while leaving it unchanged or decreasing it in other environments, G × E interactions (and potentially specialization) can increase. Bennett and Lenski (1993) tested this experimentally starting from a progenitor Escherichia coli that had been selected for 2000 generations in batch culture at 37°C. They then grew six replicate lines in each of three thermal environments (32°C, 37°C, and 42°C) for 2000 generations. For the lines evolved at 32°C, there was a significant improvement in fitness at 28°C and 32°C, but no significant improvement at any other temperature tested. One out of the six lines seems to have acquired a temperature sensitive mutation that decreased its fitness above 40°C. For the strains evolved at 42°C, there was a significant increase in fitness at 40°C and above, but no significant change below 40°C. For the strains evolved at 37°C, there was a small but significant increase in fitness at 37°C, but no change at any other temperature tested. Selection at a particular temperature improved fitness for a small range of temperatures around the selected temperature. This selection did not result in fitness decreases at other temperatures, so there are no trade-offs, yet the selection generated G × E interactions by restricting the extent of the response on the temperature gradient. Adaptation to a local environment can be expected to generate more G × E interaction than is inherent in the physiology.

All genetic differences across all environmental differences need not generate G × E interactions, as some genotypes are well buffered to genetic and environmental change. Another interesting experiment suggests that well-buffered systems may be the result of natural selection (Remold and Lenski 2001). The progenitor of this experiment was isolated from a population that had evolved for 10,000 generations in batch culture at 37°C, growing in a solution of glucose and inorganic salts. Genes were randomly knocked out from this founder, so that 26 different mutant strains were created which differed from the ancestor by a single genetic difference. The fitness of each of these 26 knockout mutations was compared to the ancestor at 28°C and 37°C for growth on glucose and on maltose, giving a total of four treatments. There is a small G × E effect for temperature on maltose, none on glucose, but a major G × E effect for the comparison of fitness between glucose and maltose. For all 26 knockout lines, the fitness on glucose is not significantly different from the ancestor at either temperature, while the fitness on maltose is significantly different from the ancestor at both temperatures for almost all the mutations. Four mutations are advantageous on maltose and the rest are detrimental. Only about a half-dozen genes are involved in the uptake and cleavage of maltose to produce two glucose molecules. While it was not checked, it is unlikely that any, let alone most, of the 26 mutations tested knocked out one of these half-dozen genes. Yet these mutations affect fitness on maltose, but not on glucose, even though maltose metabolism quickly enters the same pathway as glucose metabolism. We do not yet know why glucose metabolism is highly canalized, and maltose metabolism is not. This canalization on glucose is not due to the 10,000 generations growing on glucose, since, when the knockout mutations were transferred into the strain used to start the culture that evolved for the 10,000 generations, the resulting strains still show the canalized behavior (Elena and Lenski 2001). Presumably, the canalized behavior evolved earlier in response to the centrality of glucose metabolism and is not an inherent property of the metabolic system. If the canalized behavior is an inherent property, it is expected that all bacterial species would show it irrespective of how important glucose is in their resource base.

Based on the studies reported in this section, we propose that G × E interactions initially arise as a consequence of the properties of metabolic pathways and that selection in a particular environment will increase the G × E effect beyond what is inherent in the
metabolic system. Thus, G × E interactions are likely for any set of metabolic genotypes in any set of environments. However, selection can also decrease G × E effects by selecting for canalized genotypes. Specialization and niche differentiation will be favored by both the metabolic system which initially creates fitness differences across environments and selection by increasing fitness in favored environments. If this view is correct, then to the degree metabolism drives ecological interactions, evolutionary biologists and ecologists need to explain generalization and canalization, rather than specialization.

**The Evolution of Ecological Specialization**

The extensive G × E interactions in bacteria show that, at least with respect to metabolism, organisms are likely to have maximal fitness in only parts of their niche. If, in adapting to this portion of the niche, the species loses fitness in other portions of its niche, it will become specialized. This pattern of a single mutation increasing fitness in the current habitat while lowering it in some other habitat, known as antagonistic pleiotropy (Rose 1991, Holt 1996), has been suggested as important in the specialization of many organisms and shown to be important in some. For example, when Hawthorne and Via (2001) examined two host races of the pea aphid *Acyrthosiphon pisum pisum*, they found several regions in the genome which both increased fitness on the native host, and decreased fitness on the alternate host. Antagonistic pleiotropy has also been implicated in limiting the host range of pathogenic bacteria. Numerous species of bacteria can enter a pathogenic niche only when they lose functions that would presumably be advantageous in other niches (Sokurenko et al. 1999). While these patterns have shown that antagonistic pleiotropy can promote specialization, it is not the only mechanism that can do so.

As a species begins to specialize on a subset of its niche, it may lose fitness in the unused portion of its niche due to the fixation of mutations that are neutral or nearly neutral in the current environment but detrimental in an alternate environment. The process, known as mutation accumulation (Rose 1991, Holt 1996), can occur by genetic drift, or by hitchhiking with an advantageous mutation (Maynard Smith and Haigh 1974). Bacteria that are highly specialized to their pathogenic or symbiotic lifestyles frequently have lost many physiological abilities that would be necessary to live without their host, due to a lack of effective selection to maintain those abilities (Moran and Wernegreen 2000, Ochman and Moran 2001). Finally, fitness increases in the current environment without a concordant loss of fitness in other environments can still lead to specialization (Fry 1996). Imagine that two competing species A and B have equal fitness in environments 1 and 2. If species A increases in fitness in environment 1, and species B does likewise in environment 2, then each species will begin to exclude the other species from its specialized environment even though neither has lost fitness in either environment. Hawthorne and Via (2001) found some regions of the pea aphid genome that enhance fitness on the preferred host plant, but do not affect fitness on the alternate host.

Experimental evolution studies have three distinct advantages in examining the ecology and evolution of specialization. First, the brief generation times, large population sizes, ease of replication, and ability to compete derived and ancestral populations enables direct testing of patterns of fitness change associated with specialization. Second, a detailed understanding of the changes in characters (as opposed to fitness) which cause specialization has been facilitated by the tremendous knowledge of genetics, biochemistry, and physiology in some bacterial species. Our understanding of metabolism has been particularly useful. Finally, microorganisms provide unparalleled opportunities to use genetic manipulations to directly test hypotheses about the functional basis of specialization, although these types of experiments are still rare.

The general prevalence of specialization as a direct consequence rather than a by-product of local adaptation (i.e., antagonistic pleiotropy vs. mutation accumulation or adaptation to specific habitats) is still unclear. Selection experiments in bacteria have found that antagonistic pleiotropy, while not universal, is clearly at work in some cases. Adams and coworkers (Helling et al. 1987, Rosenzweig et al. 1994, Treves et al. 1998) provide strong evidence for antagonistic pleiotropy in specialization because they characterize the phenotypic changes involved. They found that after over 700 generations growing in glucose-limited chemostats, a stable polymorphism developed in many of their *E. coli* populations. This was surprising, as there did not appear to be multiple resources to partition in the chemostat in order to allow stable coexistence. Four separate clones were isolated that coexisted in this simple environment, and three were analyzed extensively. One clone specialized on glucose, one on acetate, and one on glycerol.

The first clone was very efficient at importing glucose, but secreted by-products of glycolysis (glycerol and acetate) into the medium (Rosenzweig et al. 1994). Secreting glycerol and acetate can increase the rate of glucose uptake, so this clone is a strong candidate for antagonistic pleiotropy leading to specialization on glucose. In order to conclude that antagonistic pleiotropy, rather than mutation accumulation, is the basis for this specialization, we would need to identify candidate mutations involved in this phenotype, then place these single mutations into the ancestral genetic background and determine if a single mutation has both phenotypic effects. One of the other clones has improved its ability to utilize glycerol, and another increased its uptake efficiency on acetate. Both retained their ability to use glucose, so these organisms may be a case of special-
ization due to fitness increases in a specialized environment. They seem to be unable to compete for glucose due to the dominant clones’ increased ability to use this sugar, and so have specialized on acetate and glycerol. In the case of the acetate specialist, a candidate mutation was identified, and competition experiments showed that acetate specialists without this mutation were unable to survive (Treves et al. 1998). These experiments are utterly amazing (D. Dykhuisen, personal observation), in that a stable community evolved from a single clone in the simplest environment we can imagine.

This simple experimental system still has a great deal to tell us about the evolution of ecological specialization. Using their knowledge of biochemistry, the original authors speculated about mechanisms promoting coexistence, and confirmed the mechanism for the case of the acetate specialist. Apart from understanding the functional basis of this metabolic polymorphism, this system could be used to explore the temporal dynamics of specialization. In particular, did the glucose specialist evolve first with the other specialists descending from that clone into response to the changed environment? If so, why did the acetate and glycerol specialists have the ancestral ability to utilize glucose? Could the enhanced ability to use these metabolic byproducts have already evolved before the glucose specialist, preadapting these strains to their specialized niches? Or is it possible that they all arose simultaneously? Repeating these experiments again with dense sampling of the dynamics could provide answers to these questions.

The evolution of specialization has also been studied extensively by Cooper and coworkers (Cooper and Lenski 2000, Cooper et al. 2001a, b). They examined 12 strains of E. coli which had grown on glucose at 37°C for 20,000 generations. All 12 strains evolved to be completely unable to use ribose as a carbon source. Cooper et al. (2001b) studied this loss of function, and demonstrated that the same mutation which eliminates the ability to use ribose increases fitness on glucose, providing strong evidence for antagonistic pleiotropy in the loss of this part of the ancestral niche. However, ribose might be an unusual case since the operon is constitutively expressed in many strains of E. coli B; a genetic loss of function might be the only way to “regulate” expression (Dykhuizen and Davies 1980).

An understanding of the general causes of specialization in these lines is still elusive, however. Cooper and Lenski (2000) showed that decreasing growth rates on a variety of alternative sugars roughly coincided with increase in fitness on glucose. In the same lines, decreases in maximum growth rates at extreme temperatures (20°C, 40°C, 41°C, and 42°C) occurred around the time of increase in maximum growth rate at 37°C (Cooper et al. 2001a). While this pattern of increase in fitness in the evolved environment coupled with the decrease in function in other environments could be the result of antagonistic pleiotropy, mutation accumulation could create the same pattern due to hitchhiking. Hitchhiking occurs when a mutation that is neutral or detrimental is swept to fixation due to its linkage with an advantageous mutation. The rate of fixation of slightly deleterious mutations increases under hitchhiking (Birky and Walsh 1988). If the loss of function on alternative resources or at alternative temperatures were slightly deleterious, we could see a pattern of fitness change indistinguishable from the pattern generated by antagonistic pleiotropy. Remold and Lenski’s (2001) experiment argues that many mutations will have no effect on glucose utilization (which is well canalized), but should effect growth rate on other carbon sources (which are not as likely to be well canalized). The time to fixation of neutral mutations will decrease with hitchhiking, so it is possible that mutations that were neutral on glucose (but not neutral on other carbon sources) were fixed rapidly when the population underwent rapid adaptation to the glucose environment in the first 2000 generations. As the rate of evolution slowed, so did the time to fixation of neutral mutations. This lengthening of the time to fixation would result in a temporally slower rate of fixation, in accord with the observed decrease in the rate of loss of function during later parts of the experiment. The relative importance of antagonistic pleiotropy versus mutation accumulation in specialization will be determined in part by the difference in canalization between phenotypes exposed to selection in the current niche, and those unused there.

Our knowledge of E. coli’s metabolism might lead to a more nuanced understanding of Cooper and coworkers’ experiments. For example, there were some carbon sources on which all lines initially decreased their growth, only to have some lines at least return to the level of the ancestor. Understanding why growth is decreased for the entire experiment on some resources, while rebounding in some lines on other resources will provide insights into the basis for long-term specialization. We also await a better understanding of the prevalence of different mechanisms of specialization. Antagonistic pleiotropy has been demonstrated by Cooper et al. (2001b) in the loss of ribose metabolism as well by Rosenzweig et al. (1994) in acetate and glycerol secretion. Other mechanisms have also been observed, such as the adaptation to a specific environment in the acetate specialist studied by Treves et al. (1998). Further tests, incorporating functional analyses in conjunction with patterns of fitness change will help us understand what the most prevalent mechanisms are. We will also be able to understand how those changes confer specific fitness consequences.

**Negative Frequency Dependence**

One frequently observed phenomenon in microbial experimental evolution is negative frequency dependence, wherein genotypes are favored when rare and
**Fig. 1.** The observed fitness relations between strains TD10C and TD2 vary across the methyl-galactoside/lactulose resource axis and are also dependent on strain frequency. Strain TD2 is a derivative of the common laboratory strain of *E. coli*, and TD10C is isogenic to TD2 except for a region around the lactose operon. The operon was transduced in from a wild strain, is constitutive, and contains a permease which is 2.16 times more active that that in TD2 (Dykhuizen and Dean 1994). Fitnesses of TD10C were estimated using initial frequencies of 0–20% (rare; filled circles) or 80–100% (common; open circles) on various proportions of the sugars lactulose and methyl-galactoside. The total concentration of sugar entering the chemostat is held constant, and the percentage of methyl-galactoside refers to the percentage of the total sugar that is methyl-galactoside. Between 23% and 30.5% methyl-galactoside, the fitness of TD10C is \( w_{TD10C} = 1 \) when rare, and \( w_{TD10C} < 1 \) when common. Here, TD10C and TD2 can coexist, maintained by frequency-dependent selection in a balanced polymorphism. The error bars designate 95% confidence intervals. When TD10C is common, fitness is not linear because the high frequency of TD10C alters the concentration of resource. This figure is redrawn from Fig. 9 of Lunzer et al. (2002).

selected against when common. These types of interactions have also been implicated in maintaining species diversity in ecological communities (Molofsky et al. 2002). In several different experimental systems, diversity has been maintained by negative frequency dependence (Paquin and Adams 1983, Rosenzweig et al. 1994, Turner et al. 1996, Elena and Lenski 1997, Rozen and Lenski 2000). Two mechanisms can explain coexistence: facilitation, where one type secretes a metabolic intermediate into the environment that can be used by another type (e.g., Rosenzweig et al. 1994, Turner et al. 1996) and demographic tradeoffs, where one type has a growth or survival advantage in a changing environment (e.g., high vs. low glucose concentrations; Turner et al. 1996). Recently, a third mechanism involving antagonism either due to a toxic secreted metabolite or elimination of a metabolite required for survival has been observed (Rozen and Lenski 2000).

The repeated observation of negative frequency-dependent interactions would suggest that such interactions are ubiquitous in maintaining diversity. A mechanistic model of negative frequency dependence in chemostats with two limiting sugars suggests that frequency dependence should only occur under a narrow range of environmental parameters when resources are interchangeable (Lunzer et al. 2002; Fig. 1). In this model, the two genotypes each prefer a different sugar, although each genotype can use the alternative carbon source. While coexistence is possible, it can only occur under a very limited subset of sugar concentrations. In more variable environments, the range of parameters that can maintain phenotypic variation might be ex-
ceed (Lunzer et al. 2002). This suggests that demographic trade-offs may not maintain ecological diversity in many natural environments. The stability of cross-feeding under conditions of high environmental variability has yet to be empirically tested. An experiment that could address this would be to determine whether negative frequency dependence can evolve in a variable environment. In most microbial experimental evolution studies, the environment is a source of variation to be eliminated. By focusing on a mechanistic basis for coexistence, we are forced to confront the role of environmental heterogeneity in evolutionary processes.

**Potential Problems with Selection Experiments**

Is the artificial simplified nature of microbial experiments so unrealistic as to be misleading? We have suggested above that this might be true for the frequency dependence experiments. In this section, we discuss how some of these apparent weaknesses are either hidden strengths or opportunities to incorporate further environmental complexity. Most of the work presented here has taken great pains to eliminate environmental heterogeneity and phenotypic plasticity. While unrealistic, the advantage of simple systems is that the causal mechanisms of selection are more easily discovered. We think that a future productive avenue of research will be to add additional complexity, particularly temporal. Does the time scale on which the environment fluctuates affect niche diversification, and will those traits that evolve differ with changes in the periodicity of the environment?

A second complaint is that microbes are not relevant to most ecologists. Not only are they not multicellular, but they are asexual haploids. Setting aside the observation that the bulk of the planet’s biomass and biochemical diversity is microbial (Whitman et al. 1998), asexual haploid organisms can serve as excellent proxies for interspecies interactions. Since bacteria are haploid and easy to manipulate genetically, we can address the mechanistic bases of adaptation and diversification. For example, determining the role of pleiotropy in adaptive change is relatively easy compared to problems when using sexual diploids (e.g., Notley-McRobb et al. 2002). A third issue is the relative simplicity of the organism and system. Much of the evolution discussed in this paper results from one or few genetic changes. It is unclear to what extent these few changes are analogous to speciation in sexual diploids, although ecological divergence can happen rapidly and suddenly in multicellular eukaryotes (Thompson 1998, Hendry and Kinnison 1999).

**Conclusions**

While the concerns of experimental evolutionists may seem distant from the concerns of most ecologists, we believe that both groups are, in the end, interested in similar questions. The use of single-celled asexual prokaryotes growing in a constant, very simple, laboratory environment may be unappealing to some ecologists. However, these simplifications have enabled researchers to piece together a deep understanding of several issues of ecological importance. Experiments in these simple environments have shown that there may be simple rules for G × E interactions arising from metabolism. Likewise, these simple environments have provided us with the ability to understand the mechanistic basis of ecological specialization. On the other hand, metabolic analysis has shown that while negative frequency dependence is common in constant laboratory environments, it is unlikely to maintain diversity in variable natural environments. If microbial experimental evolution hopes to explain ecological diversity outside of the lab, it must investigate under what conditions laboratory environments provide insight into the natural world, and when they are misleading.

A variety of new molecular genetic techniques will broaden the field in several ways. First, a number of technologies will allow post-hoc exploration of evolved organisms to understand how traits have changed. For example, microarray technologies allow investigators to simultaneously monitor the expression of thousands of genes. Similarly, new techniques to locate point mutations in bacterial genomes (Sokurenko et al. 2001) will allow investigators to pinpoint the genetic basis of adaptation. The use of genetically altered microorganisms to test hypotheses about the evolution of traits will create a more convincing understanding of how traits have evolved. While we suspect that these techniques will never be employed by most ecologists, we believe that their use in understanding organisms will be appreciated by all.

**Acknowledgments**

This is contribution number 1101 in Ecology and Evolution, State University of New York at Stony Brook. The authors received partial funding for this work from NIH grants GM6073102 and GM6380001. D. Brisson and D. M. Stoebel received funding from NSF Graduate Research Fellowships. We thank J. M. Hoch, G. Hynes, S.-R. Liou, T. L. Morelli, and two anonymous reviewers for helpful comments on this document.

**Literature Cited**


coli populations during 20,000 generations in a constant environment. Evolution 55:889–896.


Abstract. Selection experiments in the wild have greatly aided our understanding of how selection shapes phenotypic diversity in natural populations. However, these experiments have been hindered by the fact that the traits most important to adaptation and speciation may be fixed within populations. Experimental hybridization offers a means to circumvent this problem since crosses can be made between populations or species that differ for the trait of interest. Here, we discuss the advantages and limitations of this approach, review results from published studies, and suggest strategies for future experiments. Advantages associated with this approach include the ability to “generate” variation for traits of interest, as well as the increased sensitivity of selection assays because of the wider range of trait values characteristic of segregating hybrids. Moreover, experimental hybridization can be extended to crosses between near-isogenic lines, allowing the effects of major quantitative trait loci (QTLs) to be dissected in a well-defined genetic background. Limitations include widespread linkage disequilibrium created by hybridization, possible cosegregation of hybrid incompatibilities, and the large number of variable traits likely to affect hybrid fitnesses. Even with these limitations, this basic approach has been remarkably successful. We now know, for example, that many morphological differences between divergent populations or species are under selection and that the selection gradients may be surprisingly large. Also, small genetic changes in pre-mating barriers may have large effects on assortative mating. With respect to discussions about the role of hybridization in evolution, experiments indicate that some hybrid trait combinations may be advantageous, particularly when the hybrid populations are placed in a new environment. Future studies seem likely to combine selection measurements with genetic mapping of the QTLs underlying the selected traits. Natural hybrid zones hold particular promise for these kinds of studies because the only experimental manipulation required is the removal of small amounts of tissue for genotyping.

Key words: adaptive trait introgression; ecological divergence; endogeneous selection; exogeneous selection; hybridization; hybrid speciation; hybrid superiority; hybrid zones; local adaptation; species barriers; tension zones; transgressive segregation.

INTRODUCTION

The strength of selection in nature is central to many topics at the interface of ecology and evolutionary genetics. Natural selection is now widely believed to be the primary force causing phenotypic diversification in nature, leading to the evolution of functionally diverse forms within a single gene pool, and ultimately to the coexistence of phenotypically diverse species that occupy different habitats, feed on different resources, or avoid predators in different ways (Endler 1986, Wilson 1989, Schelther 2000).

Because of its central role in adaptive divergence, many researchers have attempted to measure the strength of natural selection in the wild, and numerous quantitative estimates are now available for traits that vary within natural populations (Hoekstra et al. 2001, Kingsolver et al. 2001). Although these studies are valuable, it may be that the traits or quantitative trait loci (QTLs) most important to adaptation and speciation are invariant within populations (Orr 2001). Unconditionally advantageous mutations are likely to spread rapidly to fixation within populations or species and may be the cause of many or most of the fixed differences between them (Elena et al. 1996, Rieseberg and Burke 2001). Even if these important traits are genetically variable, they may have low heritability (Gustafson 1986, Falconer and Mackay 1989), result from deleterious mutations (Mackay and Langley 1990), or have negative pleiotropic effects (Brooks 2000).

A simple solution to this general problem is to create variation for the traits that differentiate populations or species. This is sometimes done by experimental manipulation of the trait of interest (e.g., Melendez-Ackerman and Campbell 1998, Hoefer and Morris 1999). An alternative approach is to hybridize the ecotype, race, subspecies, or species in question, and examine the strength of selection on segregating traits. This approach not only increases the range of trait values,
thereby increasing the variance in fitness, but it also allows the fitness effects of multiple traits to be measured simultaneously. Thus, trait interactions, phenotypic correlations, and correlational selection may be detected and measured. In addition, the recombinant nature of hybrids may dissociate suites of traits that are taxonomically correlated, allowing the fitness effects of each trait to be assessed independently so that a distinction can be made between traits that are under selection and those that are neutral taxonomic differences. Finally, transplantation of the segregating hybrids into natural environments allows ecologists to recreate ancestral genetic backgrounds similar to those when the taxa originally diverged.

Two kinds of questions may be addressed with this approach. The first relates to the role of selection in creating and maintaining population/species differences. That is, are most of the phenotypic differences between species maintained by selection? Are the trait differences that differentiate taxa more strongly selected than traits varying within contemporary populations (Hoekstra et al. 2001, Kingsolver et al. 2001)? What traits contribute most strongly to assortative mating specifically and to reproductive isolation more generally?

The second class of questions that may be addressed with this method concerns the role of hybridization in evolution. Some authors, particularly botanists, have viewed hybridization as a kind of wide recombination that provides variation upon which selection can act (e.g., Anderson 1949, Stebbins 1957, Arnold 1997). Others have argued that hybridization is a kind of “evolutionary” noise (Wagner 1972) with mostly negative consequences for diversification in groups where it is frequent (Mayr 1963). Of course, there need not be conflict between these views. It is now widely recognized that hybridization may have a variety of evolutionary consequences (Rieseberg and Wendel 1993, Arnold 1997), and a major goal of students of hybridization is to estimate the frequency of these different outcomes in nature and the conditions under which they are likely to occur (Barton 2001, Burke and Arnold 2001, Buerkle et al. 2003).

The disagreement between these two views stems in part from differing assumptions about the nature of reproductive barriers and the fitness of hybrids. If reduced hybrid fitness is caused exclusively by intrinsic genetic factors (i.e., endogenous selection), the generation of “fit” hybrid genotypes through recombination is unlikely. In contrast, if reproductive isolation is caused, at least in part, by adaptation to differing environments, a small fraction of hybrid genotypes is likely to be fitter than either parental species (Barton 2001), particularly in novel or recombinant habitats (Rieseberg et al. 1999a). In this situation, it becomes quite plausible that hybridization could contribute positively to adaptive evolution and speciation (Barton 2001), and there is increasing evidence that it has done so in nature (reviewed in Arnold 1997, Rieseberg 1997). Selection studies involving experimental hybrids contribute most directly to this debate by providing estimates of the fitnesses of hybrids, and the nature and relationship of these fitness estimates to variation in the environment.

Selection studies with hybrids may also inform us on the role of hybridization in species decline and extinction. The fate of rare and endangered species depends in part on the strength of selection against hybrids formed with more common, related taxa (Wolf et al. 2001). However, the present literature study will focus on the creative role of hybridization in evolution rather than conservation aspects.

Here, we review a representative set of studies that have experimentally manipulated hybrids to study the strength of selection in nature. Our primary focus was on experimental hybrid populations rather than natural hybrid zones, because large numbers of well-defined hybrid genotypes (F1 s, F2 s, backcrosses, etc.) can be generated by experimental hybridization. Each cross type has features that facilitate study of certain mechanisms responsible for fitness differences, while to some extent controlling for other factors (Lynch and Walsh 1998, Rundle 2002). Also included were studies of well-characterized natural hybrids that had been subjected to some sort of experimental manipulations (e.g., reciprocal transplants).

We use the data derived from this review to provide tentative answers to the questions posed earlier about the role of selection in maintaining the phenotypic differences between taxa, the relative importance of different kinds of traits in assortative mating and reproductive isolation, the relative fitnesses of hybrids, and the interactions between hybrid fitness, genetic composition and the environment. In addition, we provide guidance with respect to experimental design and make recommendations concerning how the results from these experiments should be analyzed and interpreted. Finally, we comment on the potential power of combining selection studies with genomic approaches such as the mapping of quantitative trait loci or candidate genes. We argue that both experimental hybrids and hybrid zones may serve as venues for identifying the QTLs and, ultimately, the actual genes involved in the origin of phenotypic novelty and ecological transitions in nature.

METHODS

Literature search and data compilation

Peer-reviewed journals included in the Science Citation Index were searched using a series of keywords not restricted to either animals or plants (hybridization, hybrids, hybrid zones, tension zones, local adaptation, hybrid fitness, hybrid superiority, exogeneous selection, ecological selection, species barriers, hybrid specification, reciprocal transplant), and articles identified by
this search were screened for literature citations revealing additional studies of interest. The following criteria were required for inclusion: (1) articles were based on experimental hybrids or on natural hybrids (in the latter case, hybrids were characterized either by morphological characters or molecular markers, and were subjected to experimental manipulation such as transplantation into different environments; (2) hybrids were studied in the wild or under semi-natural conditions, i.e., in common gardens, enclosures, or outdoor experimental ponds; (3) articles reported either a direct measure of selection, or indirect measures such as fitness differences between different hybrid classes or between hybrids and their parental taxa.

Our goal was to compile a representative number of high quality articles, rather than conduct an exhaustive literature survey. The most relevant data from each study were recorded into a categorical database (Appendix A, Table A1). The records were categorized by two of the coauthors independently, then the results were compared and ambiguities were resolved by all three coauthors. A second data set was assembled for all studies that reported the strength of selection on individual phenotypic traits. These studies employed an approach that involves measuring the strength of the association between individual traits and fitness. This was achieved by tracking changes in the means of phenotypic traits before and after selection, or by multiple regression analysis (Lande and Arnold 1983). Such measures of selection are typically standardized by adopting units of standard deviation for every character to allow comparisons among studies.

Data analysis

For each variable in the categorical database, the number of experiments falling into a particular category was counted. Note that experiments were counted rather than studies, since some studies contained more than one experiment. Moreover, some of the experiments fell into more than one category, e.g., they used more than one type of cross (Appendix B, Table B1).

Examination of the second dataset (selection on individual phenotypic traits) revealed one type of selection estimate that was present in sufficient number to allow statistical analyses: directional selection gradients, computed as standardized linear regression coefficients of phenotypic traits on fitness. This estimate of directional (linear) selection measures the direct effect of selection on the mean of phenotypic traits independent of correlated characters (Lande and Arnold 1983), assuming that all relevant characters have been measured.

No formal meta-analysis of selection estimates was intended, because information on within-study variance was often insufficient and selection estimates within a study often were not independent. Selection gradients were compared across different categories with one-way ANOVA and nonparametric Kruskal-Wallis tests using SPSS (SPSS, Evanston, Illinois, USA). In a second analysis, “study” was included as a random factor in ANOVAs, to account for the lack of independence caused by the limited number of studies that reported selection on individual traits. A small number of selection gradients were omitted from the analyses, because the phenotypic characters under study were fitness characters themselves, leading to misleading estimates of selection.

Results

Summary of reviewed studies

Forty-nine studies were included in the literature survey, comprising a total of 55 selection studies representing 27 different study organisms (species or species pairs). The dominance of botanical studies (31 experiments on plants vs. 24 on animals; Appendix B, Table B1) likely results from the fact that large hybrid populations are more easily generated and maintained in plants. Moreover, plant fitness is measured more readily in the field because plants are sessile.

The majority of the experiments were aimed at studying the fitness of interspecific hybrids (65%), whereas a smaller number focused on crosses between divergent populations within a species (35%). Inter- and intraspecific studies addressed very similar types of questions, although interspecific studies more often had an additional focus on the relative role of endogenous selection (intrinsic genomic factors) versus divergent natural selection on shaping the structure and fate of hybrid zones.

Eight articles were identified that reported the strength of selection on individual phenotypic characters, and all of these were on plants. Two additional studies were omitted from the analysis due to non-standardized selection estimates (Montalvo and Ellstrand 2001), or because results were reported in a graphical form only (Campbell et al. 1998). A closer examination yielded 149 estimates of directional selection gradients, 27 estimates of directional selection differentials, one stabilizing selection gradient (computed according to Lande and Arnold 1983), and measures of correlational selection for 13 phenotypic traits (computed according to Lande and Arnold 1983, Phillips and Arnold 1989).

The number of directional selection gradients was sufficient for comparing the strength of selection across different types of fitness proxies, and across different groups of phenotypic characters. Most of the selection gradients were obtained by regressing phenotypic traits on vegetative or reproductive fitness components (Appendix B, Table B2), although 12 gradients were computed using other fitness proxies. The majority of the gradients (95 estimates) were for morphological characters, followed by 33 estimates for physiological traits and 20 estimates for characters describing life history or phenology (Appendix B, Table B2).
Key features of reviewed studies

All of the reviewed studies included experimental manipulation, as hybrid genotypes were transplanted into different natural or seminatural environments. However, 23 experiments involved additional manipulations, for example the use of trained hummingbirds as pollinators to assess the amount of pollen exported and received by individual Ipomopsis hybrid plants (Campbell et al. 1998), or cage experiments with Chrysomelid beetles on two different cottonwood (Populus) host species and their hybrids, to demonstrate that hybrid zones can act as “phenological sinks” for insect herbivores (Floate et al. 1993).

It is interesting to examine which types of hybrid classes were studied most frequently (Fig. 1A). Thirty-two of the reviewed experiments employed F1 hybrids, followed by natural hybrids characterized by molecular markers or morphology (19 experiments). A much smaller number was based on second or later generation crosses. This trend is unfortunate, since the type of hybrid cross studied may have a profound impact on estimates of hybrid fitness.

A particularly interesting feature is the target of selection chosen for analysis (Fig. 1B). In total, 48 experiments measured selection at the species/population level. That is, relative fitness values were assessed at the level of entire hybrid populations and samples of the parental species. Thirteen experiments went into greater detail by comparing the fitnesses of different hybrid classes, and 14 studied selection on individual phenotypic traits. In addition, four experiments investigated selection on different cytonuclear backgrounds, or on individual alleles at quantitative trait loci (QTL) and Mendelian genetic loci (Fig. 1B; Appendix A, Table A1).

In accordance with a focus on the population/species level, most experiments measured selection by testing the significance of fitness differences (Fig. 1C). Very few studies provided estimates of directional, stabilizing, or correlational selection for phenotypic characters (Fig. 1C; Appendix A, Table A1), and only one study addressed the evolutionary response to selection on phenotypic traits (Nagy 1997). The traits most frequently studied were reproductive and somatic morphological characters, followed by phenology. Only two experiments studied selection on physiological traits such as water-use efficiency and photosynthesis (Fig. 1D; Appendix A, Table A1).

Phenotypic integration, i.e., how phenotypic characters interact to perform a given function, or to achieve a fitness advantage in the wild, was rarely considered. Only eight studies approached this topic by examining trait correlations or reporting the effect of direct selection versus indirect selection acting on correlated characters (Lande and Arnold 1983). Only four studies addressed phenotypic integration in a more extensive fashion through path analysis (Jordan 1991, Galloway and Fenster 2001), hierarchical multiple regression (Farris and Lechowicz 1990), or multivariate correlation based on principal component analysis (Lexer et al. 2003).

Niche partitioning associated with differential resource exploitation clearly was the prevalent mechanism responsible for fitness differences (Fig. 1E). This is not surprising, since many experiments were explicitly designed to test the importance of ecological selection in shaping the structure and fate of hybrid zones, to assess the significance of divergent selection in ecological speciation, or to study local adaptation in divergent populations within a species. Assortative mating (including both mate choice in animals and pollinator preference in plants) was also frequently addressed. Finally, 18 experiments tested the importance of intrinsic genetic incompatibilities, including cytonuclear incompatibilities (Fig. 1E; Appendix A, Table A1).

Comparison of hybrid fitnesses with those of the parental populations or species (Fig. 1F) revealed that hybrid fitness often was conditional on the environment the hybrids were placed in (19 experiments), on the cross type (26 experiments), or even on the fitness proxy employed (23 experiments). Only a few studies allowed straight-forward explanations for fitness differences between hybrids and their parental taxa (Fig. 1F), and eight of these 12 experiments were measured in a single environment. All but one of these 12 studies used a single hybrid class (most often F1s).

The strength of directional selection on phenotypic characters

An analysis of 149 estimates of directional (linear) selection gradients obtained from eight different studies/plant genera revealed a mean selection gradient of 0.12 ± 0.01 (mean ± 1 se), with 20 estimates exceeding 0.25. A comparison of selection gradients across the two most commonly employed fitness proxies, reproductive and vegetative fitness, revealed no significant difference in the strength of selection (one-way ANOVA; Fig. 2A). In contrast, the strength of selection did differ significantly between the three major groups of phenotypic traits studied, life history, morphology, and physiology, when differences in the means among these categories were tested by either one-way ANOVA or nonparametric Kruskal-Wallis tests ($P < 0.05$; Fig. 2B). The most frequently studied trait category, morphology (represented by 95 selection estimates), yielded smaller selection gradients than life history (20 estimates) and physiology (33 estimates; Fig. 2B). The differences were not significant, however, when “study” was included as a random factor in ANOVAs, reflecting the small number of articles (eight studies), and stressing the need for more studies that measure selection on individual traits in experimental hybrids in the wild.
Fig. 1. Frequency diagrams for six selected key features recorded from 55 different selection experiments published in 49 different studies. Each bar gives the number of experiments assigned to a particular category. The total number of experiments in each graph may exceed 55, since many of them fell into several categories. (A) Hybrid class studied. (B) Target of selection: “other” includes one study at the QTL/Mendelian locus level and three studies at the level of different cytonuclear backgrounds. (C) Selection measure: “directional,” “stabilizing,” and “correlational” include both selection gradients and differentials according to Lande and Arnold (1983). “Stabilizing” refers to selection on the trait variances, regardless of sign (stabilizing or disruptive selection). (D) Phenotypic traits as targets of selection: “reproductive” and “vegetative/somatic” were used to group morphological characters. “Other” includes two studies measuring fitness on physiological traits and one study measuring the fitness effect of genetic/environmental/geographic distance. (E) Mechanism of selection. “Assortative mating” includes mate choice in animals and pollinator preference in plants. “Genomic factors” includes genomic or hybrid incompatibility, Haldane’s rule, cytonuclear interactions, and hybrid vigor. “Other” includes metabolic factors, density, and phenology. (F) Hybrid fitness: “contingent E,” “contingent G,” and “contingent F” represent dependency on the environment, genetic background (e.g., cross type), and on the fitness proxy employed, respectively. One experiment indicated hybrid superiority.
DISCUSSION: WHAT HAVE WE LEARNED SO FAR?

Species differences

One of the chief benefits of selection studies on hybrids is that natural selection can be measured for phenotypic characters that differ between populations or species. Also, within-population trait correlations sometimes may be broken up by recombination, allowing the fitness consequences of individual traits to be measured independently. Of 149 estimates of linear selection gradients, 83 (55.7%) differed significantly from 0 at the 95% level. These results indicate that the majority of traits differentiating populations or species are under selection in the wild. In contrast, only 25% of linear selection gradients were found to be significant for traits polymorphic within populations in a recent review of 63 studies by Kingsolver et al. (2001). The greater proportion of significant tests for differences between populations makes conceptual sense because the increased phenotypic variance associated with segregating traits in hybrid populations should lead to increased variance in fitness, i.e., increased opportunity for selection.

The observation that a majority of traits that differentiate taxa are under selection does not necessarily mean that the observed trait differences were caused by selection or even that they are currently maintained by selection. It may be, for example, that selection on a given trait is in the same direction in both parental habitats and genetic backgrounds. Also, even if selection is in different directions now, it might not have been when the parental species originally diverged. Indeed, recent models of ecological specialization have shown that such fitness trade-offs are not necessary, and that habitat race formation may be driven by the accumulation of mutations that are beneficial or deleterious in one habitat, but less strongly selected or neutral in another (Fry 1996, Kawecki 1996, 1997). Of the eight studies that reported linear selection gradients for phenotypic traits in hybrids, only two measured them in both parental habitats (Jordan 1991, Nagy 1997). Native traits were favored in seven of eight comparisons made in Gilia (Nagy 1997), compared to only seven of twelve comparisons in Diodia (Jordan 1991). However, Jordan (1991) noted that selection favoring the local form was much weaker in the year the hybrid selection study was performed compared to previous years and that relaxed selection for the local population likely also extended to the local population traits. Clearly, more studies that examine selection gradients in both parental habitats are required before generalizations can be made about the role of divergent selection in maintaining the phenotypic differences between species.

Another interesting finding of our literature survey refers to the magnitude of selection estimates. On one hand, selection on individual phenotypic characters in hybrids may be strong; 20 estimates (13.4%) of standardized linear selection gradients were larger than 0.25, and seven estimates (4.7%) exceeded 0.50. On the other hand, the mean selection gradient (0.12 ± 0.01) was only half that reported for intraspecific studies (0.22) by Kingsolver et al. (2001). It is odd that differences between populations or species appear to be on average less strongly selected than intrapopulational variants. One possible explanation for this result is that sample sizes were much larger in the hybridization studies reviewed here (median N = 327) compared to intrapopulation studies (median N = 134). As pointed out by Kingsolver et al. (2001), the magnitudes of selection estimates tend to decrease as sample sizes become larger, possibly because studies with
small sample sizes (e.g., <100) are less likely to be published unless they report strong selection. The relatively large sample sizes of the hybridization studies reviewed here indicate that these experiments are much less likely to be affected by this “file-drawer effect.”

Interestingly, selection was stronger on life history/phenology and physiological characters than on morphology (Fig. 2B). This result is concordant with evidence from two other kinds of data. First, many studies have reported lower heritability for life history traits than morphological or physiological traits within natural populations (Gustafson 1986, Charlesworth 1987, Hartl and Clark 1989). According to Fisher’s (1930) fundamental theorem, traits most closely related to fitness should have the least additive genetic variance, because they will be under strong selection. A second line of evidence supporting stronger selection for life history than morphological traits comes from the QTL literature. If a trait has a history of directional selection, allelic effects should be mostly in the same direction (Orr 1998); otherwise QTLs with opposing effects should be common. A recent review of the direction of QTL effects for 572 traits from 86 studies (Rieseberg et al. 2002) revealed that life history traits have a significantly lower proportion of QTLs with opposing effects than morphological traits, indicative of stronger directional selection on the former.

In contrast to our results, Kingsolver et al. (2001) reported significantly stronger selection on morphological than life history characters. We do not have a biological explanation for this discrepancy; however, our estimates of selection gradients are based on only eight studies, all involving plants. Clearly, more studies are needed before confident conclusions may be made about the strength of selection on different categories of traits in hybridizing populations.

Reproductive isolation

The strength of reproductive barriers depends on both the extent of assortative mating and on the strength of selection against hybrids. With respect to assortative mating, a surprising result from our survey was that small genetic changes in phenotypic traits often appear to have large effects on assortative mating. This is the case for pollinator preference in plants such as Ipomopsis (Campbell et al. 1997) or Mimulus (Schemske and Bradshaw 1999). Linear selection gradients, in effect measuring the slope of regression lines of floral characters on pollinator visitation, were >0.30 in both genera. This is remarkable, since the mean selection gradient in this literature survey was only 0.12 (±0.01). Schemske and Bradshaw (1999) also report the effects of single alleles at quantitative trait loci (QTL) on pollinator visitation, and the results are striking: one allele that increased flower pigmentation reduced bee visitation by 80%, whereas an allele that increased nectar production doubled visitation by hummingbirds. Unfortunately, no comparable estimates are available for animals, since none of the animal studies measured associations between individual phenotypic characters and fitness. Nonetheless, given the simple genetic basis of mimetic color patterns in Heliconius butterflies (Jiggins and McMillan 1997) as well as the major impact color pattern changes have on assortative mating (Jiggins et al. 2001), single gene changes must have a large effect on premating reproductive isolation in this system.

It is more difficult to make generalizations about the strength of selection against hybrids in wild populations, because hybrid fitness tends to be contingent on both hybrid genotype and on the habitat into which hybrids are placed. Indeed, of 23 studies in which hybrid fitnesses were reported for different habitats, 19 (83%) reported that hybrid fitness was conditional on the habitat the hybrids were placed in (Fig. 1F). This result is exemplified by a reciprocal transplant experiment between two subspecies of big sagebrush (Artemisia tridentata; Wang et al. 1997). That study revealed strong genotype by environment interactions (G × E) across a number of fitness components including germination, growth, and reproduction. Hybrids were significantly less fit than the parental taxa in both parental habitats, but more fit in the hybrid zone.

The finding of G × E interactions in most studies that assayed hybrid fitness in multiple environments provides support for models that rely, at least in part, on environment-dependent selection to account for maintenance, structure, and stability of hybrid zones (Moore 1977, Howard 1986, Arnold 1997), although a wider range of study organisms should be examined before reliable conclusions can be drawn. Of the four studies that failed to find G × E for hybrid fitness, only in stickleback fishes were hybrids inferior in all habitats examined (Hatfield and Schluter 1999). In Phlox, no fitness differences were found between the hybrids and parentals, regardless of the habitat tested (Levin and Schmidt 1985), and two additional studies report intermediate fitness for hybrids in all habitats (Appendix A, Table A1).

There is a widespread view that intrinsic isolating factors are more stable, predictable, and irreversible than extrinsic reproductive barriers (Futuyma 1989). However, results from natural hybridization experiments suggest that this view may be oversimplified. Although the fitness of hybrids between parents isolated by extrinsic barriers often was contingent on the environment, a similar pattern of contingent fitness effects was observed for hybrids with intrinsic genetic incompatibilities. Indeed, the relative fitness rank of hybrids under endogenous selection often varied depending on the sex of the hybrid (McMillan et al. 1997), the direction of the cross (Galloway and Fenster 2001), the hybrid genotype or genotypic class (Emms and Arnold 1997), and even the environment (Campbell and Waser 2001). Finally, there is increasing evidence that hybrid incompatibilities may be easily and quickly
purged in both experimental hybrid populations (Grant 1966, Rieseberg et al. 1996) and natural hybrid zones (Butlin 1998, Carney et al. 2000). Thus, in our view, both the stability of postzygotic reproductive barriers and the detrimental consequences of outbreeding depression are overemphasized in the literature.

Hybrid fitness and conservation issues

Although conservation aspects were not the main focus of this study, the reviewed literature may inform us on the possible role of hybridization in the decline and extinction of rare species. Fig. 1F reveals that fitness of hybrid offspring is almost never universally lower than fitness of the parental species in all habitats. Rather, hybrid fitness often appears to be contingent on the environment, as mentioned earlier. This suggests that the risk of genetic assimilation may depend to a large degree on the spatial and temporal structure of the hybridizing taxa’s habitats (e.g., patterns of disturbance; Anderson 1948). Hybrids may be just as fit as the parents in certain habitats, but less fit in others, and patterns of habitat differentiation may ultimately dictate whether stable hybrid zones are formed, or whether genetic assimilation of one or both parental taxa is more likely (Wolf et al. 2001). Clearly, assessing the extinction risk imposed by hybridization should include data on the fitnesses of hybrids and their parental taxa in different habitats.

The evolutionary role of hybridization

As alluded to in the introduction, the role of hybridization in adaptive evolution and speciation remains controversial (Anderson 1949, Wagner 1972, Arnold 1997). Selection studies with experimental hybrids in the wild may shed new light on this controversy. If fitness is contingent on both hybrid genotype and the environment, then it becomes plausible that variation generated by hybridization might contribute to adaptation. On the other hand, if all early generation hybrids are less fit than both parental species, then the adaptive valley may be too deep and broad for hybrids to recover from. Even a very strong barrier, however, cannot prevent the introgression of advantageous alleles (Barton and Hewitt 1985, Pialek and Barton 1997).

Of 80 reports of hybrid fitness, in only two instances were hybrids consistently less fit than both parental species. Rather, relative hybrid fitness was typically contingent on the fitness proxy employed (23 studies), genetic factors (26 studies), or the environment (19 studies). These results indicate that hybridization could play a role in adaptive evolution, but certainly do not prove that it has done so.

Only one study reported a general trend of hybrids having higher fitness than their parents, and the fitness character used was the quality of pollinator visits in *Ipomopsis* plants, a product of several fitness components related to pollen transfer (Campbell et al. 1998). However, fitness was only measured in one environment, and hybrid superiority was not detected when the quantity of pollinator visits was considered instead (Campbell et al. 1997).

A general impediment to understanding the evolutionary role of hybrids is the fact that most studies treated “hybrids” as some kind of amorphous mass, combining all hybrid genotypes of a given cross into one unit (population/species) or a few units (hybrid class; Fig. 1B). Only 14 experiments actually examined individuals by regressing phenotypic traits on fitness (Fig. 1B), but the differences in phenotypic values were generally not interpreted in terms of genotypic differences. Clearly, the notion that hybrid fitness may best be evaluated at the level of individual hybrid genotypes has largely been overlooked, the only available examples being the studies by Hotz and Semlitsch (2000) in animals, and Lexer et al. (2003) in plants. The latter study clearly demonstrates that evolutionary novelties such as rapid adaptation to a saline habitat can be readily achieved by the large genotypic variation underlying the relevant traits in segregating hybrid populations.

The striking paucity of studies on the genotypic level may stem from the fact that the importance of associations between individual genotypes and habitats, although first discussed more than 50 years ago (Anderson 1948), has been revisited only very recently (Johnston et al. 2001). Also, the genetic basis for superiority or extreme phenotypic values in a small number of genotypes within larger segregating populations is only beginning to be understood (Rieseberg et al. 1999a). This phenomenon is known as “transgressive segregation,” and is thought to contribute to rapid adaptive shifts in hybrid lineages, potentially leading to the origin of new, diploid, hybrid species (Rieseberg et al. 1999a; also discussed by Welch and Rieseberg 2002). One of the studies reviewed here (Lexer et al. 2003) examined the evolutionary role of transgressive segregation by examining selection on transgressive characters in sunflower hybrids (*Helianthus annuus* x *H. petiolaris*) transplanted into a novel hybrid habitat. As predicted, many of the transgressive traits detected in the segregating hybrids were favored by selection.

Hybridization can also contribute to adaptation through the introgression of advantageous alleles across hybrid zones. Theoretical models indicate that adaptive trait introgression may be a frequent outcome of hybridization, because favorable alleles may move across even very strong reproductive barriers (Pialek and Barton 1997, Buerkle et al. 2003). Unfortunately, this outcome of hybridization is difficult to document empirically because advantageous alleles should very rapidly spread to fixation and thus are difficult to “catch in the act.” However, the fact that many hybrid zones have formed recently as a result of human disturbance, suggest that advantageous trait introgression may be detectable. Indeed, Rieseberg et al. (1999b)
found several linkage blocks that appeared to be positively selected in wild sunflower hybrid zones. Another possible example comes from a study of hybridization between the Golden-collared and White-collared Manakin (Brumfield 2001). Plumage characteristics of the former taxon had moved further across a recently formed hybrid zone than either mitochondrial or nuclear markers, possibly as a result of sexual selection.

Limitations

Although the use of hybrids in selection studies has been successful, conclusions drawn from them are limited by problems with experimental design. Outbreeding depression may alter the relationships between phenotypes and fitness by reducing the mean fitness of hybrid offspring, especially for experiments that employ wide crosses. These difficulties can often be overcome by appropriate statistical means, as described in more detail below.

Perhaps a more important finding is that many studies to date have used F₁ hybrids or natural hybrids of unknown ancestry (Fig. 1A). Because F₂s tend to be heterotic, and outbreeding depression often is not expressed until the F₂ or later generations, conclusions about the relative fitness of hybrids must be tentative. Individual, well-characterized, hybrid genotypes from backcrosses, F₂, or later generations would be much more informative, as they allow us to evaluate ecological selection in the wild, while to some extent taking intrinsic genomic factors into account (Rundle 2002).

Also, in most studies selection was studied at the population/species level (Fig. 1B) by measuring fitness differences between groups (Fig. 1C). This approach proved to be successful in addressing fundamental questions of hybrid-zone dynamics, ecological speciation, or the evolution of pre- and postzygotic barriers. However, it does not provide the data required to link fitness differences to changes in phenotypic characters, or interactions among characters (Lande and Arnold 1983, Phillips and Arnold 1989), not to speak of the underlying QTLs or candidate genes.

Further, those studies that measured phenotypic characters mostly focused on morphological traits (Fig. 1D), whereas directional selection was clearly stronger on physiological traits and characters describing life history or phenology (Fig. 2B). Also, trait interactions and phenotypic integration were very rarely examined. As a result, for most study systems, data are lacking on the relationships between fitness and phenotypic traits, and on how the complex web of trait interactions actually works to produce the phenotype observed in the wild. This information is becoming increasingly important as selection studies with hybrids in the wild are discovered as a means of identifying the QTLs or even candidate genes involved in species differences (Schemske and Bradshaw 1999; discussed by Lexer et al. 2003).

Finally, selection measurements on hybrids can be seriously confounded by linkage disequilibrium induced by hybridization. Such associations among alleles at different loci can be due to population admixture, selection for particular parental gene combinations in hybrids (epistatic selection), selection against heterozygotes, or physical linkage in the genome (Gardner et al. 2000). Disequilibria can also be induced by chromosomal abnormalities segregating in interspecific backcross or F₂ hybrid populations, because different types of chromosomal rearrangements are known to suppress recombination (Rieseberg 2001). This can be a problem when the targets of selection under study are phenotypic characters or their underlying QTLs, because in either case linkage disequilibrium may result in extensive intercorrelations.

Several solutions to this problem have been suggested, such as the use of selection coefficients that are not affected by selection on other correlated characters (Lande and Arnold 1983), or appropriate means of data reduction such as summing or averaging of highly correlated characters or extraction of composite variables such as principal components (Lande and Arnold 1983, Mitchell-Olds and Shaw 1987). Moreover, the rearranged parts of the genome, and even the precise locations of chromosomal breakpoints, can be identified by suitable genetic mapping techniques (Livingstone et al. 2000). This should provide a better feel for the severity of intercorrelations in different parts of the genome (Rieseberg 2001). Also, in some situations, linkage disequilibrium due to hybridization may actually represent an advantage, e.g., if selection is studied on the genotypic level in nonmodel organisms. In such cases, the number of available genetic markers is often limited, making it difficult to collect genotypic data for the entire genome. Experimental hybridization may increase the chance to detect selection, since linkage disequilibrium may effectively widen the region of a chromosome that can be tested for selection by a given set of markers.

Outlook

An important future task will be to combine selection studies with molecular genetic mapping, as demonstrated by Schemske and Bradshaw (1999) in the present literature survey. The basic approach includes genetic map construction in segregating hybrid populations, making use of the high levels of marker polymorphism present in crosses between divergent populations or species. The QTLs underlying important phenotypic traits can then be mapped to the genome (reviewed by Mauricio 2001), and the fitness of different QTL genotypes in the wild can be tracked with the help of flanking molecular markers. Recent advances in adapting genomic methods to ecology and evolution (Gibson 2002) also raise the hope that candidate genes underlying important QTLs can be mapped for many species (Mauricio 2001). Ultimately,
this approach may yield selection differentials (Hartl and Clark 1989) for single alleles at exactly those genes that are responsible for species differences, or even for the origin of entirely new species.

When adapting genetic mapping methods to selection experiments, ecologists can build on a huge variety of experimental designs originally developed for crop species. 

\[ F_2 \] populations are ideal because they contain alleles segregating from both parents, i.e., all traits can be analyzed at the same time. However, backcrosses are better for detecting \( G \times E \) interactions (Rundle 2002). Moreover, the relatively simple segregation patterns of molecular markers in backcross populations facilitate the detection of epistasis, i.e., interactions among genetic loci (e.g., Kim and Rieseberg 2001). Recombinant inbred lines (RILs) may be even better suited for detecting \( G \times E \) interactions, because RILs have essentially no within-line genetic variance. Therefore, the same lines can be raised over many different sets of environments, much the same way lineages can be replicated in clonal species. Near isogenic lines (NILs) contain only small segments of donor genome and are therefore ideal for removing the confounding effects of other traits segregating in the population (Lynch and Walsh 1998). This should facilitate studies of multi-generation selection, i.e., the tracking of changes in allele frequencies at single QTLs over multiple generations in the field.

Some of the most challenging questions to be addressed with experimental hybrids relate to the repeatability of evolution. Although ecologists have long been aware of the importance of replication—e.g., blocks, sites per habitat, clones, or repetition over time—one type of replication has rarely been considered in selection studies involving natural populations: the type that reveals the repeatability of evolutionary change caused by selection. This may include finding the interspecific QTL combinations that confer the greatest fitness advantage in experimental hybrids in the wild, and comparing these QTL combinations with those conferring fitness differences in natural hybrid zones. In addition, the selective values of individual QTLs may be determined directly in hybrid zones and compared to replicate zones, using approaches similar to those described by Rieseberg et al. (1999b) and Rieseberg and Buerkle (2002). Experiments of this type may reveal, for instance, whether a similar selective regime prevails in experimental hybrids and across a set of hybrid zones, thereby bridging the gap between phenotypic distributions and segregation patterns in experimental hybrids, and the evolution of phenotypic novelty in nature. Replicated long-term selection studies, providing information about the repeatability of the evolutionary response to selection at the QTL level across experimental populations/environments, would be even more desirable. Finally, selection studies on experimental hybrids in the wild may serve as a tool for testing the potential role of candidate genes in producing the wealth of phenotypic diversity we see, as well as major ecological transitions in animals and plants.

**Acknowledgments**

We are thankful to John M. Burke for valuable discussions, and to Jeffrey Conner and an anonymous reviewer for insightful comments on an earlier version of the manuscript. This work was supported by Erwin-Schroedinger grant J 1960-Bio of the Austrian Science Foundation to C. Lexer, and by USDA award 2001-35320-10642 as well as NIH award GM59065 to L. H. Rieseberg.

**Literature Cited**


APPENDIX A

A table showing key features of the reviewed studies is available in ESA’s Electronic Data Archive: Ecological Archives E084-041-A1.

APPENDIX B

Tables presenting a summary of the reviewed studies are available in ESA’s Electronic Data Archive: Ecological Archives E084-041-A2.
SELECTION IN A MODEL SYSTEM: ECOLOGICAL GENETICS OF FLOWERING TIME IN ARABIDOPSIS THALIANA

MASSIMO PIGLIUCCI

Department of Botany and Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996-1100 USA

Abstract. Arabidopsis thaliana and some of its close allies have been a model system for genetics, developmental biology, and molecular biology for some time. More recently, they have been adopted by an increasing number of laboratories involved in evolutionary ecological research. In this paper, I illustrate some of the methods and advantages concerning the use of Arabidopsis to study selection and the constraints imposed on it by the genetic architecture underlying morphological and life history traits. Populations of A. thaliana and closely related species show a wider ecological variance than had been suspected, and it is increasingly clear that even such a relatively simple organism presents endless challenges to ecologists and evolutionary biologists. The study of the evolution of life history traits in this group also provides us with an invaluable opportunity to advance our search for ways to integrate biological knowledge at the organismal and molecular levels. At the same time, these efforts also yield a better understanding of the type of research that can be carried out independently at these two levels of the biological hierarchy.

Key words: Arabidopsis; genetic architecture; life history evolution; model organisms; selection.

INTRODUCTION: AN INTERESTING WEEDE

Arabidopsis thaliana and some of its wild relatives have been used as a model system in genetic and developmental studies for decades (Langridge 1957, Ratcliffe 1965, Griffing and Scholl 1991, Pyke 1994, Kunkel 1996, Anderson and Roberts 1998). More recently, these species have attracted an increasing amount of attention by ecologists and evolutionary biologists (Pigliucci 1998, Alonso-Blanco and Koornneef 2000, Mitchell-Olds 2001), and A. thaliana is rapidly on its way to become as much a household name among evolutionary ecologists as Drosophila has been throughout the 20th century.

The rapid career of A. thaliana and other familiar experimental systems, of course, is not to say that there are no inherent problems with the use of model organisms in biological research (Kellogg and Shaffer 1993), and especially in organismal biology, where a major goal is to be able to appreciate the variety of strategies that living organisms use to cope with their environment. The characteristics that make for good model organisms, such as a fast life cycle, also tend to mark those species as somewhat exceptional from the point of view of their evolutionary ecology. Exceptional, however, is a relative term, and the tantalizing prospect presented by an initial focus on model organisms is that ideas and tools that are developed by studying them can then readily be extended to their close relatives and beyond.

Despite this cautionary note, let us begin by considering the basics of Arabidopsis thaliana’s organismal biology. This species is a highly selfing, ruderal, opportunistic weed characterized by a variable pattern of life history. The observation has often been advanced that the field A. thaliana behaves as either a spring annual, with plants that overwinter as seeds, germinate early in the spring, and immediately go to flower, or as a winter annual, with plants that germinate in late summer or early fall, overwinter as rosettes, and flower in the spring (Napp-Zinn 1985, Thompson 1994). However, no clear geographical pattern can be established for the two life histories, and instances of populations with both early- and late-flowering genotypes are known (Nordborg and Bergelson 1999, Donohue 2002). The possibility of a dual life history makes A. thaliana ecologically interesting because selection leading to the two modalities is not well understood (Silvertown and Lovett-Doust 1993). Furthermore, there is a high degree of phenotypic plasticity in flowering time itself, in response to environmental cues such as temperature, photoperiod, or nutrient availability (Pigliucci and Schlichting 1998, Blazquez and Weigel 2000, Johanson et al. 2000, Sheldon et al. 2000, Pigliucci and Marlow 2001), which makes this trait an ideal focus for selection analyses. At the structural level, molecular biologists and physiologists often refer to the two types of A. thaliana as “early” and “late” flowering, a dichotomy that probably corresponds to the distinction between spring and winter annuals. However, as I shall show, the connection between our knowledge of molecular and organismal biology in this plant is more complex than one might have imagined,
and may present us with a paradigm of the advantages and pitfalls of integrative biology in general.

In this paper, I will outline the questions raised and methods employed to study selection in Arabidopsis as a model system. I will concentrate on flowering time as a key trait because of its potential ecological role, its tight relationship with many other aspects of the plant's morphology and life history, and the fact that it has been the target of different and at times innovative approaches to the study of the ecological genetics of this species. Research conducted just within my laboratory on this topic has ranged from the molecular basis of phenotypically plastic responses (Callahan et al. 1999, Pigliucci and Schmitt 1999), to their effect on reproductive fitness (Pigliucci and Schlichting 1996), and their evolution within a phylogenetic context (Pigliucci et al. 1999, Pollard et al. 2001). This scope is made possible only by the use of a model system. Many other investigators have contributed to a complex emerging picture of partially adaptive shade-avoidance responses in this species that center on alteration of the flowering schedule. These responses are mediated by several photoreceptors and auxiliary gene products, and are clearly constrained in their efficacy by both the genetic architecture underlying flowering time and the actual efficacy of this sort of phenotypic plasticity within natural ecological settings (e.g., Kowalski et al. 1994, Mitchell-Olds 1996, Kuittinen et al. 1997, Alonso-Blanco et al. 1998a, Stratton 1998, Lascève et al. 1999, Dorn et al. 2000, Munir et al. 2001). In the next sections, I will first review how phenotypic and genotypic selection are measured, and then how such techniques have successfully been applied to the specific biology of A. thaliana. The third part of the paper will discuss how to study the constraints imposed on selection by the genetic architecture of the traits of interest. I will then conclude with a discussion of the advantages and limitations of integrating molecular and organismal biology in a model system. In the Appendix, I present a brief "how-to" guide to Arabidopsis thaliana ecological genetics.

Measuring Selection in the Field and Under Controlled Conditions

Natural selection has been one of the holy grails of evolutionary biology ever since Darwin’s publication of On the Origin of Species, and the measurement of its mode and intensity is crucial to evolutionary theory. Surprisingly, while methods to estimate the magnitude of natural selection have been practiced for a long time (Endler 1986), it has been only in the last few decades that we have developed approaches that allow a direct link between empirical measures of selection and the theoretical quantities characteristic of evolutionary quantitative genetics. Since ecologists tend to be more familiar with the classical approaches rather than those framed within the context of quantitative genetics theory, it will be useful here to briefly summarize the standard modus operandi and the limitations of measuring phenotypic and genotypic selection, in order to better understand how research on selection in Arabidopsis can be conducted.

In a landmark paper, Lande and Arnold (1983) started out from Fisher’s (1930) suggestion that phenotypic selection on a given trait can be thought of as the covariance between fitness and that trait. This leads to measuring phenotypic selection on several traits by simply carrying out a multiple regression analysis using an estimate of fitness as the dependent variable and the individual traits as independent variables (this, of course, assumes that one can actually come up with a reasonable measure of fitness, often not a trivial task in itself; de Jong 1994). Lande and Arnold then showed that the magnitude of the linear and quadratic coefficients in such regression analysis can be translated into quantities used by theoretical quantitative geneticists to describe phenotypic evolution. With some caveats, linear coefficients indicate directional selection, while quadratic coefficients indicate the operation of stabilizing (when they are negative in sign) or disruptive (when they are positive) selection. Given that one uses a multiple regression analysis, the effects of several “independent” traits on each other is also accounted for.

There are several limitations to this approach, and I will briefly discuss a few because they have a major impact on how we actually carry out and interpret the results of selection analyses in any system, including Arabidopsis. First, as Lande and Arnold immediately recognized, it is always possible that the results of such analyses will be biased by the fact that an important independent variable has not been factored in. For example, one might measure apparent selection on branch production, but this in fact could be the indirect result of selection for a generalized increase in the size of the plants. This is a rather unavoidable problem, the likelihood of which can only be minimized by augmenting our knowledge of the biology of the system, so that one can measure as many biologically significant traits as possible. The practical upshot of this problem is that, at most, we can say that certain traits or their correlates are under selection; we can never conclude with certainty that the traits we measured directly are the ones being selected.

Rausher (1992) focused on another problem with the original Lande-Arnold framework, which had been raised by other authors before (e.g., Mitchell-Olds and Shaw 1987, Wade and Kalisz 1990): if the analyses are conducted on phenotypic data, i.e., data gathered by measuring individual plants, then environmental factors might dramatically influence the results. For example, suppose that, unbeknownst to the investigator, some of the plants are growing on a patch of soil with unusually high levels of nutrients. As a consequence, all plants might be taller and produce more biomass, thereby increasing their reproductive fitness.
One would then measure selection on height and biomass, but this apparent selection would be caused by phenotypic plasticity in response to the local environmental conditions, and may or may not have a genetic basis. To correct for this problem, Rausher suggested carrying out the regression analysis using the breeding values, i.e., the character means across replicates of the same genotype, instead of calculating simply the raw phenotypic value. This ameliorates the problem, except for the possibility of unaccounted for genotype–environment interactions, since the environment may not have the same effect on each genotype; however, Rausher’s solution also creates the new obstacle of a much reduced sample size left to the investigator to play with, because one needs many raw phenotypic measurements in order to derive a single breeding value; this leads obviously to a reduced statistical power of the analysis. There is no simple solution to the problem, and each researcher faces a trade-off between augmenting statistical power and at least partially accounting for environmental effects.

A third major class of problems raised by the Lande-Arnold approach has also been pointed out by several researchers: multiple-regression analyses tend to be a poor means to represent the often complex and hierarchical relationships among traits (Crespi and Bookstein 1989, Kingsolver and Schemske 1991, Sinervo and DeNardo 1996, Scheiner et al. 2000). The solution often proposed is to use path analysis instead. Path analysis is an extension of multiple-regression and multivariate techniques such as principal factor analysis. It allows the researcher to build a series of alternative models of the hypothesized causal structure underlying the observed correlations among traits as well as their relationship with fitness. These models can then be confronted with the data (Hilborn and Mangel 1997) and discarded or improved based on how well they fit the available evidence. This is a very powerful technique, which is still largely underutilized in biology. However, it does have its drawbacks, the chief one being that for any realistic situation there is a large number of models that could be fit to the data, which makes it very difficult to carry out exhaustive analyses. This means that the individual investigator has to play a much more direct, inevitably subjective, role in the analysis of data. However, one could argue that doing statistical analysis in a more involved and less mechanical way than usual is not a problem at all, but on the contrary provides benefits to the scientific enterprise.

An additional problem with the Lande-Arnold approach for measuring selection is that it relies on parametric statistics, which assumes certain properties in the distribution of the data. When these assumptions are not met, alternative nonparametric methods of analysis have been proposed (Schluter and Nychka 1994). Of course, these have their own problems, including the fact that the resulting output cannot be plugged into the standard quantitative genetic equations describing phenotypic evolution, a major advantage of the original Lande-Arnold method.

There is one more major challenge involving studies of selection that has been rarely discussed in the literature, but concerning which A. thaliana and a few other model systems may be able to play a decisive role. This is the suggestion that certain statistical approaches, such as some classes of path analytic and structural equation models, can directly test causal hypotheses—in spite of the “correlation is not causation” mantra we often hear repeated rather mechanically—through the use of concepts such as d-separation (Shipley 2000). The definition of d-separation is the set of necessary and sufficient conditions for two vertices of a path diagram or other kind of graph intended to represent causal relationships to be observationally independent, conditional on some other set of vertices. “Conditioning” on a vertex means to alter the value of that variable, either experimentally or by means of statistical controls. It is a property of these graphs that d-separation in a causal graph must be mirrored by an equivalent statistical independency in the observational data if the causal model is correct. In other words, causal graphs and d-separation allow us to make a direct link between specific hypotheses about causal independence of factors and observations of statistical relationship between variables. We need to be clear on what this means: even a combination of structural equation modeling and d-separation cannot tell us if a particular model of causal relationship between phenotypic traits and fitness is true; however, and crucially, this technique can eliminate models that are false on the basis that their predictions do not match closely enough the observed correlation structure among the data. I cannot possibly do justice here to this ongoing discussion on the possibility of testing causal models statistically, but Shipley’s (2000) book is a must read for anybody seriously interested in causal modeling in biology, and the first papers applying these techniques have already begun to appear (Magwene 2001).

The role that Arabidopsis and other model systems can play in all of this is in the rare possibility of actually moving, at least in theory, from the level of statistical modeling to that of direct causal links by way of our understanding of the molecular and developmental biology of these species, as discussed below. The idea here is that portions of causal models of, for example, the genetic architecture of different traits, could be tested not only statistically by way of path analyses or d-separation, but also by direct knowledge of the molecular genetics of the organisms in question. For example, if a statistical model predicts a causal connection between, say, leaf production and flowering time, this connection has to be reflected in a common genetic-developmental basis underlying the expression of the two traits. For most organisms, the latter prediction will
go untested, but in Arabidopsis it can be investigated by means of available molecular genetic techniques.

**STUDYING SELECTION IN ARABIDOPSIS**

It is now time to turn our attention to actual examples of research focusing on estimating selection on Arabidopsis’ phenotype to get a more concrete idea of how the above-mentioned considerations actually translate into practice. Two important pieces of information are crucial when one is interested in studying selection from an evolutionary standpoint: on the one hand, one needs a good understanding of the multivariate structure of selection under field and laboratory conditions. On the other hand, we must have estimates of the genetic constraints affecting the actual evolutionary response. This section will deal with the first issue, while the next one will delve into the second one.

Some of the approaches to study phenotypic selection outlined in the previous section have been applied to studies of selection in A. thaliana (Pigliucci et al. 1995, Mauricio and Rausher 1997, Mauricio et al. 1997, Scheiner et al. 2000, Callahan and Pigliucci 2002, Donohue 2002). As an example, Callahan and Pigliucci (2002) have investigated selection on flowering time and on a genetically correlated character, leaf production, in natural populations of A. thaliana grown both in the field and under controlled conditions. The typical outcome from selection analyses in the field, which were conducted on two populations during two consecutive years, is presented in Fig. 1: both using phenotypic and genotypic data, the latter to meet Rausher’s objection to the standard Lande-Arnold approach, we found evidence of strong directional selection favoring early flowering in our populations, consistent across years and sites. This was accompanied by strong directional selection for an increase in leaf production, also consistent across sites and years. As I shall show below, because of the genetic correlation underlying the two traits, populations of A. thaliana cannot move in the area of phenotypic space favored by selection, i.e., a combination of early flowering and increased leaf production. Depending on the relative intensity of the selection coefficients on the two traits and on the magnitude of the genetic correlation itself, these populations are stuck in some middle-ground position that balances the impetus of selection and the friction of the genetic constraint.

Because of our and others’ previous research on flowering time, Callahan and I hypothesized that the causative agent behind directional selection for early flowering was competition for sunlight (Halliday et al. 1994, Dorn et al. 2000). A. thaliana has a sophisticated mechanism that presides over its light perception capability, and the so-called “shade avoidance” component is controlled in part by one of its five phytochromes, molecules that are sensitive to the red and far red portions of the solar spectrum. We compared the field data with suitable controlled-condition experiments to test the idea that phytochrome-mediated shade avoidance was responsible for the observed directional selection. While we did observe selection for early flowering in both sets of experiments, it was also clear that the strength of selection was as strong under shade as it was under more open conditions (Callahan and Pigliucci 2002), rather than being stronger in the shade, as one might expect under the shade-avoidance hypothesis. These results therefore point to an as yet unidentified selective factor. This is an interesting example of how knowledge of the molecular basis of a trait (the mechanics of phytochrome-based shade avoidance) has generated a testable hypothesis at the organismal level of study (is it really shade from plant neighbors that is causing the observed selection?). However, the hypothesis happened to be incorrect, and it is possible that the shade avoidance response plays less of a clear-cut role in the adaptation of A. thaliana to changing environmental conditions than it was previously thought (but see Dorn et al. 2000). The question of what, then, caused selection for early flowering in our experiments, is still open, although the necessity to avoid drought is a likely candidate.

As I have mentioned above, the issue of selection on flowering time is related to the life history of Arabidopsis, of which we still know surprisingly little. Donohue (2002) investigated the effect of the time of germination on natural selection on life history char-

---

**FIG. 1.** This is an example of selection analysis under field conditions, measuring the relationship between flowering time and relative reproductive fitness in two populations of A. thaliana. The two lines are the result of linear regression analyses for the two populations; their slopes were not significantly different from each other ($F_{1,441} = 0.40, P > 0.05$). Ag. Station and Sharp Ridge are the names of the two localities, near Knoxville, Tennessee (see Callahan and Pigliucci 2002).
acters in *A. thaliana*. She started from the observation that in the United States, where *A. thaliana* has been introduced from Europe, at least southern populations tend to behave as winter annuals, whereas more northern populations can behave either as winter or spring annuals. Donohue collected seeds from five naturalized populations from Kentucky and forced them to germinate in November, December, and early March. The result was that the selective pressures on these populations were altered dramatically. Not surprisingly, plants germinating in November were larger than those germinating in December, and the latter were subjected to stronger selection for increased size. Interestingly, all spring-germinating plants died before reproducing, suggesting that indeed the spring annual life style is not well suited to the conditions prevalent in the southeast of the US. It is also remarkable that Donohue found no evidence of directional selection on flowering time in any of her populations, in marked contrast with the consistent and strong directional selection Callahan and I found in Tennessee populations, only a few hundred miles away.

Another twist concerning selection on flowering time is provided by a recent experimental field study of *A. thaliana* by Courtney Murren and me (*unpublished manuscript*), where we explored the effect of exposure to a novel environment on the impact of herbivory in this species. We found evidence for tolerance and concluded that it may play an important role in the adaptation of recently established populations to novel habitats. Our results indicate that cauline leaves in particular play an important role in the compensation for loss of resources due to herbivory. In general, Murren and I found supporting evidence for the hypothesis that directional selection will favor increased plant size in a nonnative habitat (Thebaud and Simberloff 2001), even in the presence of novel herbivores. We also found dramatically different patterns of selection on the same traits depending on whether the plants had experienced fungus gnat larvae damage, aphid damage, or did not experience herbivory at all (Fig. 2). Importantly, since this research was conducted on a series of recombinant inbred lines (Alonso-Blanco et al. 1998b) for which a detailed genetic map is available, work is in progress to correlate the observed selection under the three different conditions with quantitative trait loci (QTL). Research is also ongoing to identify candidate genes affecting the natural genetic variation underlying plasticity to herbivory and to match it with the available knowledge base on the molecular basis of response to herbivores in *A. thaliana* (van Wees et al. 2000, Jander et al. 2001). For example, Mitchell-Olds’ group has recently identified QTL underlying the response to generalist and specialist herbivores in *A. thaliana*, which uses defense compounds such as glucosinolates and myrosinase (Kliebenstein et al. 2002).

In general, the papers briefly discussed in this section show that the patterns of selection on flowering time in *A. thaliana* can be complex because they are affected by a variety of factors, including seasonal variation in external conditions and not necessarily predictable events such as herbivore attacks. There does not seem to be a simple set of answers, but rather distinct patterns that depend on the particular contingencies of the populations studied and the environments in which they live.

**The Other Side of the Selection Question: Genetic Architecture**

So far, we have considered the study of selection independently of the underlying genetic architecture on which it has to act. But if one wishes to understand the evolutionary outcome of selection, one needs an understanding of the number and relationship (e.g., linkage, epistasis) of the loci affecting quantitative trait variation germane to the characters found to be under selection (Lynch and Walsh 1998, Phillips 1999, Otto and Jones 2000). Indeed, studies of the natural genetic architecture of *A. thaliana* have become increasingly common in the recent literature (Kowalski et al. 1994, Mitchell-Olds 1996, Kuittinen et al. 1997, van der Schaar et al. 1997, Alonso-Blanco et al. 1998a, Stratton 1998, Juenger et al. 2000).

Let us therefore briefly examine some attempts at characterizing the genetic architecture of life history and other phenotypic traits in *A. thaliana* and discuss how it affects our understanding of selection. The research mentioned below is based on so-called quan-
tative trait loci (QTL) mapping, which is a statistical technique that allows one to pinpoint genomic regions associated with the expression of a particular phenotype (Phillips 1999). Essentially, the researcher identifies genotypes that diverge for one or more traits of interest and crosses them. Segregating populations (F2's, backcrosses, etc.) are then both mapped for random molecular markers along the genome, which represents the labor-intensive step that has been carried out mostly in a few model systems, and scored for the phenotypic attributes of interest. A variant of multiple regression analysis is then carried out to associate the phenotypic scores with genomic regions that are inferred to be causally linked to the expression of those phenotypic scores.

One of the early studies of QTL affecting flowering in Arabidopsis was conducted on an F2 progeny of a cross between an early and a late flowering natural populations (Kowalski et al. 1994) and found two putative genomic regions, both on chromosome 5 (A. thaliana has a haploid chromosome number of five). One of these regions contained two known mutants affecting flowering time, co and flc, thereby opening the possibility of follow-up studies using these as candidate loci. Mitchell-Olds (1996) also found two QTL affecting flowering time in a different cross between two laboratory lines of A. thaliana. These were both early flowering, but he used the rather unnatural condition of a 24-hr photoperiod, which is known to artificially accelerate flowering in this species and may have affected the results in an unpredictable fashion. The QTL found by Mitchell-Olds mapped in different regions from the previous study (on chromosomes 1 and 2), with one of them coinciding with the location of another candidate gene, the GIGANTEA locus. More recently, Ungerer et al. (2002) investigated the QTL affecting a variety of traits in the progeny of two different crosses between two laboratory lines and one of these and a natural population. Five QTL were found to affect flowering time in both crosses, but their chromosomal locations and effects were different in the two crosses for all but possibly one of the QTL.

A complication of QTL studies, which however also has the advantage of making them more realistic, occurs when they are conducted in multiple environments to investigate the possibility of genotype-environment interactions shaping the genetic architecture of the traits of interest. For example, a cross between a laboratory line and a natural population with similar flowering schedule was studied by Alonso-Blanco and collaborators (1998a), who grew their plants under different combinations of photoperiod and vernalization treatments. Four distinct QTL were identified, one on chromosome 1 and three on chromosome 5; one of these has been putatively associated with a known mutant affecting flowering time, fha, a second one with the same flc or co loci mentioned above, a third one to a known natural allele causing late flowering, while the fourth QTL did not map close to any known candidate gene. Stratton (1998) exposed one hundred recombinant lines from a cross between two laboratory strains of A. thaliana to a gradient of light intensity and found seven QTL affecting flowering time and eight affecting the genetically correlated trait of leaf production. Only two of these genomic regions affected the response of the plants to the environmental gradient, and the low amount of variance they explained suggested the existence of additional loci involved in the plasticity in response to light, with individual effects too weak to be picked up by this sort of analysis.

What can one learn from QTL studies of flowering time in Arabidopsis so far? The first observation is that things are more complicated than it was thought a few years ago. While we now know quite a bit more about several genomic regions presumably causally involved in the variation in flowering schedule of different genotypes of A. thaliana, it is also clear that very few general patterns emerge. For one thing, different crosses highlight different QTL, as one would expect if the genetic bases of the trait in question are diversified among natural populations of this species. One could conclude that there is no unified block of "flowering time genes," though any such conclusion must be tempered by the fact that these experiments are limited by changes in marker positions and coverage, recombination fractions, sample sizes and observable phenotypic variance between different crosses. Second, changes in environmental conditions also affect the number and effect of the QTL related to flowering time, as expected if genotype-environment interactions play a significant role in the phenomenon, which had been ascertained independently through ecological genetic studies. Third, while some of the natural variation for flowering time may be linked to candidate genes that have been identified by mutagenesis and molecular genetic studies, several others seem to represent additional loci, and probably many more genes play a role that cannot be detected by the relatively low power typical of many QTL studies.

As in the case of phenotypic selection analyses discussed above, QTL studies have inherent limitations that need to be understood (for a recent review see: Asins 2002). As far as A. thaliana in particular is concerned, it is not clear how ecologically informative studies of the progeny of crosses among highly inbreeding parents really are. While, of course, a significant amount of phenotypic variation is generated, and this can be associated with polymorphisms at the genetic level, most of this variation would not occur under natural conditions and may not be reflective of the actual evolutionary paths that have led to the observable (fixed) differences between parental genotypes.

More broadly, perhaps the major concern about QTL research is conceptual: even though this research uses molecular markers, it is still based on a statistical ap-
approach to the problem. From a pragmatic point of view, the efficacy of QTL studies depends on the sample size used as well as on the degree of resolution of the available molecular map. Both of these pose an upper limit to the power of the approach. On the one hand, it is still labor intensive and costly to build an accurate molecular map for a new organism, which is a necessary step if a new cross is made between genetically different enough populations or species. While this is probably going to get easier in the near future because of technical improvements, it will still not represent a trivial task for ecologists. Second, and more important, the limit imposed by the number of offspring of a cross that can reasonably be expected to be studied will preclude QTL mapping on many large organisms and will bound the power to detect QTL characterized by small effects. Since it is possible that small effect QTL are in fact widespread and account for a good portion of the natural genetic variance in a trait, other approaches will need to be devised in order to obtain a picture less biased in favor of genes with large effects. Finally, and related to the problem of power, QTL mapping does not really pinpoint individual genes, but only fairly large genomic regions that can include hundreds or even thousands of genes. This leaves a huge task to the molecular geneticist to bridge the gap between quantitative genetics and a more fine molecular analysis of the genetic architecture of a given trait. These limitations notwithstanding, QTL studies are obviously a valuable addition to the toolbox of the evolutionary ecology detective and need to be integrated with the other approaches developed so far to answer the biological questions of interest to our community.

A different problem related to the study of genetic architecture has been pointed out by Wagner and Altenberg (1996): studies conducted on current populations only reveal what these authors refer to as genetic variation, which they distinguish from genetic variability. The latter indicates the potential of the genetic architecture to be altered over a relatively short period of time, and it is therefore crucial to estimate it if we wish to gain some insights into the likelihood of a population escaping what at the moment might look like a genetic constraint.

Studying variability instead of variation is not a trivial matter, conceptually or empirically; however, an example concerning again flowering time in *A. thaliana* is provided by work that Mark Camara and I have conducted using a rather unusual tool for evolutionary biologists. We have applied selection on natural populations whose genetic variation had been enhanced by exposure to ethyl-methane-sulfonate (EMS), a mutagen routinely used in *Arabidopsis* and other genetic research (Camara and Pigliucci 1999, Camara et al. 2000). The idea was to see if it were possible to break the constraint represented by the genetic correlation linking flowering time and leaf production, the same correlation that makes it impossible for some natural populations to respond to the combined selective forces sometimes acting on these two traits that we have seen in the previous section. The results were very clear (Fig. 3): when we pushed the populations along the direction of the genetic correlation, e.g., toward early

![Fig. 3](image-url)
flowering and lower leaf production, they responded dramatically, with the EMS-enhanced ones responding faster than the base populations, due to their larger genetic variation. However, all attempts to move the populations away from the area of phenotypic space identified by the genetic correlation between the two traits, thus mimicking natural selection, yielded results no different from a control situation with no selection at all.

These results are in agreement with expectations from quantitative genetics theory and with the intuitive idea that there must be a trade off at the physiological level between flowering time and leaf production: after all, it does take more time to produce more leaves. However, they are in contrast to the fact that natural “late” flowering populations of *A. thaliana* actually show no correlation at all between these two characters (M. Pigliucci, personal observation). How, then, did the two traits become decoupled in the course of evolution? Or did they start that way and became coupled in the ecologically specialized “early” flowering populations? Under which natural selective pressures does *A. thaliana* switch from one pattern to the other? And how did this dichotomy originate from what appears to be a widespread winter annual or biennial habit in *A. thaliana’s* close relatives? As I shall argue next, at least some of these organismal-level questions may be illuminated by what we know of the molecular genetics of flowering time in this species.

**IF WE ONLY KNEW SOMETHING ABOUT THE GENES...**

One chief, and often repeated, reason to use *Arabidopsis* as a model system in evolutionary ecology is that it has been well developed as a model system in physiology and molecular biology. Historically, organismal biologists have often wished to know something more about the genetic underpinnings of traits in which they were interested, and studies of the molecular genetics of flowering time in *Arabidopsis* have the potential to fulfill this dream beyond expectations.

Fortunately for evolutionary ecologists, flowering time happens to be a trait of high interest to molecular and developmental biologists, so much so that there is a very large literature on its molecular basis as well as on how environmental conditions interact with the basic machinery of the cell to yield a remarkable variety of outcomes (e.g., Alonso-Blanco et al. 1998a, Arai et al. 1998, Stratton 1998, Blazquez and Weigel 2000, Johanson et al. 2000, Mas et al. 2000, Sheldon et al. 2000). Another area of research that has generated similar excitement and that can be seen as a promising contact between molecular and organismal biology concerns plant–pathogen or plant–herbivore interactions. While I do not have space for a separate review of this research, the reader may be interested in some of the key papers (Mauricio and Rausher 1997, Peters 1999, Warren et al. 1999, van Wees et al. 2000, Jander et al. 2001).

This is also obviously not the place for a review of the literature on the molecular biology of flowering time, but a sketch of the current state of the art in the field should be sufficient as a basis to discuss to what extent knowledge of the molecular basis of a trait can inform our studies of selection on the same trait. Blazquez and Weigel (2000), for example, have investigated the importance of floral meristem-identity genes such as LEAFY and of hormones like gibberellin in integrating environmental and endogenous signals that control flowering time in *Arabidopsis*. The “decision” to flower is affected by at least four pathways (Michaels and Amasino 2000): one accelerates flowering in response to long days (photoperiod pathway); one is independent of photoperiod but sensitive to other environmental factors such as vernalization; a third one, the so-called “autonomous” pathway, makes it possible for the plant to eventually flower regardless of environmental conditions, presumably according to an internal clock (*A. thaliana* is a facultative long-day plant, meaning that it flowers earlier under extended photoperiods, but that it will eventually flower even under short days); a fourth pathway also responds to photoperiod, this time to short days, and depends on the action of the gibberellin hormone.

These pathways have been separated through the identification of mutants that block each of them separately from the others, and Michaels and Amasino (2000) have highlighted the similarities between these mutants and spring-flowering or winter-annual wildtype strains of *A. thaliana*. They even demonstrate how it is possible to construct a mutant that does not flower at all in the absence of a cold treatment, essentially transforming *A. thaliana* into a biennial plant, a remarkable ecological transition obtained under laboratory conditions! This is important because some of the close relatives of *A. thaliana* (Mitchell-Olds 2001) are in fact biennial, a fact that immediately suggests the as-yet unexploited possibility of using the biennial-inducing genes of *A. thaliana* as molecular probes to identify genes underlying the evolution of life history in its phylogenetic neighborhood.

A link between the molecular and organismal aspects of the flowering time puzzle is provided by a discussion by Simpson and Dean (2002). First, they describe a model of the changing effect of different floral induction pathways with seasonality. According to this model (Fig. 4, left), plants don’t flower during the winter because the FRIGIDA locus stimulates the production of *FLOWERING LOCUS C* product, which inhibits flowering. Neither photoperiod, which is short during this time of the year, nor the autonomous, environmentally independent, pathways are sufficient to overcome this effect. During the spring, however, the effect of photoperiod becomes more important because days get longer, the accumulated days of cold are enough to
reduce the expression of *FLOWER LOCUS C*, and the plants flower. How might this information illuminate ecological considerations about natural selection? If this mechanistic scenario has any relevance at the population level, one would expect the major genetic elements of this system to be under stabilizing selection in winter annual natural populations of *Arabidopsis*. This is an eminently testable prediction made possible by the fact that one can now study selection on quantitative trait loci using recombinant inbred lines. What one has to do then is to look for populations that differ at one or more of the loci involved and measure selection on those QTL. To my knowledge, no such research has been conducted so far.

Simpson and Dean (2002) have also proposed an evolutionary pathway explaining the transition from winter annual to what they call “rapid cycler” habit (i.e., spring annual populations). It turns out that different populations of rapid cyclers have independently lost the *FRIGIDA* function, which removes the brakes to activation of the floral pathway, normally antagonized by vernalization (Fig. 4, right). As a consequence, these plants flower without a vernalization requirement and, under controlled laboratory conditions, behave as “early flowering” genotypes. Again, the molecular scenario can stimulate venues of research at the organismal level: if this is in fact the preferential evolutionary route, then one should expect to see the repeated independent evolution of rapid cyclers from winter annuals, but not vice versa, because loss-of-function mutations tend to be more frequent than gain-of-function ones. Data on the genetic and historical relationships among populations, which are available for some accessions (Erschadi et al. 2000, Vander Zwan et al. 2000, Pigliucci et al. 2003), could be used to test this possibility. Again, to my knowledge, this has not been done yet.

The examples mentioned so far illustrate the idea that molecular and organismal biology can in fact converge toward a common research program, at least when model systems for which one has abundant information at the mechanistic level are concerned. However, I would also like to draw the attention of the reader to the limitation of integrative biology, something that is rarely addressed in this era of enthusiasm for genomics and molecular reductionism. So far, the assumption has been that molecular and organismal levels of analysis are complementary pieces of information necessary to gain a complete picture of the puzzle of, say, the evolution and mechanics of flow-
ering time. However, there are serious reasons to think that the two levels of investigation might be more decoupled than normally granted (Dupré 1993; Pigliucci, in press): organismal and structural questions may be epistemologically largely independent of each other and the points of contact may represent more the exception than the rule. Flowering time in Arabidopsis may be used again as an example. While I have shown how knowledge of molecular mechanisms may generate some interesting evolutionary hypotheses, the majority of research questions concerned with the ecology and evolution of flowering time need only assume that there is a variable genetic basis to the trait in order to be carried out: questions concerning which environmental conditions or population dynamics favor the evolution of different life histories are largely independent of the genetic details, and have in fact been successfully pursued as such. Conversely, molecular biologists interested in the structural aspects of the genetic machinery underlying certain traits have gained very little from organismal-level studies of selection, demography, and microevolution. All they need to know is what effects the genes they focus on have on the phenotype at large, not under what specific circumstances these effects are favored by natural selection.

The point I wish to drive home here is that, while it is, of course, interesting to explore the borderlands between structural molecular biology and functional organismal biology and to foster cross-benefits whenever one can find them, there is a danger of overemphasizing the points of contact between the two areas of research. Such an overemphasis can result in the unwelcome outcome of constraining researchers in both fields to pursue a significantly more narrow research program than they would be otherwise inclined to do. There is much molecular biology that can be done without addressing evolutionary ecological questions, and vice versa.

**Coda: A Lot of Work to Do**

My goal in this brief overview of research on selection and the genetic architecture of selected traits in Arabidopsis was to make the case for a use of this model system in ecological and evolutionary genetic studies by exploring the question of what one learns about these systems and how readily it applies to other groups of organisms. This is, of course, an open empirical question.

Finally, I think studying Arabidopsis can teach us some humility. It is certainly more exciting to go into the rainforest to do ecology than to visit roadsides and railroad tracks. However, I have given the reader a glimpse of the fact that even in the case of such a deceptively simple weed, we actually know much less ecology and evolution than we would have expected to. There is indeed a lot of work ahead of us.

**Acknowledgments**

I wish to thank Anurag Agrawal, Jeff Dole, and Paula Kovar for comments on previous versions of this manuscript. I could not have written this without the help of several people in my lab, especially three of my former postdocs, Mark Camara, Hilary Callahan, and Courtney Murren. This work was in part supported by NSF grant DEB-0089493.

**Literature Cited**


**APPENDIX**

**A BRIEF “HOW-TO” GUIDE TO ARABIDOPSIS THALIANA ECOLOGICAL GENETICS**

If someone wished to use Arabidopsis thaliana to study selection, or any other topic in evolutionary ecology, where would they start? Here are a few pointers for students and colleagues who are actually interested in pursuing the use of this model system in practice. Notice that many of these techniques can, of course, be implemented in other species as well. What makes A. thaliana special at this point is simply the wealth of materials and protocols already available to the researcher. Indeed, it is exactly the eventual transferability of the approaches discussed in this review to other species, starting with the close relatives of A. thaliana, that makes this endeavor especially interesting for an evolutionary ecologist.

Perhaps the first set of questions that come to mind concern very basic aspects of the logistics of these experiments: e.g., what sample sizes should one use? Of course, the simple answer is “the more the better,” and with organisms like A. thaliana one can in fact work with fairly large numbers of individuals. Nevertheless, even under ideal conditions, one does run into limitations of space, personnel, and time, especially when engaging in quantitative genetics experiments. My experience has been that experiments with up to 1000 plants are fairly manageable, and that one can reach the 2000 without too much trouble. The question of how to subdivide this sample size among families, replicates, and possibly different environmental treatments does not have a general answer because it depends on the specific question at hand. As a rule of thumb, one should set up experiments that include two replicates with highly genetically homogenous material such as A. thaliana and still maintain statistical power because of the reduced error variance. Other than the number of replicates, one needs to keep in mind that selection experiments often require multiple controls, and should always be repeated at least twice (in parallel, if possible) to account for the effects of random drift. Once that constraints such as number of replicates at the genotypic level and number of independent lines of selection and controls have been established, one can then maximize the representation of different genotypes within a population or of populations within A. thaliana.
thaliana until the capacity of the available facility has been reached. One of the basic things to realize about A. thaliana, which is mostly not true of its close relatives, is that its highly selfing mating system generates practical trade-offs that make it very convenient for some kinds of research but rather impractical for others. For example, it should be obvious that classical quantitative genetic experiments such as diallelic crosses (Lynch and Walsh 1998) are painful in this case because they require the emasculation of thousands of flowers before the very small buds open (although Westerman [1970a, b] and Westerman and Lawrence [1970] carried out several diallele crosses). Unfortunately, this precaution is necessary if one wishes to have some control on what pollinates what (but see Jones [1971], who shows that this species is protogynous and has some degree of natural outcrossing). By the same token, however, QTL experiments (Kowalski et al. 1994, Kuittinen et al. 1997, van der Schaar et al. 1997, Alonso-Blanco et al. 1998a, Stratton 1998, Juenger et al. 2000) are relatively easy not just because they can take advantage of already available detailed genetic maps, but because after the initial cross and perhaps some later backcrosses, it is a straightforward matter to obtain recombinant inbred lines: all one has to do is to let the plants self for however many generations one needs to obtain quasi-isogenic lines.

As far as selection experiments in particular are concerned, there are again several approaches, some more convenient or appropriate, depending on the question to be investigated. Mass selection experiments in which plants are simply left to self and occasionally to outcross by mechanical contact, or by being exposed to generalist pollinators in the open, can be carried out. In this case the goal is to simulate how selection occurs under natural conditions. However, because of the relatively small amount of intrapopulation genetic variation, the response to selection can be limited, unless one uses widely diverging collections of populations, which would not occur simultaneously in the same location in nature. Seeds of accessions collected throughout the world can be obtained from the Arabidopsis Information Resource.1 It is, on the other hand, easy to obtain a highly variable population by crossing quasihomozygous strains and let the progeny self as if one were at the early stages of deriving recombinant inbred lines. These highly variable populations can be used to test specific hypotheses about the relationship between phenotypic characteristics and reproductive fitness when studying selection, but they are obviously not reflective of genetic and phenotypic variation as found in natural populations.

Field or greenhouse experiments on selection can be conducted following the standard approach of regressing phenotypes of either individual plants or the averages of genetically related families (Lande and Arnold 1983, Rausher 1992) against a measure of reproductive fitness (Callahan and Pigliucci 2002). The latter is usually simply a count of fruit production, since seed production is directly proportional to it (Westerman 1970c), and fruit numbers are far easier to count. At any rate, since this is an annual plant with no vegetative reproduction, output in terms of reproductive progeny is a good measure of success, though one should take into account variation in seeds’ viability. A particularly interesting kind of field or greenhouse selection experiment can be conducted by exposing genetically mapped recombinant inbred lines to whatever selection agent one is interested in. In this case, not only one can gather valuable insights into the effects of selective forces on particular aspects of the phenotype, but it is possible to move on to the next step of mapping the genomic regions involved in the observed variation in reproductive fitness as well as in the traits under selection. However, there are many issues of proper statistical analysis that should be considered when embarking in this sort of research. One of these is that obviously a set of recombinant inbred lines is not a random sample of a natural population since their parents in turn are not randomly selected, so that mixed-model analyses of variance that treat genotype as a random effect may not be appropriate.

In this paper, I mentioned the use of mutagens to produce additional genetic variation useful for studies of genetic correlations and response to selection. The techniques for doing this are very well worked out in Arabidopsis (Kooiinenee et al. 1997), so much so that one can even acquire already mutagenized stocks with different genetic backgrounds for a few dollars or, for a higher price, one can order custom made mutagenized lines from Lehle Seeds.2 If one wishes to prepare lines from scratch, the paper by Kooiinenee et al. contains pointers on what dosages of ethyl-methane-sulfonate (EMS) to use. The procedure simply calls for soaking seeds in a low concentration solution of EMS for a variable period of time, depending on how many “hits” one wishes to obtain on average (Camara et al. 2000). It is fairly simple also to check the efficacy of the mutagenic treatment by inspecting fruits for embryos with albino mutations and compare their frequency with the known spontaneous frequency of mutations causing the same phenotype.

Of course, a great advantage of plants in general for evolutionary ecological experiments is that it is easy to follow individuals under field conditions. With A. thaliana, my laboratory has done it over the winter at the rosette stage, and once one has developed a search image it is possible to identify new individuals coming up at early stages of rosette growth. Perhaps more importantly, seeds can be stored for long periods of time (at least two years, often more, in the case of A. thaliana) and maintain their germinability and viability, which makes selection experiments more manageable. This means that it is possible to store the seeds of the base population and then grow it alongside the nth generation of a selection experiment, thereby reducing the necessity of making questionable assumptions on the stability of the genetic constitution of the base population when running the pertinent statistical analyses. Of course, this is true if the selection experiment does not last for several years so that stored seeds are no longer viable, or that selection for viability significantly changes the genetic makeup of the population.

To achieve all this it is crucial to collect seeds at the proper stage and to handle them with care. In A. thaliana, the best time to collect seeds is when the siliques (the fruits) turn yellowish.

Since one of the main advantages of working with Arabidopsis is that it joins a large and diverse community of researchers, it is always a good idea to send seeds of whatever new mutagenic line, recombinant inbred line, or natural population one develops or collects to one of the two international Arabidopsis stock centers, one in the United States (see footnote 1) and the Nottingham Arabidopsis Stock Centre in the United Kingdom.3 It is this continued use of the same material and the progressive extension of such collections that make research on different aspects of Arabidopsis increasingly convenient and fruitful.

1 URL: (www.arabidopsis.org)

2 URL: (www.arabidopsis.com)

3 URL: (http://nasc.nott.ac.uk/)