Methods

Statistical convenience vs biological insight: consequences of data transformation for the analysis of fitness variation in heterogeneous environments

Maureen L. Stanton and Denise A. Thiede

Section of Evolution and Ecology and The Center for Population Biology, University of California, Davis, CA 95616, USA

Author for correspondence:
Maureen L. Stanton
Tel: +1 530 7522405
Fax: +1 530 7521449
Email: mlstanton@ucdavis.edu

Received: *31 July 2004*
Accepted: *1 November 2004*

Summary

• In plants, more favourable environmental conditions can lead to dramatic increases in both mean fitness and variance in fitness. This results in data that violate the equality-of-variance assumption of ANOVA, a problem that most empiricists would address by log-transforming fitness values.
• Using heuristic data sets and simple simulations, we show that ANOVA on log-transformed fitness consistently fails to match the outcome of selection in a heterogeneous environment or its sensitivity to environmental frequency. Only ANOVA based on relative fitness within environments accurately predicts the sensitivity of genotype selection to the frequency of alternative environments.
• Parallel analyses of variance based on absolute fitness and relative fitness can bracket the expected success of alternative genotypes under hard and soft selection, respectively. For example, for *Sinapis arvensis* growing in full sun and partial shade treatments, families achieving high fitness in the best environment are favoured under hard selection, whereas soft selection favours different families that achieve consistently good performance across environments.
• Based on these findings, we recommend that log-transformation of fitness should no longer be standard practice in ecological genetics studies. Weighted ANOVA is a preferable method for dealing with unequal variances, and investigators should also make greater use of techniques such as quantile regression or resampling to describe and evaluate fitness variation across heterogeneous environments.

Key words: analysis of variance (ANOVA), genotype-by-environment interaction, hard selection, logarithmic transformation, phenotypic plasticity, soft selection.

New Phytologist (2005) **166**: 319–338

© *New Phytologist* (2005) **doi**: 10.1111/j.1469-8137.2004.01311.x

Introduction

One of the most conspicuous features of plants and other modular organisms is the extraordinary degree to which their morphology, physiology, and fitness change with environmental conditions (Bradshaw, 1965). Until quite recently, the genetic bases for such plastic responses were almost entirely unknown, but new analyses at the interface of genomics, development, and ecology are beginning to shed light on how specific genetic variations influence plant responses to spatial environmental heterogeneity. Most notably, experiments using well-characterized mutants in model systems such as *Arabidopsis thaliana* are not only showing how specific genes control plant responses to light quantity, spectral composition, and spatial distribution, but are also demonstrating that these plastic responses have significant fitness costs and/or benefits across a range of experimental treatments (Ballare *et al.*, 1994; Schmitt *et al.*, 1995; Galen *et al.*, 2004; Pigliucci & Schmitt, 2004). These studies have laid the groundwork for empirical research that will fully integrate ecology with evolutionary and developmental

biology in an effort to document how naturally occurring environmental heterogeneity influences evolution (Jackson *et al.*, 2002; Sultan, 2003). A central aim of such studies will be to characterize the fitness consequences of genotypic variation across both natural and manipulated levels of environmental variation. In this paper we discuss the use of analysis of variance for data interpretation in such studies.

Natural environments and selection regimes vary over a wide range of spatial and temporal scales (Antonovics *et al.*, 1987; Stewart & Schoen, 1987; Prout & Barker, 1989; Bell & Lechowicz, 1991; Stratton, 1994; Galloway, 1995; Stratton & Bennington, 1996; Juenger & Bergelson, 2002). Population and quantitative genetics models suggest that evolutionary responses to such environmental heterogeneity will be determined by the presence of positive or negative genetic correlations in relative fitness across selection regimes, as well as by the relative scales at which differential selection and gene flow operate (Levins, 1968; Gillespie, 1974; Felsenstein, 1976; Via & Lande, 1985, 1987; Gillespie & Turelli, 1989; Van Tienderen, 1991). These models have yet to be tested exhaustively in any ecological system, although many of their elements have been examined in organisms as diverse as pea aphids (Via, 1999; Caillaud & Via, 2000; Via *et al.*, 2000), sticklebacks (Hatfield & Schluter, 1999; Peichel *et al.*, 2001; Bolnick, 2004), and jewelweed (Donohue *et al.*, 2000a,b).

To study the genetic consequences of environmental heterogeneity, evolutionary ecologists often conduct experiments in which individuals from two or more groups (genotypes, ecotypes, lineages, varieties, or populations) are exposed to contrasting environmental regimes. (To simplify our presentation, we use the term 'genotype' to refer to this broader set of genetic groupings in this paper.) Environment-dependent changes in performance (e.g. the 'norm of reaction' for fitness: Schmalhausen, 1949) can then be measured for each genotype. Ideally, the results of such experiments should indicate whether overall fitness varies among genotypes, if there is evidence for fitness tradeoffs between habitats, and the extent to which the success of alternative genotypes depends on the frequencies of contrasting selection regimes.

To date, most experiments have contrasted the responses of genotypes to just two or three environmental extremes. For example, experiments designed to test for local adaptation typically compare the fitness that putative ecotypes achieve in distinct 'home' vs 'away' conditions (e.g. Antonovics *et al.*, 1971; Schemske, 1984; Lively, 1989; Via, 1991), while laboratory experiments aimed at understanding the consequences of environmental gradients for adaptation often expose individuals to conditions representative of the opposite environmental extremes their populations are likely to encounter in nature (e.g. Björkman & Holmgren, 1963; Bennett & Grace, 1988; Hoffman & Watson, 1993; Sultan & Bazzaz, 1993; Schmitt *et al.*, 1995). For the majority of empirical studies, in which environmental variation is treated as a fixed categorical factor, analysis of variance (ANOVA) has become a standard tool for determining how genotype, environment, and genotype-by-environment interaction contribute to overall variance in fitness or phenotype (e.g. Via, 1991; Sultan, 1992; Simms & Triplett, 1994; Shaw *et al.*, 1995; Galen *et al.*, 2004; Pigliucci & Schmitt, 2004). Here we consider the simplest example of this type of analysis and experimental design, a two-factor ANOVA in which genotype and environment are crossed effects.

When absolute fitness, or a major component of fitness, is the dependent variable in such an analysis, each factor in the ANOVA has a specific evolutionary interpretation. A significant main effect of environment, averaged across genetic groups, implies differences in environmental quality (e.g. Dingle, 1992; Via & Conner, 1995; Galen *et al.*, 2004; Stanton *et al.* in press). A main effect of genotype, averaged across test environments, suggests that some lineages will have an overall selective advantage if the alternative selection regimes are equally frequent (Allard & Bradshaw, 1964; Dingle, 1992; Andersson & Shaw, 1994). Significant genotype-by-environment interaction may indicate that genotypes vary with respect to the difference in their mean fitness between environments, yielding nonparallel reaction norms when plotted on an arithmetic scale (Lynch & Walsh, 1998).

Genotype-by-environment interaction for fitness can be a potent force maintaining genetic variation in populations (Gillespie & Turelli, 1989). Models of evolution in heterogeneous environments consider two types of population regulation, soft and hard selection. Under soft selection, relative fitness within habitat types determines genotype frequencies after selection, but the contribution of each habitat type to the total population pool is independent of local genotype frequencies or mean fitness (Christiansen, 1975; Wallace, 1981). In other words, population regulation occurs after the selection episode and within habitat patches. Under hard selection, selection in high-quality environments dominates the evolutionary trajectory of the population as a whole, and genotypes that have the greatest absolute fitness across environments are likely to increase in frequency. Single locus models (Levene, 1953; Dempster, 1955; Gillespie, 1974; Christiansen, 1975) and models assuming quantitative trait inheritance (Via & Lande, 1985, 1987; Gillespie & Turelli, 1989; Van Tienderen, 1991) substantially agree in predicting that the conditions for maintenance of genetic variation are much less stringent under soft selection.

The purpose of this paper is to consider how our conclusions from ecological genetics experiments may be altered by the choice to use absolute fitness, relative fitness within environments, or log-transformed fitness as the outcome variable in ANOVA. Most investigators will analyse absolute fitness values unless doing so violates assumptions of ANOVA, but in studies of plant responses to contrasting environments, individual fitness data usually violate those assumptions egregiously in being strongly L-shaped and heteroscedastic. Standard biometry texts prescribe log-transformation as a cure for this affliction (Sokal & Rohlf, 1995; Zar, 1996), and so log-transformed fitness is almost universally used in ANOVA

models for plant ecological genetics experiments (e.g. Galen *et al.*, 2004; Pigliucci & Schmitt, 2004).

Although the statistical rationales for data transformation are familiar to most evolutionary ecologists, it is not widely appreciated that manipulating data values to optimize statistical hypothesis-testing may impair biological insight (reviewed in Grissom, 2000). Using four hypothetical data sets as examples, we show how (and why) the use of absolute fitness, log-transformed fitness, or relative fitness within environments as the outcome variable in such experiments changes the variance terms associated with genotype, environment, and the genotype-by-environment interaction. These ANOVA results are then compared to simulations of hard and soft selection that mirror the fitness differences in each hypothetical data set. The comparisons show that the common practice of log-transforming individual fitness values results in main effect and interaction terms that do not consistently predict genotype performance under either hard or soft selection. As an alternative, the use of parallel ANOVA models, based on absolute fitness and relative fitness within environments, can be used to bracket the possible responses to hard and soft selection. In experiments on plants, variances in fitness are often extremely different across treatments, and this presents a serious, and still imperfectly resolved, challenge for statistical hypothesis-testing (Potvin & Roff, 1993; Dutilleul & Potvin, 1995). Here, we demonstrate the use of weighted ANOVA (Welch, 1951), rather than log-transformation, as one possible way to deal with heteroscedasticity in fitness across environments.

Materials and methods

Overview of approach

We begin by demonstrating how proportional changes in fitness across environments, changes in genotype ranking across environments, or a combination of these factors influence the significance of the main effect of genotype and the G × E interaction in an ANOVA. The design of our hypothetical factorial experiment is one in which replicate individuals of two genotypes are exposed to two environments that differ in average quality. We consider four potential outcomes of the experiment, each representing different fitness responses of genotypes to the environmental regimes. In the first two cases, the mean fitness of genotypes increases additively (#1) or proportionately (#2) in environment 2, but the genotypes do not change in rank across environments. In such situations, G × E interaction may alter the rate at which one genotype replaces the other, but is not expected to maintain genetic variation over the long-term (Via & Lande, 1987). Next we consider two cases (#3 and #4) in which the fitness rankings of genotypes change between the two environments. For each of these four hypothetical scenarios, we compare the results of ANOVA using absolute fitness, relative fitness within each environment, or log-transformed fitness as the outcome variable. To investigate whether the ANOVA results for different fitness measures inform us about expected changes in genotype frequency, we use a simple model of clonal selection to examine short-term changes in genotype frequency under both soft selection and hard selection. Finally, to illustrate how transformation of fitness values can change the interpretation of ecological genetics experiments, we compare ANOVA based on alternative fitness measures in a study of how full sun and partial shade treatments influence lifetime fertility in paternal sibships of an annual plant, *Sinapis arvensis*.

Construction and analysis of mock data sets

The four cases that we consider in our hypothetical experiment were selected to illustrate most dramatically: (i) how changes in scale might motivate an evolutionary ecologist to transform fitness for statistical and/or biological reasons; (ii) how either log-transformation or the calculation of relative fitness affects the significance of the G × E interaction term; and (iii) which fitness measure most accurately predicts short-term evolutionary dynamics under soft vs. hard selection. In each of the four cases we assume a balanced design in which 20 individuals belonging to each of two genotypes are assigned to each of the two environments. We scale the variance in fitness among replicate individuals in each environment to the genotype's mean, since a positive correlation between the mean and the variance is often observed in biological data sets (Snedecor & Cochran, 1989). Twenty fitness values were assigned to individuals within each of the four genotype-environment combinations, as follows. The mean genotype fitness was assigned to seven of the 20 individuals, while the remainder were assigned fitness values of +1 SD, −1 SD, +2 SD, or −2 SD to approximate a normal distribution where the standard deviation is approx. 20% of the genotype's mean fitness.

In analysing experiments of this type, most evolutionary ecologists interested in understanding the consequences of environmental variation for selection among genotypes would focus on the significance of the main effect of genotype and the genotype-by environment (G × E) interaction in a fully factorial two-way ANOVA. Main effects in the ANOVA model may be treated as fixed or random factors, depending on the study system and the question of interest. Treating genotype as a fixed effect is appropriate if test genotypes are selected for their particular characteristics (e.g. Bennington & Thayne, 1994; Newman *et al.*, 1997; Galen *et al.*, 2004). In contrast, analysing randomly selected genotypes allows one to estimate variance components associated with the evolutionary response (Andersson & Shaw, 1994; Donohue *et al.*, 2000a). Treatments selected to represent specific environmental conditions are typically regarded as fixed effects (e.g. Galen *et al.*, 2004), whereas treatments (or blocks) distributed across natural environmental gradients (e.g. Juenger & Bergelson, 2002) can be considered random. In the four simple hypothetical cases that we consider here, we treat both genotype and environment as fixed factors.

Our hypothetical data sets were constructed to reflect the biological reality that phenotypic variability often scales to the mean phenotypic value, a pattern that is particularly striking for plant growth and fitness responses to environments of different quality (e.g. Stanton *et al.* in press). To meet the assumptions of ANOVA, most investigators would log-transform these heteroscedastic fitness data (Falconer, 1989). To verify the statistical justification for transforming fitness, we employed a modification of Levene's test for heteroscedasticity (Levene, 1960) by running the factorial two-way ANOVA, obtaining the residuals and calculating their absolute value, and then using this value as the dependent variable in a second, identical ANOVA model. Highly significant main effects or interactions in this analysis indicate significant heteroscedasticity. In each of our four hypothetical cases, we show that log-transformation does enhance equality of variance, but it does not accurately predict genotype performance.

To conduct appropriate significance tests without log-transforming fitness, we used a weighted least-squares approach to reduce the impact of heteroscedasticity (Welch, 1951; Neter *et al.*, 1996). In practice, this is done by first identifying the main effect or interaction showing the highest level of significance in the Levene's test. Next, for each level of this main effect or interaction, the inverse of residual variance (known as 'precision') is calculated. Finally, this precision value is used to weight each observation in ANOVA, e.g. by including the WEIGHT statement in the GLM or MIXED procedures of SAS (SAS Institute, 1999, for an example, see Stanton *et al.*, 2000). For balanced data sets and fixed-effect ANOVA models like those we present, heteroscedasticity has little effect on significance-testing. However, in a random effects model, heteroscedasticity can have pronounced effects on variance component estimation (Yamada, 1962; Neter *et al.*, 1996). Significance-testing based on re-sampling techniques, including permutation tests or bootstrapping, is not a panacea for this problem, since these methods are also based on the homoscedasticity assumption (Potvin & Roff, 1993; Hayes, 2000).

To investigate how transforming fitness influences the evolutionary predictions of ANOVA, and to determine which fitness measure most closely matches the evolutionary outcome predicted by the clonal selection model, we considered three fitness measures as outcome variables in each of the four hypothetical cases. In addition to absolute fitness, we analysed natural log-transformed absolute fitness (+1) to meet the equality of variance assumption (a statistical justification), as well as to eliminate simple proportional or scaled responses to the environment (a biological justification). Relative fitness within environments, a third fitness measure, represents an alternative approach to eliminating differences in scale between environments. Relative fitness for each individual was calculated by dividing its fitness value by the mean fitness achieved in that environment, as would be done in selection analyses (Lande & Arnold, 1983). Calculating relative fitness eliminates the difference in mean fitness between environments, but does not homogenize fitness variances.

Modeling clonal selection in a spatially heterogeneous environment

We used two simple algorithms to model deterministic, clonal selection in the two-genotype, two-environment universe depicted in each of the four cases of the hypothetical experiment. In the pure soft selection model, each individual produces its assigned number of propagules (as determined by the mean fitness for that genotype in that environment), and then selection reduces the pool of propagules within that environment to the size of the original parental subpopulation. The survivors of soft selection within environments then re-distribute themselves proportionately across environments to form the next generation. Under hard selection, each individual contributes propagules to a common pool, as determined by its fitness value. Selection then reduces the total size of the propagule pool to that of the previous adult population, and the surviving propagules are proportionately re-distributed across environments to form the next generation. Our very simple algorithms for clonal selection assume no sampling variance; mean fitness values and environmental frequencies exactly determine the fractional contribution each genotype makes to subsequent generations. Because the outcome of each episode of selection is deterministic, a genotype's fitness is solely determined by the total number of propagules its representatives produce within each environment, and not by the variance in fitness within genotypes.

Each simulation begins with equal frequencies of genotypes *A* and *B* within each of the two environments. If the number of individuals of genotype A in environment 1 in generation 0 is A_{10}, then $A_{10} = B_{10}$ and $A_{20} = B_{20}$. Thus, the starting genotype frequencies, within and across environments, are as follows: $a_0 = b_0 = 0.50$. The mean fitness for genotype i in environment j (W_{ij}) is the average number of asexual propagules its representatives produce in that environment. Overall population size is constant from one generation to the next, and generations are nonoverlapping; adults die after they produce asexual propagules. The following recursions, for either soft or hard clonal selection, were carried out for 10 generations to generate the frequencies of each genotype over time.

In generation 0, the proportion of propagules produced in environment j that belong to genotype A is given by:

$$\frac{A_{j0} \cdot W_{Aj}}{A_{j0} \cdot W_{Aj} + B_{j0} \cdot W_{Bj}} \qquad \text{Eqn 1}$$

Under soft selection, before migration, selection reduces the size of the propagule pool in each environment to the sizes of the parental subpopulations, N_1 and N_2, for environments 1 and 2, respectively. Accordingly, the number of genotype A propagules surviving selection in environment j is:

$$\frac{N_j \cdot A_{j0} \cdot W_{Aj}}{A_{j0} \cdot W_{Aj} + B_{j0} \cdot W_{Bj}} \quad \text{Eqn 2}$$

With free migration, the surviving propagules produced in each environment combine into the common pool of individuals that forms the next generation. The proportional representation of genotype A in that pool is given by a_1,

$$\frac{\frac{N_1 \cdot A_{10} \cdot W_{A1}}{A_{10} \cdot W_{A1} + B_{10} \cdot W_{B1}} + \frac{N_2 \cdot A_{20} \cdot W_{A2}}{A_{20} \cdot W_{A2} + B_{20} \cdot W_{B2}}}{N_1 + N_2} \quad \text{Eqn 3}$$

These propagules are proportionately re-distributed between the environments without error, and so the new frequency of genotype A in both environments is a_1, and the numbers of genotype A individuals in environments 1 and 2 are $a_1 N_1$ and $a_2 N_2$, respectively.

Under hard selection, the propagules generated by each adult are combined into a single pool before population regulation takes place. Selection then reduces the size of the combined propagule pool to that of the previous generation's adult population. Survivors are distributed proportionately between the two environments to form the next generation. The frequency of genotype A in the next generation (a_1) is calculated as its frequency within the combined pool of surviving progeny, as follows:

$$\frac{A_{10} \cdot W_{A1} + A_{20} \cdot W_{A2}}{A_{10} \cdot W_{A1} + A_{20} \cdot W_{A2} + B_{10} \cdot W_{B1} + B_{20} \cdot W_{B2}} \quad \text{Eqn 4}$$

Among the adults in generation 1, the numbers that belong to genotype A in environments 1 and 2 are $a_1 N_1$ and $a_2 N_2$, respectively.

Where one genotype has an advantage in both environments, the distribution of genotypes among environments can potentially alter the rate, but not the direction, of genotype frequency change. In contrast, when genotype fitness rankings are reversed between environments, both the rate and direction of genotype frequency change depend on the frequencies with which environments 1 and 2 are encountered. To illustrate how evolutionary trajectories change with the frequencies of alternative selection regimes, we vary the numbers of individuals experiencing the alternative environments so that either 10, 50, or 90% of the population is exposed to environment 1 in each generation.

Example data analysis

As part of a large-scale experiment examining the evolutionary consequences of abiotic stresses in the out-crossing mustard *Sinapis arvensis* (Roy & Stanton, 1999; Roy *et al.*, 1999; Stanton *et al.*, 2000), we investigated how paternal families performed in a field experiment in which we manipulated light intensity. For purposes of illustration, only a small subset of the overall design and data are used here; details of the complete experiment are published elsewhere (Stanton *et al.*, in press). The subset of plants analysed here were derived from three generations of selection under low light stress in the greenhouse, followed by a single generation grown under near-optimal conditions, in which controlled crosses were performed to produce paternal half-sibships, each including two maternal parents. We created two contrasting light environments in a disturbed field at the University of California, Davis: two blocks each of full sun vs partial shade created by a 60% neutral shade canopy. After preparing the soil for planting, we laid down a water-permeable weed barrier to minimize uncontrolled spatial variation in the level of competition, and created small holes for planting our experimental seedlings at 20 cm spacing. On 7–8 March 1997, we transplanted 5- to 6-d-old *S. arvensis* seedlings from each paternal half-sibship into random locations within each treatment block. For this analysis, we include all plants from one replicate of the greenhouse shade stress selection treatment (263 plants from six paternal half-sibships). After plants had matured their fruits and senesced, we counted the total number of viable seeds produced, an estimate of lifetime female fitness, and used this value as our measure of absolute fitness.

We used mixed model analysis of variance to characterize genetic fitness variation across the full sun and partial shade treatments for these six paternal half-sib families. Parallel analyses were run for absolute fitness, log-transformed fitness, and relative fitness within each environment, using the RANDOM/TEST statement within the GLM procedure of SAS (SAS Institute, 1999). To facilitate comparison with analyses of the hypothetical data sets discussed earlier, we used the simplest possible two-way ANOVA model, including sire (a random genetic factor), light (a fixed environmental treatment), and their interaction. As the fitness variance attributable to block was nearly zero, data were pooled for the two blocks of each light treatment. To test whether the average heritability for fitness variance is different from zero, we tested genotype over the error variance (Fry, 1992). Because Levene's test revealed significant heteroscedasticity, we conducted a weighted ANOVA for hypothesis-testing, as described in the section 'Construction and analysis of mock data sets,' above. We then discuss how the analysis of different transformations of fitness alter our evolutionary interpretations.

Results

Additive response, Case #1

Consider a hypothetical scenario where both genotype and environment have strictly additive effects on fitness (Fig. 1a). In this example, genotype A produces two more propagules than genotype A in both environments. Both genotypes respond to environment 2 by producing one additional propagule, so the reaction norms for genotypes A and B are parallel when fitness is plotted on an arithmetic scale (Fig. 1a), and the G × E term is negligible in the ANOVA for absolute fitness (Table 1a). Typically, this result would be interpreted as

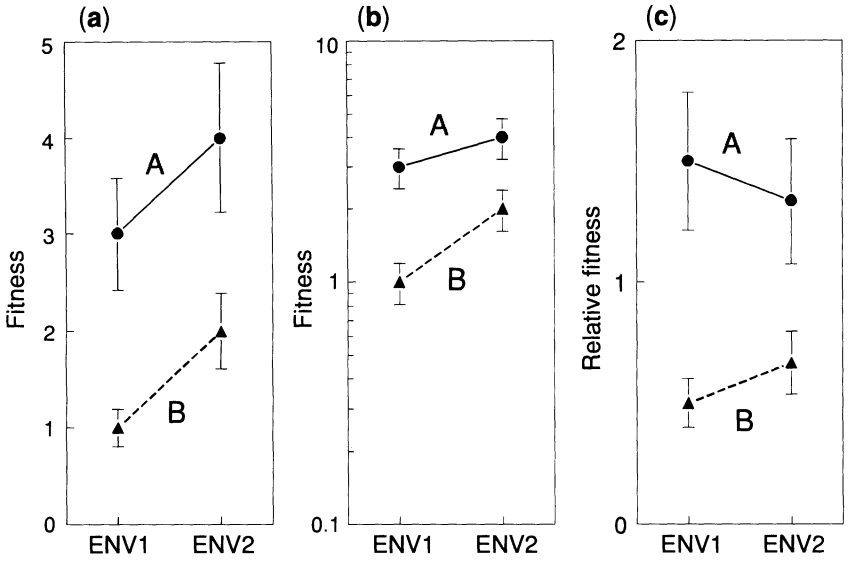

Fig. 1 Case #1 – additive effects of environment on genotype fitness. Norms of reaction for the mean fitness (± SD) of genotypes A and B, in response to 2 contrasting environments (ENV1 and ENV 2), are displayed in three formats. (a) Absolute fitness values are plotted on an arithmetic scale. (b) Absolute fitness values are plotted on a logarithmic scale. (c) Relative fitness (the fitness of each genotype, relative to the mean fitness averaged across genotypes, is plotted for each environment).

Table 1 Additive response, case #1: results of fixed-effects ANOVA for a hypothetical data set

	Fixed-effects ANOVA			Levene	Weighted ANOVA		
Source of variation	MS	*F*	*P*	*P*	MS	*F*	*P*
(a) Absolute fitness:			$R^2 = 0.82$				$R^2 = 0.84$
Genotype	80.00	281.48	< 0.0001	0.0008	288.89	281.48	< 0.0001
Environment	20.00	70.37	< 0.0001	0.0859	72.22	70.37	< 0.0001
G × E	0.00	0.00	1.0000	1.0000	0.00	0.00	1.0000
Error	0.28				1.03		
(b) Log (fitness +1):			$R^2 = 0.87$				
Genotype	7.13	380.77	< 0.0001	0.1400			
Environment	1.94	103.76	< 0.0001	0.4051			
G × E	0.16	8.67	0.0043	0.6620			
Error	0.02						
(c) Relative fitness:			$R^2 = 0.81$				$R^2 = 0.82$
Genotype	13.89	310.46	< 0.0001	0.0005	318.63	310.46	< 0.0001
Environment	0.00	0.00	1.0000	1.0000	0.00	0.00	1.0000
G × E	0.56	12.42	0.0007	0.4671	12.75	12.42	0.0007
Error	0.04				1.03		

Significance of heteroscedasticity is indicated by the *P*-value from the Levene test. Weighted mixed-model ANOVA is presented as an alternative analysis for significance testing when heteroscedasticity is indicated. Three measures of fitness were used as outcome variables in separate analyses: (a) absolute fitness (b) log-transformed absolute fitness, and (c) relative fitness. Mean fitness (± SD) for 20 replicates of the two genotypes in each environment are graphically shown in Fig. 1(a). DF = 1, 76 for each factor in the ANOVA model. Mean squares are presented for constructing alternative *F*-tests under a random effects model.

evidence that there is no genetic variation for fitness plasticity, i.e. that A and B are equally responsive to the two environments.

To meet the statistical assumption of equality of variances in ANOVA, an evolutionary ecologist might use log-transformation to eliminate the significant heteroscedasticity among genotypes indicated by the Levene test (Table 1a). Log-transformation does decrease variance heterogeneity (Table 1b), but also changes the significance of the G × E interaction in the ANOVA. The main effects of genotype and environment continue to be significant, but a significant G × E interaction now highlights the fact that genotypes show proportionately different responses to the two environments. The fitness of *A* increases by 33% between environments 1 and 2, while that for genotype B increases by 100%. This difference in proportional sensitivity is shown by nonparallel norms of reaction when fitness is plotted on a logarithmic scale (Fig. 1b).

An alternative way to examine proportional differences in fitness is to analyse relative fitness, calculated by dividing an individual's performance by the mean performance in each environment. This transformation eliminates mean fitness differences between environments. In the ANOVA based on relative fitness (Table 1c), the significant main effect of genotype

indicates that genotypes differ in average relative fitness across environments, while a significant G × E interaction indicates that selection will act differently within the two environments (also see Fig. 1c). Significant heteroscedasticity requires that an alternative method of significance testing be employed, but the weighted ANOVA shows the same pattern of significance (Table 1c). In these hypothetical cases, where all factors are considered fixed, alternative methods of significance-testing have little impact on the interpretations. This is unlikely to be the case in mixed model analyses, where the testing of F-ratios for fixed effects will tend to be overly sensitive to data in treatments causing high variability (Neter *et al.*, 1996).

An evolutionary ecologist is interested not only in the evolutionary consequences of environmental heterogeneity, but also in the sensitivity of this outcome to the frequency of the different selective environments. A main effect of genotype suggests that one genotype should replace the other when the environments are equally frequent, and for all three fitness measures we predict the same outcome: genotype A will replace genotype B (Table 1). The sensitivity of this outcome to the frequency of the environments is indicated by the G × E interaction, and this is where transforming fitness has dramatic consequences. While the absence of a significant G × E term for absolute fitness suggests that the frequency of environments will not affect the outcome of selection (Table 1a), highly significant G × E interaction for both proportional fitness transformations suggests that the outcome will be sensitive to the frequency of the two environments (Table 1b,c).

Over multiple generations, the clonal selection model shows that the frequency of genotype A increases monotonically and at a nearly identical rate under both hard and soft selection (Fig. 2), but also demonstrates that the population's expected trajectory depends on the distribution of genotypes across the alternative environments. When environment 1 is more frequent, clonal selection results in a more rapid increase in frequency of genotype A. The absence of G × E interaction in the ANOVA based on absolute fitness fails to predict this sensitivity to environmental frequencies. In contrast, ANOVA models based on the proportional fitness measures, log-transformed and relative fitness, qualitatively agree in revealing a significant main effect of genotype (indicating that genotype A will replace B), and a significant G × E interaction (indicating that the fitness advantage of A should depend on environmental frequency).

Although the ANOVA models based on log-transformed and relative fitness are in *qualitative* agreement with predictions of the clonal selection model in this case, we also seek to understand how the three alternative fitness measures vary in their *quantitative* relationship to the rate of genotype frequency change predicted under clonal selection. A simple calculation can show why log-transformed fitness values fail to mirror the quantitative evolutionary dynamics of the study system under either hard or soft selection. For haploid selection or clonal selection between two genotypes, the proportional change in frequency per generation is given by the proportional difference

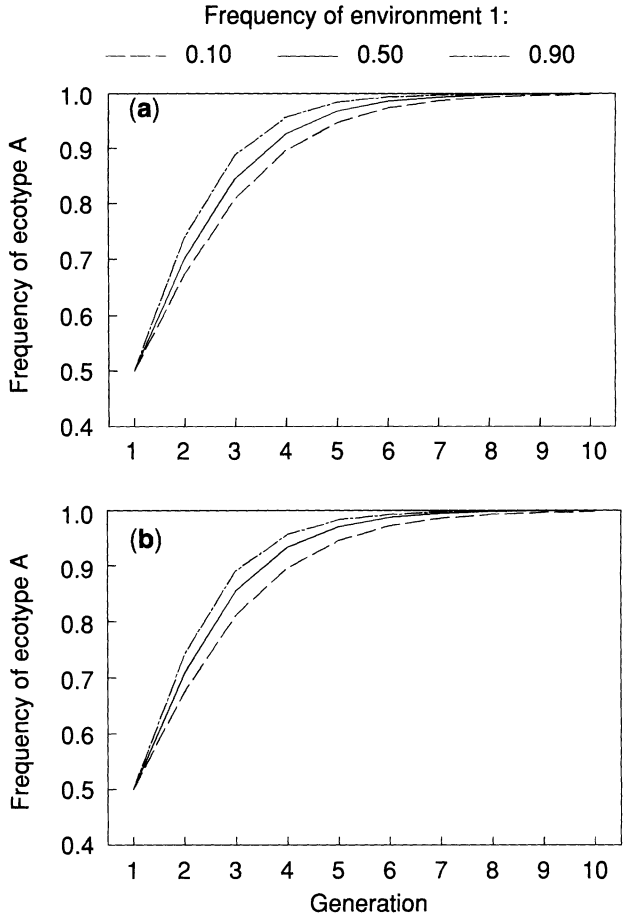

Fig. 2 Case #1 – the expected change in frequency of genotype A during 10 generations of clonal selection in a habitat that is divided between two environmental selection regimes. Genotypes are assigned the mean fitness values shown in Fig. 1(a). The selection model assumes that genotypes A and B begin at equal frequency in each environment. Three curves in each panel show the expected trajectory for three different frequencies of the alternative environments under: (a) hard selection and (b) soft selection. For both selection scenarios, migration equalizes genotype frequencies across environments in each generation.

in genotype fitness; this can be calculated as the difference in average fitness between the genotypes, divided by the sum of their average fitness. To show how fitness transformation influences this proportional difference, we divided the difference between genotype means, averaged across environments, for each of the three fitness measures (absolute, log-transformed, and relative) by the sum of the two genotype means for that same fitness measure. This proportional difference between genotype fitness values is then multiplied by the starting frequency of genotype A to calculate the predicted change in genotype frequency after one generation. For simplicity, we examine only the situation in which the two environments are encountered with equal frequency and the two genotypes begin at equal frequency. In Table 2, these calculated frequency changes are compared to the genotype frequency changes predicted by the clonal selection model.

Table 2 Predicted changes in the frequency of genotype A for four hypothetical fitness scenarios. (a) Additive case #1 (see Figs 1 and 2). (b) Multiplicative case #2 (see Figs 3 and 4). (c) Change in genotype rank case #3 (see Figs 5 and 6). (d) Change in genotype rank case #4 (see Figs 7 and 8)

	Frequency of genotype A from the clonal selection model:			Predicted change in frequency of genotype A based on proportional difference between genotype means for:		
	generation 0	generation 1	change	absolute fitness	ln (fitness +1)	relative fitness
(a) Case #1:						
Soft selection	0.500	0.708	**+0.208**	+0.200	+0.126	**+0.208**
Hard selection	0.500	0.700	**+0.200**	**+0.200**	+0.126	+0.208
(b) Case #2						
Soft selection	0.500	0.750	**+0.250**	**+0.250**	+0.150	**+0.250**
Hard selection	0.500	0.750	**+0.250**	**+0.250**	+0.150	**+0.250**
(c) Case #3						
Soft selection	0.500	0.452	**−0.048**	0.000	−0.019	**−0.048**
Hard selection	0.500	0.500	**0.000**	**0.000**	−0.019	−0.048
(d) Case #4						
Soft selection	0.500	0.500	**0.000**	+0.083	+0.015	**0.000**
Hard selection	0.500	0.583	**+0.083**	**+0.083**	+0.015	0.000

For each case, the frequency of genotype A begins at 0.50 in generation 0. The new genotype frequency and the change in frequency after one generation of soft or hard selection are derived from a simple clonal selection model in which the two test environments are equally frequent. For hard and soft selection in each case, these expected changes in frequency (shown in bold) are compared to the frequency change calculated using proportional differences between genotype means for three fitness variables (absolute fitness, log-transformed fitness, and relative fitness) that could be used as outcome variables in ANOVA (see text). For each selection scenario, the proportional fitness difference that that matches the change in the frequency of genotype A under clonal selection is also shown in bold.

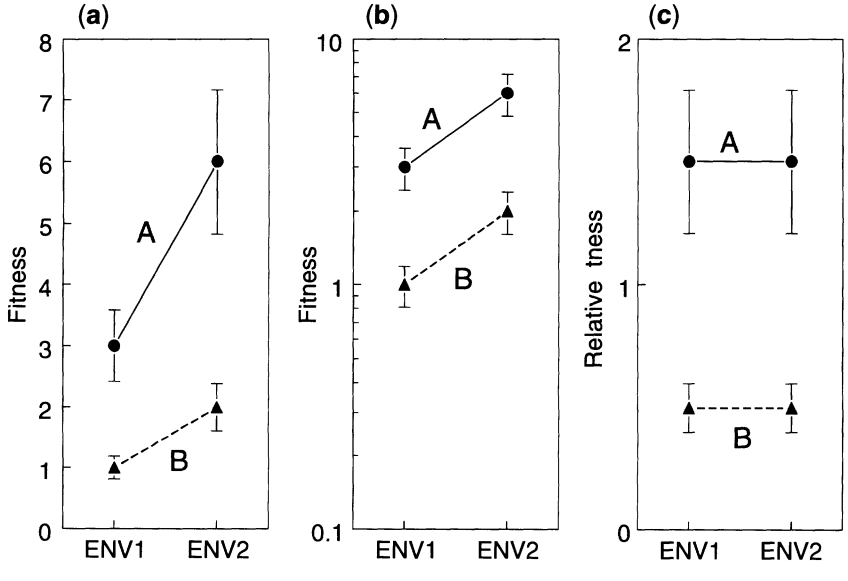

Fig. 3 Case #2 – proportionate effects of environment on genotype fitness. Norms of reaction for the mean fitness of genotypes A and B, in response to two contrasting environments are displayed in three formats. (a) Absolute fitness values are plotted on an arithmetic scale. (b) Absolute fitness values are plotted on a logarithmic scale. (c) Relative fitness (the fitness of each genotype, relative to the mean fitness averaged across genotypes), is plotted for each environment.

The comparison for Case #1 (Table 2a) shows that the single-generation increase in the frequency of genotype A under hard selection is predicted from absolute fitness values, whereas relative fitness values predict the rate of genotype frequency change under soft selection. The proportional difference in log-transformed fitness between genotypes substantially underestimates the rate at which genotype A will increase in frequency.

Proportional response, Case #2

As in Case #1, this scenario is one in which environment 2 is more favourable than environment 1 for both genotypes, leading to a significant main effect of environment in an ANOVA based on absolute fitness (Fig. 3; Table 3a). Genotype A has greater fitness in both environments, as indicated by a significant main effect of genotype. However, the effects of environment

Table 3 Proportional response, case #2: results of fixed-effects ANOVA for a hypothetical data set

	Fixed-effects ANOVA			Levene	Weighted ANOVA		
Source of variation	MS	*F*	*P*	*P*	MS	*F*	*P*
(a) Absolute fitness:			$R^2 = 0.89$				$R^2 = 0.89$
Genotype	180.00	380.00	< 0.0001	0.0001	390.00	380.00	< 0.0001
Environment	80.00	168.89	< 0.0001	0.0086	173.33	168.89	< 0.0001
G × E	20.00	42.22	< 0.0001	0.1817	43.33	42.22	< 0.0001
Error	0.47				1.03		
(b) Log (fitness +1):			$R^2 = 0.92$				
Genotype	11.69	592.28	< 0.0001	0.0972			
Environment	4.59	232.57	< 0.0001	0.3011			
G × E	0.12	6.04	0.0162	0.8395			
Error	0.02						
(c) Relative fitness:			$R^2 = 0.85$				$R^2 = 0.85$
Genotype	20.00	422.22	< 0.0001	< 0.0001	433.33	422.22	< 0.0001
Environment	0.00	0.00	1.0000	1.0000	0.00	0.00	1.0000
G × E	0.00	0.00	1.0000	1.0000	0.00	0.00	1.0000
Error	0.05				1.03		

Mean fitness (± SD) for 20 replicates of the two genotypes in each environment are graphically shown in Fig. 3(a). Where Levene's test detects significant heteroscedasticity, a weighted ANOVA is presented as an alternative model for hypothesis-testing (see text).

and genotype Are no longer strictly additive, as indicated by the significant G × E interaction. The nonparallel reaction norms for fitness plotted on an arithmetic scale (Fig. 3a) suggest that genotype A is more sensitive to the difference in environmental quality, as it produces three additional propagules in environment 2, while genotype B produces only a single additional propagule in environment 2, compared with environment 1. At first glance, this highly significant G × E interaction might suggest that the success of the two genotypes will depend on the frequencies with which the alternative selection regimes are encountered.

An empiricist might be tempted to use log-transformation for statistical reasons to eliminate the significant heteroscedasticity for genotypes indicated by the Levene test (Table 3a). An interest in the proportional changes in fitness might also lead one to use log-transformation (e.g. Johnston & Schoen, 1994; Dean, 1995). Both genotypes show the same proportional change in fitness between environments, so that when mean fitness is plotted on a log scale, the norms of reaction for genotypes A and B are parallel (Fig. 4b). For this data set, log-transformation does eliminate heteroscedasticity, but there is little change in the significance of the main effects or interactions in ANOVA (Table 3).

When we plot relative fitness within each environment to examine proportional changes in fitness, we see that the relative fitness of genotypes A and B is the same in the two environments (Fig. 3c). In the ANOVA for relative fitness, the main effect of genotype remains significant, but the G × E interaction term is negligible (Table 3c). Thus, ANOVA based on all three fitness measures predicts that genotype A will eventually replace genotype B, as indicated by the significance of the main effect of genotype in all three models (Table 3a–c). However, the analyses of absolute and log-transformed fitness indicate sensitivity to environmental frequency (i.e. significant G × E interaction), whereas the ANOVA for relative fitness does not.

The ANOVA for relative fitness is the only model that correctly predicts both genotype performance and its insensitivity to environmental frequency. Over multiple generations, the clonal selection model shows that the outcome of selection is identical under hard and soft selection, and is not influenced by the frequencies of the alternative environments (Fig. 4). Genotype A shows the same rate of monotonic increase in environments 1 and 2, and the rate of genotype frequency change overall is unaffected by frequencies of the two environments. This insensitivity to environmental frequency contrasts with the significant G × E interaction term seen in ANOVA for both absolute and log-transformed fitness (Table 3a,b).

For this special (and admittedly unrealistic) case of perfectly proportionate responses to environmental heterogeneity, the proportional differences between genotype means calculated from both relative and absolute fitness values accurately predict the rate at which genotype A increases in frequency (Table 2b). This is to be expected, since the trajectories for hard and soft selection do not differ in this case. However, the same calculation based on log-transformed fitness underestimates the selective advantage for genotype A (Table 2b). As for Case #1, log-transforming fitness leads to inaccurate biological interpretation.

Genotype rank changes, Case #3

Now we consider a case in which the fitness rankings of genotypes change between the two environments under

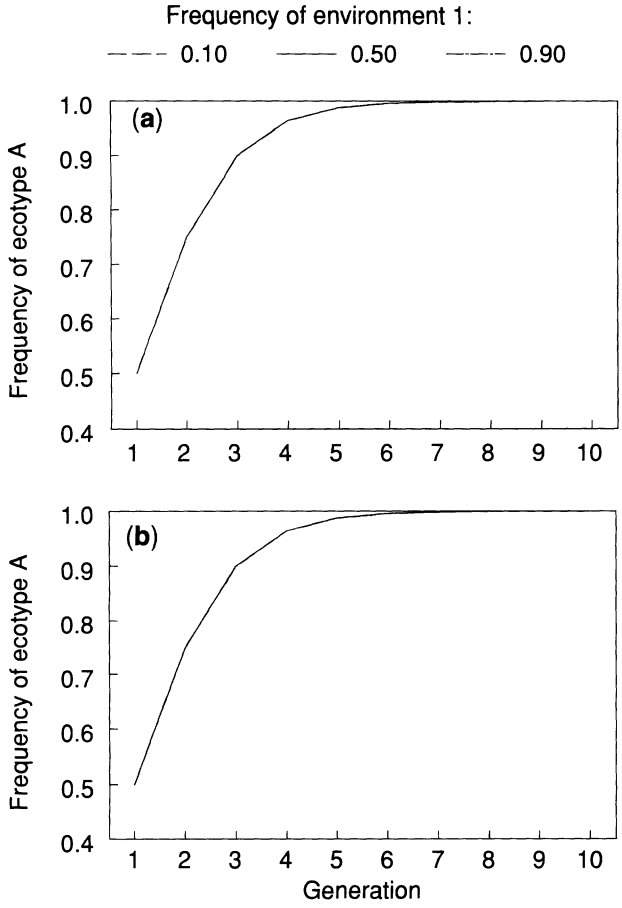

Fig. 4 Case #2 – the expected change in frequency of genotype A during 10 generations of clonal selection in a habitat that is divided between two environmental selection regimes. Genotypes are assigned the mean fitness values shown in Fig. 3(a). The selection model assumes that genotypes A and B begin at equal frequency in each environment. Because selection dynamics are identical in the two environments, a single curve shows the expected trajectory for 3 different frequencies of the alternative environments, under (a) hard selection and (b) soft selection.

consideration. In this hypothetical example, genotype A shows a 3-propagule increase in fitness between environments 1 and 2, whereas genotype B produces only one additional propagule in environment 2 (Fig. 5a). These are the same absolute changes in fitness used in Case #2 (Fig. 3), but now the norms of reaction cross, so that the rank order of genotype fitness reverses between environments. One might consider log-transformation based on: (i) statistical considerations, because significant heteroscedasticity among environments is indicated by Levene's test (Table 4a); or (ii) for biological purposes, to focus on proportional changes in fitness across environments. Analysis of relative fitness would be an alternative method for examining proportional changes in performance.

The choice of fitness measures used in ANOVA profoundly affects whether a main effect of genotype is detected. A main effect of genotype implies an overall fitness difference between genotypes, averaged across environments. Conversely, if environments are equally frequent, the absence of a main effect of genotype should indicate that selection will lead to no net change in genotype frequency. For Case #3, an ANOVA conducted on absolute fitness detects no difference in performance between genotypes (Table 4a); the arithmetic mean fitness of genotypes A and B are equal, so long as the genotypes are equally represented in both environments. In contrast, analyses of relative fitness and of log-transformed fitness both indicate a significant fitness advantage to genotype B, reflecting the fact that the relative fitness advantage experienced by genotype B in environment 1 is greater than the relative fitness advantage of genotype A in environment 2 (Fig. 5b,c). Not surprisingly, ANOVA conducted on any of the three fitness measures reveals significant G × E interaction (Table 4), indicating that the population's response to selection will depend on the distribution of genotypes across environments. However, because inferences about the eventual evolutionary outcome vary among ANOVA based on the three fitness measures, we use the clonal selection algorithms to compare the accuracy with which different ANOVA models predict evolutionary changes under hard and soft selection in a freely migrating population.

Under hard selection, the frequency of genotypes A and B will not change when environments are equally frequent (Fig. 6a). When environment 1 is more common than environment 2, genotype A will increase in frequency monotonically until it reaches fixation. When environment 2 is more common, genotype B will approach fixation at an even greater rate. Only the proportional differences in absolute fitness predict short-term changes in genotype frequency under hard selection (Fig. 6a, Table 2c), and only the ANOVA for absolute fitness accurately predicts the result of selection when the two environments are equally frequent.

When soft selection is operating, the evolutionary outcome and the rate of change differ from the hard selection case. When both environments are equally common, soft selection will cause the frequency of genotype A to decline (Fig. 6b), and only the proportional difference for relative fitness matches the single generation rate of genotype frequency change (Table 2c). When the frequency of environment 2 is greater than 50%, the overall frequency of genotype A increases monotonically from 0.5 to its final value, the frequency of environment 2. Conversely, when environment 1 is more common than environment 2, the frequency of genotype A rapidly declines towards the frequency of environment 2. As in previous cases, although log-transforming fitness reduces heteroscedasticity in the data (Table 4b), proportional differences in log-transformed fitness do not match changes in genotype frequency under either hard or soft selection (Table 2c).

Genotype rank changes, Case #4

Last, we consider a second scenario in which the fitness rankings of genotypes change between environments. As in all

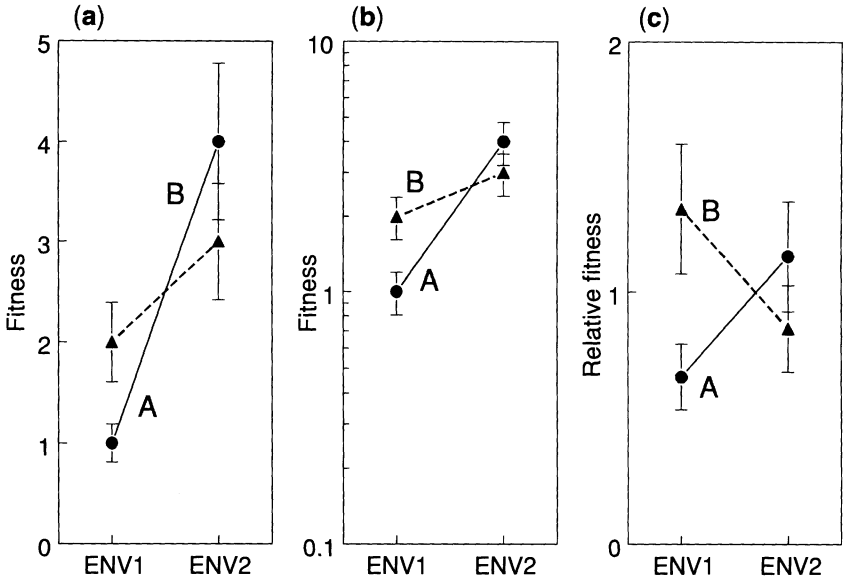

Fig. 5 Case #3 – arithmetically equal reversal of genotype fitness responses to two environments. Norms of reaction for the mean fitness of genotypes A and B, in response to 2 contrasting environments are displayed in three formats. (a) Absolute fitness values are plotted on an arithmetic scale. (b) Absolute fitness values are plotted on a logarithmic scale. (c) Relative fitness (the fitness of each genotype, relative to the mean fitness averaged across genotypes), is plotted for each environment.

the examples considered here, environment 2 is, on average, more favourable than environment 1 (Fig. 7a). However, genotype A experiences a much greater increase in absolute fitness in environment 2, compared with genotype B. The crossing norms of reaction indicate that the response to selection will depend on the distribution of genotypes across environments. A Levene's test based on the ANOVA for absolute fitness reveals highly significant hetreoscedasticty for the main effect of environment and for the G × E interaction, a violation of ANOVA assumptions that can be ameliorated by log-transforming individual fitness values (Table 5a,b). Plots of both log-transformed and relative fitness retain the crossing reaction norms (Fig. 7b,c).

Choosing to analyse different fitness measures dramatically affects whether a main effect of genotype is detected under this selection regime, but the pattern is the reverse of that seen in Case #3. Here, ANOVA on absolute fitness reveals a strong main effect of genotype, since the arithmetic mean fitness of genotype A exceeds that of genotype B when both environments are equally frequent. ANOVA on log-transformed fitness shows a smaller main effect of genotype. In contrast, an analysis of relative fitness within environments reveals no difference between genotypes A and B (Table 5c), since the relative fitness advantage of genotype B in environment 1 is equal to the relative fitness advantage of genotype A in environment 2 (Fig. 7c). These ANOVA results suggest different evolutionary outcomes, depending on the fitness measure. Based on analyses of absolute fitness and log-transformed fitness, we expect genotype A to replace B, although the rate at which it does so will be sensitive to the frequency of alternative environments. Based on the ANOVA on relative fitness, we expect no change in genotype frequency when environments are equally frequent.

As for Case #3, the outcome of clonal selection differs under hard and soft selection, and depends in both cases on the frequency of the environments (Fig. 8). Under hard

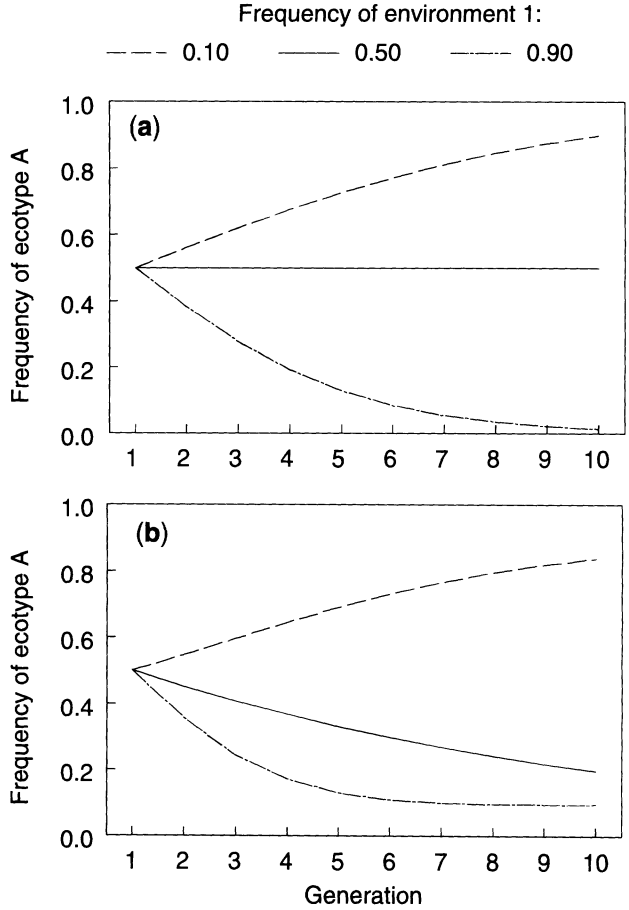

Fig. 6 Case #3 – the expected change in frequency of genotype A during 10 generations of clonal selection in a habitat that is divided between two environmental selection regimes. Genotypes are assigned the mean fitness values shown in Fig. 5(a). The selection model assumes that genotypes A and B begin at equal frequency in each environment. Three curves in each panel show the expected trajectory for three different frequencies of the alternative environments under (a) hard selection and (b) soft selection.

Table 4 Change in genotype rank, case #3: results of fixed-effects ANOVA for a hypothetical data set

	Fixed-effects ANOVA			Levene	Weighted ANOVA		
Source of variation	*MS*	*F*	*P*	*P*	*MS*	*F*	*P*
(a) Absolute fitness:			$R^2 = 0.82$				$R^2 = 0.84$
Genotype	0.00	0.00	1.0000	1.0000	0.00	0.00	1.0000
Environment	80.00	281.48	< 0.0001	0.0008	288.89	281.48	< 0.0001
G × E	20.00	70.37	< 0.0001	0.0859	72.22	70.37	< 0.0001
Error	0.28				1.03		
(b) Log (fitness +1):			$R^2 = 0.87$				
Genotype	0.16	8.67	0.0043	0.6620			
Environment	7.13	380.77	< 0.0001	0.1400			
G × E	1.94	103.76	< 0.0001	0.4051			
Error	0.02						
(c) Relative fitness:			$R^2 = 0.63$				$R^2 = 0.65$
Genotype	0.73	17.97	< 0.0001	0.3821	17.97	17.97	< 0.0001
Environment	0.00	0.00	1.0000	1.0000	0.00	0.00	1.0000
G × E	4.54	112.29	< 0.0001	0.0310	112.29	112.29	< 0.0001
Error	0.04				1.00		

Mean fitness (± SD) for 20 replicates of the two genotypes in each environment are graphically shown in Fig. 5(a). Where Levene's test detects significant heteroscedasticity, a weighted ANOVA is presented as an alternative model for hypothesis-testing (see text).

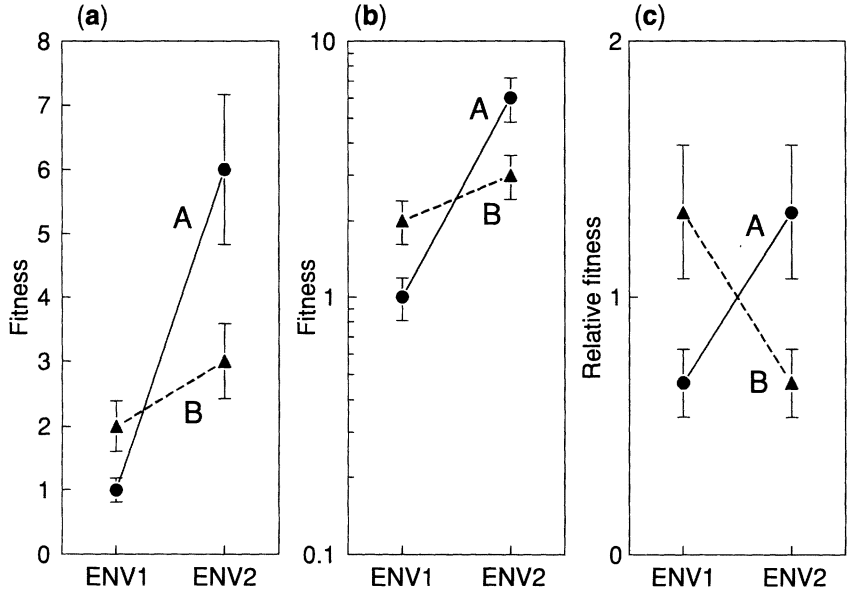

Fig. 7 Case #4 – proportionately equal reversal of genotype fitness responses to two environments. Norms of reaction for the mean fitness of genotypes A and B, in response to 2 contrasting environments are displayed in three formats. (a) Absolute fitness values are plotted on an arithmetic scale. (b) Absolute fitness values are plotted on a logarithmic scale. (c) Relative fitness (the fitness of each genotype, relative to the mean fitness averaged across genotypes) is plotted for each environment.

selection, genotype A replaces genotype B when environment 1 is equally frequent or less common, while B replaces A when environment 1 is more common (Fig. 8a). ANOVA based on both absolute fitness and log-transformed fitness predict this pattern *qualitatively*; the main effect of genotype is consistent with the outcome that one genotype replaces the other, while a significant G × E term is consistent with the fact that the outcome depends on environmental frequency. However, only the proportional difference in absolute fitness matches the rate of genotype frequency change in the first generation of hard selection (Table 2d). The same calculation for log-transformed fitness drastically underestimates the advantage accruing to genotype A.

In contrast, under soft selection, the clonal simulation predicts no change in genotype frequency when environments are equally frequent (Fig. 8b), a result that is consistent with the absence of a main effect of genotype in the ANOVA on relative fitness (Table 5c). Because population regulation occurs within each environment after selection, the contribution of each environment to the global pool of propagules is independent of mean fitness within environments. As a result, the increased frequency of genotype A propagules in environment

Table 5 Change in Genotype Rank Case #4: results of fixed-effects ANOVA for a hypothetical data set

	Fixed-effects ANOVA			Levene	Weighted ANOVA		
Source of variation	MS	*F*	*P*	*P*	MS	*F*	*P*
(a) Absolute fitness:			$R^2 = 0.89$				$R^2 = 0.89$
Genotype	20.00	42.22	< 0.0001	0.1817	43.33	42.22	< 0.0001
Environment	180.00	380.00	< 0.0001	0.0001	390.00	380.00	< 0.0001
G × E	80.00	168.89	< 0.0001	0.0086	173.33	168.89	< 0.0001
Error	0.47				1.03		
(b) Log (fitness +1):			$R^2 = 0.92$				
Genotype	0.12	6.04	0.0162	0.8395			
Environment	11.69	592.28	< 0.0001	0.0972			
G × E	4.59	232.57	< 0.0001	0.3011			
Error	0.02						
(c) Relative fitness:			$R^2 = 0.74$				$R^2 = 0.74$
Genotype	0.00	0.00	1.0000	1.0000	0.00	0.00	1.0000
Environment	0.00	0.00	1.0000	1.0000	0.00	0.00	1.0000
G × E	8.89	211.11	< 0.0001	0.0035	211.11	211.11	< 0.0001
Error	0.04				1.00		

Mean fitness (± SD) for 20 replicates of the two genotypes in each environment are graphically shown in Fig. 7(a). Where Levene's test detects significant heteroscedasticity, a weighted ANOVA is presented as an alternative model for hypothesis-testing (see text).

2 is exactly offset by selection against *A* in environment 1. Neither the analysis of absolute fitness nor of log-transformed fitness matches this prediction. Moreover, the rate of change in genotype frequency after one generation of soft selection uniquely matches that predicted by the proportional difference in relative fitness between genotypes (Table 2d). Patterns of selection, population regulation and migration that are intermediate between pure hard and soft selection will result in changes in genotype frequency that are bracketed by the two extremes modelled here.

Example analysis: genetic variation for fitness across light treatments

The amount of ambient light dramatically affected lifetime seed production in *Sinapis arvensis* (Fig. 9), as shown by the highly significant main effect of light treatment in the ANOVA based on absolute fitness (Table 6a). Due to an especially hot and dry spring, plants in full sun experienced drought stress, resulting in average seed production in the partial shade approximately three times that in full sun. This striking heterogeneity in environmental quality creates a situation in which the outcome of hard selection and of soft selection are likely to differ.

Variation in lifetime fertility among the six paternal half-sib families differed dramatically between light environments (absolute fitness: Fig. 9a). Plotting the norms of reaction for relative fitness (Fig. 9b) and for absolute fitness on a logarithmic scale (Fig. 9c) suggest that transformation will affect the variance among paternal half-sib families, as well as the G × E interaction term. Log-transformation reduces, but does not eliminate, heteroscedasticity in this data set (Table 6b). Evolutionary ecologists interested in estimating genetic variation for fitness within each light environment, or the extent to which different genotypes show local adaptation to light level, might consider relative fitness as the appropriate proportional measure, although this measure is also plagued by significant heteroscedasticity. We contrast the results of a weighted ANOVA on absolute, log-transformed, and relative fitness to illustrate how transformation affects our interpretation of these data.

First we determine whether there is a change in family fitness rankings between environments. A highly significant sire-by-light interaction in an ANOVA on fitness ranking within environments (d.f. = 5, 250; F = 4.63; $P = 0.0005$) indicates that a change in rank contributes to the G × E interaction. The weighted ANOVA based on absolute fitness reveals significant fitness variation among sires across the two environments, as well as significant genetic variation for plasticity in absolute fitness (Table 6a). Family S4 has by far the greatest mean fecundity, averaged across environments, and would be expected to increase in frequency under hard selection if sun and shade environments are equally frequent. The ANOVA based on relative fitness also reveals significant genetic variation for fitness among families (Table 6b). However, in contrast to the hard selection scenario, under soft selection we would expect that family S3 would increase in frequency when the environments are equally frequent, but by a narrow margin over families S4 and S5. These parallel analyses of absolute fitness and relative fitness suggest that different families will win under hard vs soft selection. However, the two analyses concur in calculating low fitness for families S1 and S6; these lineages should decrease in frequency under either

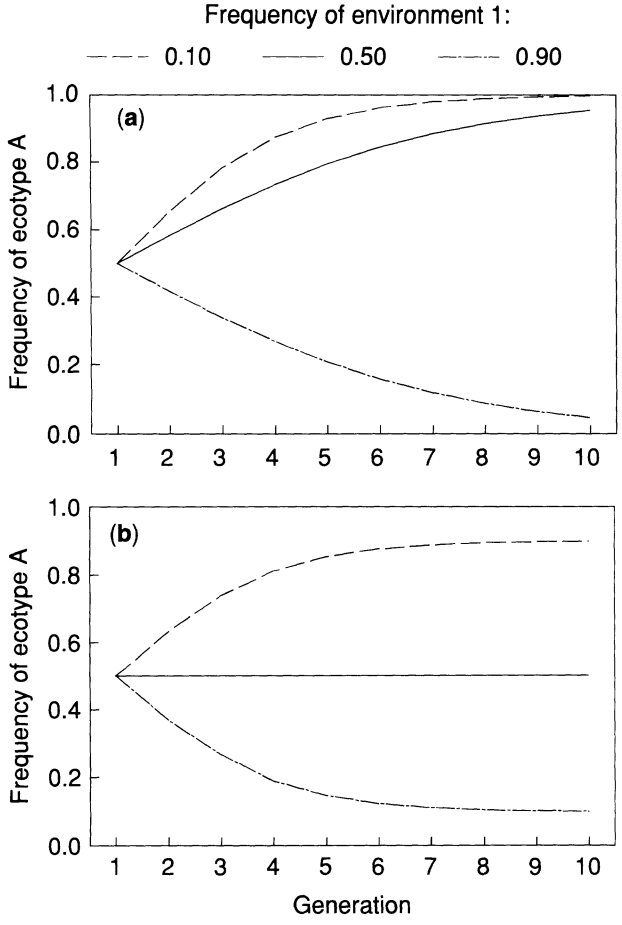

Fig. 8 Case #4 – the expected change in frequency of genotype A during 10 generations of clonal selection in a habitat that is divided between two environmental selection regimes. Genotypes are assigned the mean fitness values shown in Fig. 7(a). The selection model assumes that genotypes A and B begin at equal frequency in each environment. Three curves in each panel show the expected trajectory for three different frequencies of the alternative environments under (a) hard selection and (b) soft selection.

hard or soft selection regimes. Although we did not do so here, one could run these ANOVA models with different weights applied to observations in different environments to explore the breadth of light conditions under which these families would remain the least fit.

Based on our analyses of the four hypothetical cases, we would not recommend that investigators log-transform fitness in a data set like this one, if their aim is to use the results for make inferences about genotype performance, or about the sensitivity of selection to environmental frequencies. However, since most evolutionary ecologists would have log-transformed these fitness values to reduce heteroscedasticity, we show the results of that analysis here for comparison. Unlike either of the previous analyses, ANOVA on log-transformed fitness does not indicate significant genetic variation for plasticity in fitness and reveals only weakly significant genetic variation for fitness among families. In this analysis, family S5 has the greatest mean fecundity, averaged across environments. These patterns do not match the analyses of either absolute or relative fitness, and fitness rankings derived from the ANOVA on log-transformed individual fitness values do not correspond with either the hard or soft selection scenarios.

Closer examination of the fitness responses of individual families yields some insights into why the outcomes of hard and soft selection should be different in this study system. Note that family S3 has the highest fecundity in full sun and the lowest fecundity in partial shade. Because the shaded environment is of much higher quality overall quality, the expected fitness ranking of family S3 is lower under hard selection (Table 6a) than under soft selection (Table 6b). Similarly, family S4 should have an enormous fitness advantage under hard selection, despite its mediocre performance in full sun, because it has by far the highest fecundity in the high quality shade environment. Hard selection places a premium on strong performance under near-optimal conditions, whereas soft selection creates opportunities for genotypes that do well consistently, even under marginal conditions.

Discussion

Analysis of variance is a standard technique used by evolutionary ecologists to characterize the fitness responses of different genetic groups to contrasting treatments or to natural environmental variation. Although it is standard practice to transform fitness values in these analyses, if required to meet the assumptions of ANOVA, the consequences of fitness transformation for our evolutionary interpretation of such experiments are rarely discussed. Substantial heterogeneity for environmental quality is typically mirrored by substantial heteroscedasticity in fitness between environments, especially in plant ecological genetics studies, leading most investigators to log-transform fitness before conducting ANOVA. Through the use of heuristic data sets and simple simulations, we have shown here that log-transformation of fitness leads to inaccurate assessment of genotypic fitness variation and genotype-by-environment interaction in a wide range of selection scenarios. In other words, manipulating fitness data to meet the ANOVA assumption of variance homogeneity can compromise our evolutionary interpretations.

While there may be both statistical and biological justifications for transforming fitness in analysis of variance, choosing the most evolutionarily informative analysis depends on heterogeneity in environmental quality, as well as the scales at which selection operates and population size is regulated. When selection is occurring within microenvironments of different quality, the evolutionary consequences of hard selection and soft selection are likely to differ. In plant studies, these environmental effects are often dramatic. For example, in a recent study of phototropin mutants of *A. thaliana*, Galen

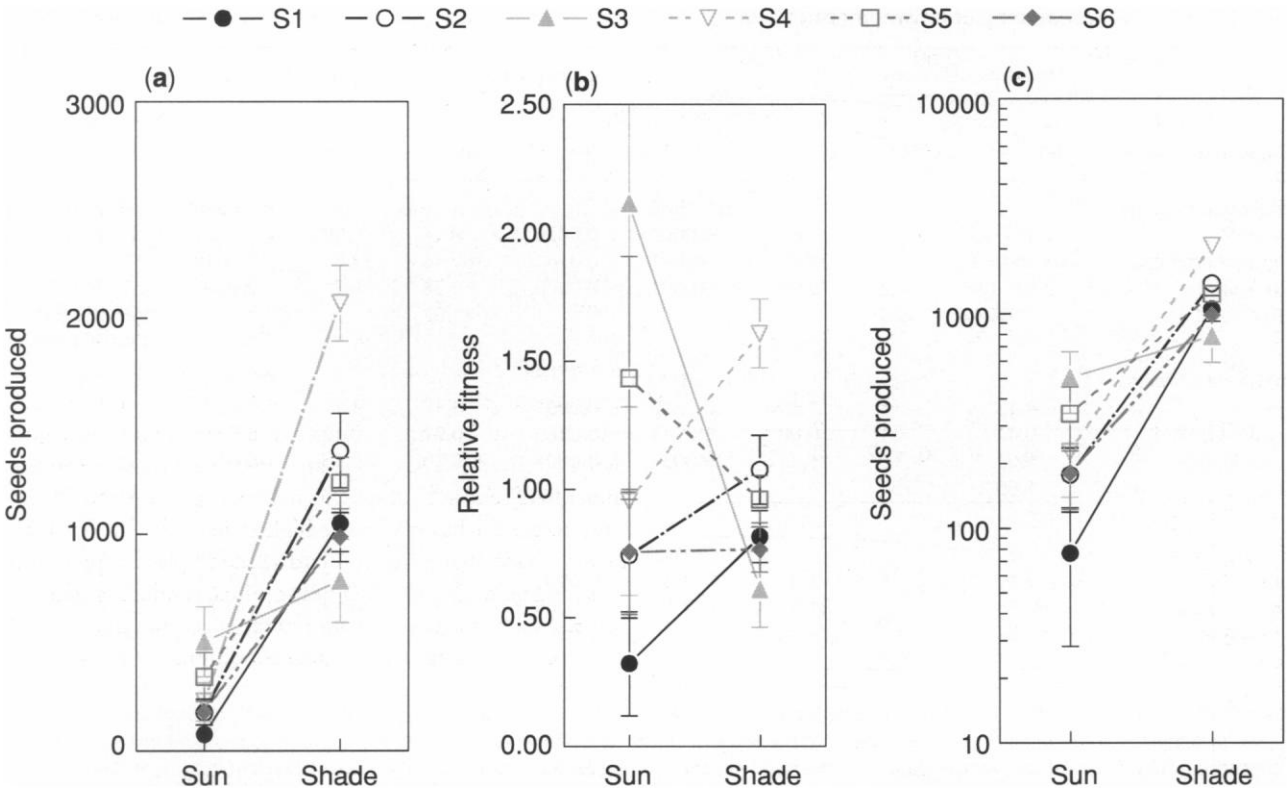

Fig. 9 Lifetime fertility of six paternal half-sibships of *Sinapis arvensis* in response to full sun and partial shade treatments. Neutral shade cloth ameliorated stress due to unusually hot and dry spring conditions in 1997. Norms of reaction for family fitness are displayed in three formats. (a) Sibship means for lifetime fertility are plotted on an arithmetic scale. (b) The relative fitnesses of each half-sibship are plotted for each light treatment. (c) Sibship means for lifetime fertility are plotted on a logarithmic scale.

et al. (2004) found that mean fitness (and fitness variance) varied by up to two orders of magnitude among shading and ground cover treatments in the field. Moreover, different components of fitness may be regulated in different ways (Van Tienderen, 1991). For example, hard selection might apply to female fecundity in species with highly dispersed seeds, whereas soft selection might dominate for juvenile survival or for competition among males for mating opportunities. Both the nature of the fitness component and differences in environmental quality are relevant when considering how the modes of population regulation influence the appropriate fitness measure.

Our analyses suggest that the G × E interaction term from an ANOVA on absolute fitness can fail to capture important evolutionary consequences of variation in environmental quality. When mean fitness varies among the environments being compared, then two genotypes that show the same absolute fitness change in response to contrasting environments will experience different relative fitnesses in those environments, even if their fitness rankings do not change (e.g. Case #1). Thus, even if one finds a null G × E interaction term based on an analysis of absolute fitness, the expected rate of genotype frequency change under selection will still be sensitive to the frequencies of alternative environments. Conversely, if environments vary in quality, but consistently ranked genotypes respond to that heterogeneity with proportionately equal changes in fitness, then relative fitnesses will remain unchanged across environments (e.g. Case #2). In this case, the G × E interaction term from an ANOVA on absolute fitness can be substantial, even though the evolutionary trajectory of the population is insensitive to the frequencies of alternative environments. In contrast, our analyses indicate that conducting ANOVA on relative fitness within environments yields a G × E interaction term that reliably indicates whether changing frequencies of alternative selection regimes will alter the population's genetic trajectory.

Despite potentially important differences between hard and soft selection, we rarely have complete knowledge of the spatial or temporal scales over which selection and population regulation occur. Our results suggest that conducting parallel analyses of absolute and relative fitness can be a useful way to bracket the likely genotype responses to heterogeneous environments based on experimental data. Across the test environments being considered, a main effect of genotype from an ANOVA on absolute fitness can predict differences in overall performance under hard selection, while that from ANOVA on relative fitness within environments can predict the outcome

Table 6 Example analysis: lifetime fecundity of six paternal sibships of Sinapis arvensis, grown under two levels of sunlight in the field

	Mixed-model ANOVA				Levene's test	Weighted ANOVA			
Source of variation	MS	d.f.	*F*	*P*	*P*	MS	*F*	*P*	Sire fitness rankings: LSM
(a) Absolute fitness:				$R^2 = 0.54$				$R^2 = 0.53$	S4: 1148.42a
Sire	2 167 136	5	6.61	< 0.0001	0.0076	6.14	5.99	< 0.0001	S5: 789.64b
Light treatment	63 554 881	1	24.48	0.0041	0.0141	189.46	33.61	0.0018	S2: 781.57b
Sire × light	2 627 086	5	8.01	< 0.0001	0.0220	6.36	6.21	< 0.0001	S3: 644.69b
Error	328 042	250				1.02			S6: 583.68b
									S1: 563.95b
(b) Relative fitness:				$R^2 = 0.11$				$R^2 = 0.17$	S3: 1.36a
Sire	4.13	5	2.50	0.0310	0.0001	3.17	3.17	0.0086	S4: 1.28a
Light treatment	0.42	1	0.08	0.7943	0.0185	0.22	0.12	0.7357	S5: 1.20a
Sire × light	5.63	5	3.41	0.0053	0.0001	2.35	2.35	0.0405	S2: 0.91a
Error	1.65	250				1.00			S6: 0.76a
									S1: 0.57a
(c) Ln (fitness +1):				$R^2 = 0.49$				$R^2 = 0.60$	S5: 5.29a
Sire	9.19	5	1.91	0.0934	0.0001	2.19	3.17	0.0556	S4: 5.20a
Light treatment	960.65	1	81.81	0.0002	0.0005	172.91	0.12	< 0.0001	S2: 4.78a
Sire × light	11.94	5	2.48	0.0325	0.0007	1.81	2.35	0.1120	S6: 4.70a
Error	4.81	250				1.00			S3: 4.39a

Sire is treated as a random effect in this mixed model ANOVA. In the presence of heteroscedasticity indicated by Levene's test, we use a weighted ANOVA for significance-testing, as described in the text. For analysis of absolute fitness, for example, the sire term shows the most significant heteroscedasticity. For each sire, we calculated the inverse of variance among residuals, and use those values to weight individual observations in a weighted least-squares ANOVA. For each model, the least-squares mean fitness values for the six paternal half-sibships (S1–S6) are shown in rank order in the right-hand column. Sibships with significantly different mean fitness at the 0.05 level, as demonstrated by Tukey's test, are followed by different letters.

of soft selection among genotypes. In *Sinapis arvensis* for example, parallel analyses indicate that some families would do poorly under either hard or soft selection across full sun and partial shade microenvironments. In contrast, rankings among the better-performing families shift dramatically between selection regimes. Genotypes that reproduce luxuriantly in the less stressful partial shade environment should have an advantage under hard selection, whereas genotypes that consistently maintain high relative fitness in both full sun and partial shade should increase in frequency under soft selection.

Although main effects of genotype are commonly reported and interpreted in ecological genetics studies, and are invaluable in teaching us about potential patterns of selection, it is sometimes not clear how these findings inform us about the evolutionary consequences of environmental heterogeneity in nature. This is because, in most experiments, investigators measure genotypic responses to controlled environments or to experimental treatments in the field, and any variation in fitness among genotypes revealed in ANOVA has predictive value only with respect to those specific environments. Because we still know relatively little about the scale of variation in relative fitness variation in nature (Barton & Turelli, 1989; Mitchell-Olds, 1992), a number of investigators are now advocating experiments in which genotype performance is measured over a range of environments that reflects natural environmental heterogeneity (e.g. Jackson *et al.*, 2002; Juenger & Bergelson, 2002; Sultan, 2003). For accurate evolutionary interpretation of such experiments, it is particularly important to analyse fitness measures that reflect patterns of selection.

Consider an experiment in which replicate plant genotypes are transplanted into a series of test gardens (blocks) dispersed at random across a natural environmental gradient spanned by a population (for examples of this approach, see Antonovics *et al.*, 1987; Bell & Lechowicz, 1991; Stratton, 1994; Stratton, 1995; Stratton & Bennington, 1996, 1998; Juenger & Bergelson, 2002). Each block represents one level of random environmental heterogeneity, and so a basic analysis of variance would include main effects of genotype (which could be either random or fixed), environmental block (a random factor), and the genotype-by-environment interaction. To the extent that blocks adequately sample the underlying heterogeneity in selection pressures, one could use ANOVA to infer whether there will be net selection favouring some genotypes over others (a main effect of genotype) or, alternatively, whether most variance in fitness occurs as G × E interaction, a situation that would tend to maintain genetic variation in the population. Our analyses suggest that log-transforming fitness to deal with heteroscedasticity would yield relatively uninformative ANOVA results. In

contrast, analysing both absolute fitness and relative fitness within blocks should reflect the outcome of hard vs soft selection, respectively, within the context of this natural environmental heterogeneity.

In plant populations, mean fitness and variance in fitness typically respond concordantly to environmental variation, so heterogeneity of fitness variances is a ubiquitous feature in studies of plant responses to contrasting conditions (e.g. Stanton *et al.*, in press). The existence of such heteroscedasticity poses challenges for investigators interested in testing for significant effects of genotype and G × E interaction on fitness variation. Log-transformation and rank-transformation are two common methods used to reduce heteroscedasticity (Potvin & Roff, 1993; Neter *et al.*, 1996), but both will compromise evolutionary interpretations of ecological genetic experiments, since tests based on reducing heteroscedasticity are problematic when treatment effects on variance are of intrinsic interest (Grissom, 2000). Log-transformation leads to serious biases in analyses of fitness variation because it discounts the true genetic contributions of very fit individuals, relative to those which are less fit. Rank transformation gives misleading results in data sets with strong nonadditive interactions (Seaman *et al.*, 1994) and, by its very nature, cannot possibly capture the full range and impact of fitness variation among individuals.

We have shown here that evolutionary interpretation can be enhanced by parallel analyses based on both absolute fitness values and relative fitness values within environmental treatments. When these relative or absolute fitness distributions show very unequal variances between environments, as is common for plants, weighting observations by the residual of variance within a given cell of the experimental design will more precisely estimate differences in means, and will reduce Type 1 error (Welch, 1951; Neter *et al.*, 1996). However, weighting will also tend to discount the fitness contributions of individuals in high variance environments, since these an be estimated with less precision, and so the use of weighting in ANOVA can produce estimates of least squared means (LSM) and standard errors (SE) for fitness that do not perfectly predict the expected dynamics of selection. As a check for such bias, one can compare the LSM and SE values obtained from weighted ANOVA to those calculated from raw values of absolute and relative fitness. If the differences are subtle, then the predictive value of the ANOVA should be relatively high.

If fitness parameter estimates from raw data and from weighted ANOVA are notably different, then investigators may need to use descriptive methods that are less restrictive in their assumptions about error variance distributions. The development of robust heteroscedastic methods, especially for datasets containing both fixed and random independent variables, is an active area of research in applied statistics (reviewed in Grissom, 2000), but most of these methods are not yet incorporated into software programs capable of handling complex experimental designs. Two distributions with different variances may not only differ with respect to their central location (mean or median), but also in their dispersion and symmetry, and any of these parameters may be of importance to evolutionary dynamics. Quantile regression, as applied within a general linear models framework, is a powerful technique that can be used to more fully describe such differences in distribution (see Cade & Noon, 2003), but will result in some loss of information because it is based on fitness ranks rather than on actual fitness values. Monte Carlo tests for significant differences in central location, including permutation and bootstrapping, are based on the assumption that samples are drawn from the same underlying distribution, but the compromising effects of heteroscedasticity can be reduced by having nearly equivalent and relatively large sample sizes in each group sampled (Hayes, 2000).

When applied to an appropriately robust experiment, Monte Carlo techniques can be useful in an evolutionary genetics framework, because they use the raw fitness values that determine the outcome of selection. For example, to evaluate effects of genotype and environment on fitness in a dataset containing at least 15 replicates within each genotype-environment combination, one could perform many iterations of the following procedure. First, bootstrap independently within each genotype-environment combination to construct a resampled dataset. Next, to describe absolute fitness differences between each pair of genotypes, calculate the difference in their mean (or total) fitness, and save those values to an output file. Once all iterations are complete, a given pair of genotypes can be judged to have unequal fitness if 95% of their difference values from bootstrapped datasets are either greater than or less than zero. The same basic approach can be used to test for differences in mean fitness between environments and for G × E interaction, and can also be used to analyse relative fitness within environments. As for quantile regression, this kind of resampling approach can quickly become cumbersome to implement and/or interpret if there are many genotypes and/or environments in the experimental design. There is no question that empirical work on plant fitness responses to heterogeneous environments will be significantly advanced by the increased implementation of user-friendly, robust methods within standard statistical software packages.

Although it is not very satisfying to elaborate a problem for which there is not yet an ideal solution, we feel that it is important for evolutionary ecologists to understand that there can be significant tension between our standard statistical techniques and our search for biological insight. Log-transforming fitness may reduce heteroscedasticity in fitness datasets, but also produces biased fitness estimates that do not predict genetic contributions to future generations. For the same reason, transforming fitness is not recommended when calculating selection gradients or differentials because it alters the relationship between phenotype and fitness, obscuring the

nature of selection (Lande & Arnold, 1983; Arnold & Wade, 1984a,b). The key issue to keep in mind is that transforming variables to meet assumptions of ANOVA or regression complicates interpretation when the raw data are already expressed on a scale that is of inherent biological interest. Log-transformation continues to be a valuable technique when analysing variables that show mean-variance correlation, and for which there is no *a priori* reason to favour one scale of measurement over another. In evolutionary analyses, because it is fitness itself that determines the outcome of natural selection, investigators should employ statistical techniques that do not require fitness transformation.

Acknowledgements

This paper is dedicated to the dozens of colleagues with whom we have shared the agonies and frustrations of analysing plant fitness data over the past decade. Our ideas were especially clarified by discussions with Sharon Strauss, Laura Galloway, Anurag Agrawal, Bitty Roy, Jessica Wright, David Posner, David Boose, Sarah Eppley, Jenny Davidson, Caroline Christian, Eric Baack, Ken Whitney, Nancy Emery, Lizzie King, and Michael Turelli. We are very grateful to statisticians Neil Willits, Brian Cade, and Cliff Lunneborg for advice on ways to deal with heteroscedastic data when testing hypotheses. Sharon Strauss, Rick Grosberg, John Gillespie, Michael Wade, Joe Travis, and anonymous reviewers provided constructive criticisms on earlier drafts of the manuscript. This work was supported by NSF grants IBN 94-19800 and DEB 97-26663, and by a Bridge Grant from the University of California Davis.

References

Allard RW, Bradshaw AD. 1964. Implications of genotype–environmental interactions in applied plant breeding. *Crop Science* **4**: 503–508.

Andersson S, Shaw RG. 1994. Phenotypic plasticity in *Crepis tectorum* (Asteraceae): Genetic correlations across light regimens. *Heredity* **72**: 113–125.

Antonovics J, Bradshaw AD, Turner RG. 1971. Heavy metal tolerance in plants. *Advances in Ecological Research* **7**: 1–85.

Antonovics J, Clay K, Schmitt J. 1987. The measurement of small-scale environmental heterogeneity using clonal transplants of *Anthoxanthum odoratum* and *Danthonia spicata*. *Oecologia (Berlin)* **71**: 601–607.

Arnold SJ, Wade MJ. 1984a. On the measurement of natural and sexual selection: Theory. *Evolution* **38**: 709–719.

Arnold SJ, Wade MJ. 1984b. On the measurement of natural and sexual selection: Applications. *Evolution* **38**: 720–734.

Ballare CL, Scopel AL, Jordan ET, Vierstra RD. 1994. Signaling among neighboring plants and the development of size inequalities in plant populations. *Proceedings of the National Academy of Sciences, USA* **91**: 10094–10098.

Barton NH, Turelli M. 1989. Evolutionary quantitative genetics: how little do we know? *Annual Review of Genetics* **23**: 337–370.

Bell G, Lechowicz MJ. 1991. The ecology and genetics of fitness in forest plants. I. Environmental heterogeneity measured by explant trials. *Journal of Ecology* **79**: 663–685.

Bennett BD, Grace J. 1988. Shade tolerance and its effect on the segregation of two species of Louisiana *Iris* and their hybrids. *American Journal of Botany* **77**: 100–107.

Bennington CC, Thayne WV. 1994. The use and misuse of mixed model analysis of variance in ecological studies. *Ecology* **75**: 717–722.

Björkman O, Holmgren P. 1963. Adaptability of the photosynthetic apparatus to light intensity in ecotypes from exposed and shaded habitats. *Physiologia Plantarum* **16**: 889–913.

Bolnick DI. 2004. Can intraspecific competition drive disruptive selection? An experimental test in natural populations of sticklebacks. *Evolution* **58**: 608–618.

Bradshaw AD. 1965. Evolutionary significance of phenotypic plasticity in plants. *Advances in Genetics* **13**: 115–155.

Cade BS, Noon BR. 2003. A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment* **1**: 412–420.

Caillaud MC, Via S. 2000. Specialized feeding behavior influences both ecological specialization and assortative mating in sympatric host races of pea aphids. *American Naturalist* **156**: 606–621.

Christiansen FB. 1975. Hard and soft selection in a subdivided population. *American Naturalist* **109**: 11–16.

Dean AM. 1995. A molecular investigation of genotype by environment interactions. *Genetics* **139**: 19–33.

Dempster ER. 1955. Maintenance of genetic heterogeneity. *Cold Springs Harbor Symposia in Quantitative Biology* **20**: 25–32.

Dingle H. 1992. Food level reaction norms in size-selected milkweed bugs (*Oncopeltus fasciatus*). *Ecological Entomology* **17**: 121–126.

Donohue K, Messiqua D, Pyle EH, Heschel MS, Schmitt J. 2000a. Evidence of adaptive divergence in plasticity: Density- and site-dependent selection on shade-avoidance responses in *Impatiens capensis*. *Evolution* **54**: 1956–1968.

Donohue K, Pyle EH, Messiqua D, Heschel MS, Schmitt J. 2000b. Density dependence and population differentiation of genetic architecture in *Impatiens capensis* in natural environments. *Evolution* **54**: 1969–1981.

Dutilleul P, Potvin C. 1995. Among-environment heteroscedasticity and genetic autocorrelation: Implications for the study of phenotypic plasticity. *Genetics* **139**: 1815–1829.

Falconer DS. 1989. *Introduction to Quantitative Genetics*. New York, USA: Longman Scientific & Technical and. John Wiley & Sons.

Felsenstein J. 1976. The theoretical population genetics of variable selection and migration. *Annual Review of Genetics* **10**: 253–280.

Fry JD. 1992. The mixed-model analysis of variance applied to quantitative genetics: Biological meaning of the parameters. *Evolution* **46**: 540–550.

Galen C, Huddle J, Liscum E. 2004. An experimental test of the adaptive evolution of phototropins: Blue-light photoreceptors controlling phototropism in *Arabidopsis thaliana*. *Evolution* **58**: 515–523.

Galloway L. 1995. Response to natural environmental heterogeneity: Maternal effects and selection on life-history characters and plasticities in *Mimulus guttatus*. *Evolution* **49**: 1095–1107.

Gillespie J. 1974. Polymorphism in patchy environments. *American Naturalist* **108**: 145–151.

Gillespie JH, Turelli M. 1989. Genotype–environment interactions and the maintenance of polygenic variation. *Genetics* **121**: 129–138.

Grissom RJ. 2000. Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology* **68**: 155–165.

Hatfield T, Schluter D. 1999. Ecological speciation in sticklebacks: Environment-dependent hybrid fitness. *Evolution* **53**: 866–873.

Hayes AF. 2000. Randomization tests and the quality of variance assumption when comparing group means. *Animal Behaviour* **59**: 653–656.

Hoffman AA, Watson M. 1993. Geographical variation in the acclimation responses of *Drosophila* to temperature extremes. *American Naturalist* **142**: S93–S113.

Jackson RB, Linder CR, Lynch M, Purugganan M, Somerville S, Thayer SS. 2002. Linking molecular insight and ecological research. *Trends in Ecology and Evolution* **17**: 409–414.

Johnston MO, Schoen DJ. 1994. On the measurement of inbreeding depression. *Evolution* **48**: 1735–17421.

Juenger T, Bergelson J. 2002. The spatial scale of genotype by environment interaction (GEI) for fitness in the loose-flowered gilia, *Ipomopsis laxiflora* (Polemoniaceae). *International Journal of Plant Sciences* **163**: 613–618.

Lande R, Arnold SJ. 1983. The measurement of selection on correlated characters. *Evolution* **37**: 1210–1226.

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

Levene H. 1960. Robust tests for equality of variance. In: Olkin I, ed. *Contributions to Probability and Statistics.* Palo Alto, California, USA: Stanford University Press, 278–292.

Levins R. 1968. *Evolution in Changing Environments.* Princeton, CT, USA: Princeton University Press.

Lively CM. 1989. Adaptation by a parasitic trematode to local populations of its snail host. *Evolution* **43**: 1663–1671.

Lynch M, Walsh B. 1998. *Genetics and Analysis of Quantitative Traits.* Sunderland, MA, USA: Sinauer Associates.

Mitchell-Olds T. 1992. Does environmental variation maintain genetic variation? A question of scale. *Trends in Ecology and Evolution* 7: 397–398.

Neter J, Kutner MH, Nachtsheim CJ, Wasserment W. 1996. *Applied Linear Statistical Models.* Boston, MA, USA: McGraw-Hill.

Newman JA, Bergelson J, Grafen A. 1997. Blocking factors and hypothesis tests in ecology: Is your statistics text wrong? *Ecology* **78**: 1312–1320.

Peichel CL, Nereng KS, Ohgi KA, Cole BLE, Colosimo PF, Buerkle CA, Schluter D, Kingsley DM. 2001. The genetic architecture of divergence between threespine stickleback species. *Nature (London)* **414**: 901–905.

Pigliucci M, Schmitt J. 2004. Phenotypic plasticity in response to foliar and neutral shade in gibberellin mutants of *Arabidopsis thaliana. Evolutionary Ecology Research* **6**: 243–259.

Potvin C, Roff DA. 1993. Distribution-free and robust statistical methods: Viable alternatives to parametric statistics? *Ecology* **74**: 1617–1628.

Prout T, Barker JSF. 1989. Ecological aspects of the heritability of body size in *Drosophila buzzatii. Genetics* **123**: 803–814.

Roy BA, Stanton ML. 1999. Asymmetry in wild mustard, *Sinapis arvensis* (Brassicaceae), in response to severe physiological stresses. *Journal of Evolutionary Biology* **12**: 440–449.

Roy BA, Stanton ML, Eppley SM. 1999. Effects of environmental stress on leaf hair density and consequences for selection. *Journal of Evolutionary Biology* **12**: 1089–1103.

SAS Institute. 1999. *SAS/STAT User's Guide,* release 8. ed. Cary, NC, USA: SAS Institute.

Schemske DW. 1984. Population structure and local selection in *Impatiens* (Balsaminaceae), a selfing annual. *Evolution* **40**: 817–829.

Schmalhausen II. 1949. *Factors of Evolution.* Philadelphia, USA: Blakiston.

Schmitt J, McCormac AC, Smith H. 1995. A test of the adaptive plasticity hypothesis using transgenic and mutant plants disabled in phytochrome-mediated elongation responses to neighbors. *American Naturalist* **146**: 937–953.

Seaman JW, Walls SC, Wise SE, Jaeger RG. 1994. Caveat emptor: rank transform methods and interaction. *Trends in Ecology and Evolution* **9**: 261–263.

Shaw RG, Platenkamp GJ, Shaw FH, Podolsky RH. 1995. Quantitative genetics of response to competitors in *Nemophila menziesii*: a field experiment. *Genetics* **139**: 397–406.

Simms EL, Triplett J. 1994. Costs and benefits of plant responses to disease: Resistance and tolerance. *Evolution* **48**: 1973–1985.

Snedecor GW, Cochran WG. 1989. *Statistical Methods.* Ames, IA, USA: Iowa State University Press.

Sokal RR, Rohlf FJ. 1995. *Biometry.* New York, USA: W.H. Freeman and Company.

Stanton M, Roy B, Thiede D. 2000. Evolution in stressful environments. I. Phenotypic variability, phenotypic selection, and response to selection in five distinct environmental stresses. *Evolution* **54**: 93–111.

Stanton ML, Thiede DA, Roy BA. in press. Consequences of intraspecific competition and environmental variation for selection in the mustard *Sinapis arvensis*: contrasting ecological and evolutionary perspectives. *American Naturalist* **in press**.

Stewart SC, Schoen DJ. 1987. Pattern of phenotypic viability and fecundity selection in a natural population of *Impatiens pallida. Evolution* **41**: 1290–1301.

Stratton DA. 1994. Genotype-by-environment interactions for fitness of *Erigeron annuus* show fine-scale selective heterogeneity. *Evolution* **48**: 1607–1618.

Stratton DA. 1995. Spatial scale of variation in fitness of *Erigeron annuus. American Naturalist* **146**: 608–624.

Stratton DA, Bennington CC. 1996. Measuring spatial variation in natural selection using randomly-sown seeds of *Arabidopsis thaliana. Journal of Evolutionary Biology* **9**: 215–228.

Stratton DA, Bennington CC. 1998. Fine-grained spatial and temporal variation in selection does not maintain genetic variation in *Erigeron annuus. Evolution* **52**: 678–691.

Sultan SE. 1992. Phenotypic plasticity and the neo-Darwinian legacy. *Evolutionary Trends in Plants* **6**: 61–71.

Sultan SE. 2003. Commentary: The promise of ecological developmental biology. *Journal of Experimental Zoology* **296B**: 1–7.

Sultan SE, Bazzaz FA. 1993. Phenotypic plasticity in *Polygonum persicaria.* II. Norms of reaction to soil moisture and the maintenance of genetic diversity. *Evolution* **47**: 1009–1031.

Van Tienderen PH. 1991. Evolution of generalists and specialists in spatially heterogeneous environments. *Evolution* **45**: 1317–1331.

Via S. 1991. The genetic structure of host plant adaptation in a spatial patchwork: demographic variability among reciprocally transplanted pea aphid clones. *Evolution* **45**: 827–852.

Via S. 1999. Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution* **53**: 1446–1457.

Via S, Bouck AC, Skillman S. 2000. Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution* **54**: 1626–1637.

Via S, Conner J. 1995. Evolution in heterogeneous environments: Genetic variability within and across different grains in *Tribolium castaneum. Heredity* 74: 80–90.

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

Via S, Lande R. 1987. Evolution of genetic variability in a spatially heterogeneous environment: Effects of genotype–environment interaction. *Genetical Research* **49**: 147–156.

Wallace B. 1981. *Basic Population Genetics.* New York, USA: Columbia University Press.

Welch BL. 1951. On the comparison of several mean values: An alternative approach. *Biometrika* **38**: 330–336.

Yamada Y. 1962. Genotype by environment and genetic correlation of the same trait under different environments. *Japanese Journal of Genetics* **37**: 489–509.

Zar JH. 1996. *Biostatistical Analysis.* Upper Saddle River, NJ, USA: Prentice Hall.