

Evolutionary Quantitative Genetics

Bruce Walsh

Depts. of Ecology and Evolutionary Biology¹,
Plant Sciences, and Molecular and Cellular Biology
University of Arizona, Tucson, AZ 85721

Phone: 520 621-1915

FAX: 520 621-9190

email: jbwalsh@u.arizona.edu

¹ Mailing address: Dept of Ecology and Evolutionary Biology, Biosciences West,
University of Arizona, Tucson, AZ 85721

CONTENTS

- 1 Introduction
 - 1.1. Resemblances, Variances, and Breeding Values
 - 1.2. Single Trait Parent-offspring Regressions
 - 1.3. Multiple Trait Parent-offspring Regressions
- 2 Selection Response Under the Infinitesimal Model
 - 2.1. The Infinitesimal Model
 - 2.2. Changes in Variances
 - 2.3. The Roles of Drift and Mutation Under the Infinitesimal Model
- 3 Fitnesses
 - 3.1. Individual Fitness
 - 3.2. Episodes of Selection
 - 3.3. The Robertson-Price Identity, $S = \sigma(w, z)$
 - 3.4. I , The Opportunity for Selection
 - 3.5. Some Caveats in Using the Opportunity for Selection
- 4 Fitness Surfaces
 - 4.1. Individual and Mean Fitness Surfaces
 - 4.2. Measures of Selection on the Mean
 - 4.3. Measures of Selection on the Variance
 - 4.4. Gradients and the Geometry of Fitness Surfaces
 - 4.5. Estimating the Individual Fitness Surface
 - 4.6. Linear and Quadratic Approximations of $W(z)$
- 5 Measuring Multivariate Selection
 - 5.1. Changes in the Mean Vector: The Directional Selection Differential S
 - 5.2. The Directional Selection Gradient β
 - 5.3. Directional Gradients, Fitness Surface Geometry and Selection Response
 - 5.4. Changes in the Covariance Matrix: The Quadratic Selection Differential C
 - 5.5. The Quadratic Selection Gradient γ
 - 5.6. Quadratic Gradients, Fitness Surface Geometry and Selection Response
 - 5.7. Estimation, Hypothesis Testing and Confidence Intervals
 - 5.8. Geometric Aspects
 - 5.9. Unmeasured Characters and Other Biological Caveats
- 6 Multiple Trait Selection

- 6.1. Short-Term Changes in Means: The Multivariate Breeders' Equation
- 6.2. The effects of genetic correlations: direct and correlated responses
- 6.3. Evolutionary constraints imposed by genetic correlations
- 6.4. Inferring the Nature of Previous Selection
- 6.5. Changes in G under the infinitesimal model
- 7 Phenotypic Evolution Models
 - 7.1. Selection vs. Drift in the fossil record
 - 7.2. Stabilizing Selection
- 8 Theorems of Natural Selection: Fundamental and Otherwise
 - 8.1. The Classical Interpretation of Fishers' Fundamental Theorem
 - 8.2. What Did Fisher Really Mean?
 - 8.3. Heritabilities of Characters Correlated With Fitness
 - 8.4. Robertson's Secondary Theorem of Natural Selection

Final Remarks and Acknowledgments

References

1. INTRODUCTION

Evolutionary quantitative genetics is a vast field, ranging from population genetics on one extreme to development and functional ecology on the other. The goal of this field is a deeper understanding of not only the genetics and inheritance of complex traits in nature (traits whose variation is due to both genetic and environmental factors) but also of the nature of the evolutionary forces that shape character variation and change in natural populations. Given the limitations of space, we have chosen to focus this review on evolutionary change, in particular estimation of the nature of selection and aspects of the evolutionary response to selection. A detailed treatment of the inheritance of complex traits in nature can be found in Lynch and Walsh (1998). Roff (1997) provides a good overview of the entire field, while the reader seeking detailed treatments on specific issues should consult Bulmer (1980) and Walsh and Lynch (2003). Bürger (2000) provides an excellent (but, in some places, highly technical) review of the interface between population and quantitative genetics.

1.1. Resemblances, Variances, and Breeding Values

A few summary remarks on quantitative genetics are in order to give the reader the appropriate background. A good introduction is Falconer and MacKay (1996), while a very detailed treatment is given by Lynch and Walsh (1998). Chapters by Gianola, Jansen, Hoeschele, and Whittaker in this Handbook also examine some of these issues in greater detail.

Fisher's Genetic Decomposition

Although Yule (1902) can be considered the first paper in quantitative genetics, the genesis of modern quantitative genetics is R. A. Fisher's classic 1918 paper. Fisher made a number of key insights, notably that (i) with sexual reproduction, only a specific fraction of an individual's genotypic value (the mean value of that genotype when averaged over the distribution of environments) is passed onto its offspring, (ii) we can estimate the variances associated with these various components by

looking at the phenotypic covariances between appropriate relatives.

Fisher decomposed the observed phenotypic value z of an individual into a genotypic G and environment E value,

$$z = G + E \quad (1a)$$

The genotypic value can be thought of as the average phenotypic value if the individual was cloned and replicated over the universe of environments it is likely to experience. As mentioned, Fisher noted that a parent does not pass along its entire G value to its offspring, as for any given locus a parent passes along only one of its two alleles to a particular offspring. Thus, the genotypic value can be further decomposed into a component passed on to the offspring A (the **additive genetic value**) and a non-additive component, which includes the dominance D and any epistatic effects. We ignore the effects of epistasis here, whose complex features are extensively examined by Lynch and Walsh (1998) and Walsh and Lynch (2003). Thus,

$$z = \mu + A + D + E \quad (1b)$$

where μ is the population mean (by construction, the mean values of A , D , and E equal zero). The additive values A are often referred to as **breeding values**, as the average value of an offspring is just the average breeding value of its parents. Hence, the expected value of an offspring is $\mu + (A_f + A_m)/2$, where A_f is the paternal (or sire) and A_m the maternal (or dam) breeding values.

Additive Genetic Variances and Covariances

When the phenotype can be decomposed as in (1b), Fisher showed that the phenotypic covariance between parent and offspring is half the population variance of breeding values, σ_A^2 ,

$$\sigma^2(z_p, z_o) = \frac{\sigma_A^2}{2} \quad (2a)$$

Thus, twice the parent-offspring phenotypic covariance provides an estimate of σ_A^2 , which is called the **additive genetic variance**. There are a number of potential complications in blindly applying

(2a), such as shared environmental effects, maternal effects when considering the mother-offspring covariance, etc. See Falconer and MacKay (1996) and Lynch and Walsh (1998) for details on handling these complications. Gianola (this Handbook) considers how to use not just parent-offspring information, but to simultaneously use all the information in general pedigrees in estimating the additive genetic variances.

The covariance in breeding values between two traits (say x and y), $\sigma_A(x, y)$, is needed for considering evolution of multiple traits. These **additive genetic covariances** can also be estimated from parent-offspring regressions, as the phenotypic covariance between one trait in the parent (p) and the other trait in the offspring (o) estimates half the additive genetic covariance of the traits, e.g.,

$$\sigma^2(x_p, y_o) = \sigma^2(y_p, x_o) = \frac{\sigma_A(x, y)}{2} \quad (2b)$$

Two different mechanisms can generate a (within-individual) covariance in the breeding values of two traits: linkage and **pleiotropy**. Linked loci show an excess of parental gametes, creating a correlation between alleles at these loci. (If a parent has alleles A and B on one chromosome and a and b on the other homologous chromosome, this parent will produce more AB and ab parental gametes than Ab and aB recombinant gametes.) In this case, even if each linked locus affects only a single trait, the correlation between alleles (linkage disequilibrium) generates a short-term correlation between the breeding values of the traits. Over time these associations decay through recombination randomizing allelic associations across loci. Alternatively, with pleiotropy a single locus can affect multiple traits. Covariances due to pleiotropic loci are stable over time.

1.2. Single trait Parent-offspring regressions

Phenotypic and Genetic Regressions

Most of the theory of selection response in quantitative genetics assumes linear parent-offspring regressions that have homoscedastic residuals (i.e., the variance about the expected value is inde-

pendent of the values of the parents). The simplest version considers the phenotypic value of the midparent, $z_{mp} = (z_m + z_f)/2$, with the offspring value z_o predicted as

$$z_o = \mu + b(z_{mp} - \mu) \quad (3a)$$

where μ is the population mean. Recalling that the slope b of the best linear regression of y on x , $y = a + bx$, is given by

$$b = \frac{\sigma(x, y)}{\sigma_x^2}$$

the slope of the midparent-offspring regression becomes

$$b = \frac{\sigma(z_o, z_{mp})}{\sigma^2(z_{mp})} = \frac{\sigma(z_o, z_m)/2 + \sigma(z_o, z_f)/2}{\sigma^2(x_m/2 + x_f/2)} = \frac{2\sigma_A^2/4}{2\sigma_z^2/4} = \frac{\sigma_A^2}{\sigma_z^2} \quad (3b)$$

This ratio of the additive genetic variance to the phenotypic variance is usually denoted h^2 and is called the (narrow-sense) **heritability**. Thus, the midparent-offspring regression is given by

$$z_o = \mu + h^2(z_{mp} - \mu) = \mu + h^2 \left(\frac{z_m + z_f}{2} - \mu \right) \quad (3c)$$

(3c) is the expected value for an offspring. The actual value for any particular offspring is given by

$$z_o = \mu + h^2(z_{mp} - \mu) + e \quad (4a)$$

where the residual e has mean value zero and variance

$$\sigma_e^2 = \left(1 - \frac{h^4}{2} \right) \sigma_z^2 \quad (4b)$$

Selection differentials and the Breeders' Equation

The parent-offspring regression allows us to predict the response to selection. Suppose the mean of parents that reproduce (μ_*) is different from the population mean before selection μ . Define the **directional selection differential** as $S = \mu_* - \mu$ (S is often simply called the selection differential). From (3c), the difference R between the offspring mean of these parents and the original mean of the population before selection is

$$R = h^2 S \quad (5)$$

This is the **breeders' equation** which transforms the between-generation change in the mean (the selection response R) into the within-generation change in the mean (S). Notice that strong selection does not necessarily imply a large response. If the heritability of a trait is very low (as occurs with many life-history traits), then even very strong selection results in very little (if any) response. Hence, selection (nonzero S) does not necessarily imply evolution (nonzero R).

1.3. Multiple trait parent-offspring regressions

The (single-trait) parent-offspring regression can be generalized to a (column) vector of n trait values, $\mathbf{z} = (z_1, \dots, z_n)^T$. Letting \mathbf{z}_o be the vector of trait values in the offspring, \mathbf{z}_{mp} the vector of midparent-trait values (i.e., the i th element is just $(z_{f,i} + z_{m,i})/2$), and $\boldsymbol{\mu}$ be the vector of population means, then the parent-offspring regression for multiple traits can be written as having expected value

$$\mathbf{z}_o = \boldsymbol{\mu} + \mathbf{H}(\mathbf{z}_{mp} - \boldsymbol{\mu}) \quad (6)$$

where the ij -th element in the matrix \mathbf{H} is the weight associated with the value of trait i in the offspring and trait j in the midparent.

The Genetic \mathbf{G} and Phenotypic \mathbf{P} Covariance Matrices

In order to further decompose \mathbf{H} into workable components, we need to define the phenotypic covariance matrix \mathbf{P} whose ij th element is the phenotypic covariance between traits i and j . Note that \mathbf{P} is symmetric, with the diagonal elements corresponding to the phenotypic variances and off-diagonal elements corresponding to the phenotypic covariances. In a similar manner, we can define the symmetric matrix \mathbf{G} , whose ij th element is the additive genetic covariance (covariance in breeding values) between traits i and j . Using similar logic as that leading to (3b) (the slope of the single-trait regression), it can be shown the

$$\mathbf{H} = \mathbf{G}\mathbf{P}^{-1} \quad (7a)$$

giving the multitrait parent-offspring regression as

$$\mathbf{z}_o = \boldsymbol{\mu} + \mathbf{GP}^{-1}(\mathbf{z}_{mp} - \boldsymbol{\mu}) \quad (7b)$$

The Multivariate Breeders' Equation

Letting \mathbf{R} denote the column vector of responses (so that the i th element in the between-generation change in the mean of trait i), and \mathbf{S} the vector of selection differentials, then (7b) allows us to generalize the breeders' equation to multiple traits,

$$\mathbf{R} = \mathbf{HS} = \mathbf{GP}^{-1}\mathbf{S} \quad (8)$$

(8) forms the basis for discussions about selection on multiple characters (Section 6).

2. SELECTION RESPONSE UNDER THE INFINITESIMAL MODEL

2.1. The Infinitesimal Model

The breeders' equation predicts the change in mean following a single generation of selection from an unselected base population. The only assumption required is that the parent-offspring regression is linear, but the question remains as to when this linearity holds. Further, the breeders' equation focuses only on the change in mean, leaving open the question of what happens to the variance. The latter concern is of special relevance in evolutionary quantitative genetics, as stabilizing selection (which reduces the variance without necessarily any change in the mean) is thought to be common in natural populations. The infinitesimal model, introduced by Fisher (1918), provides a framework to address these issues.

Allele frequency changes under the infinitesimal model

Under the infinitesimal model, a character is determined by an infinite number of unlinked and nonepistatic loci, each with an infinitesimal effect. A key feature of the infinitesimal model is that

while allele frequencies are essentially unchanged by selection, large changes in the mean can still occur by summing infinitesimal allele frequency changes over a large number of loci. To see this feature, consider a character determined by n completely additive and interchangeable loci, each with two alleles, Q and q (at frequencies p and $1 - p$), where the genotypes QQ , Qq , and qq contribute $2a$, a , and 0 (respectively) to the genotypic value. The resulting mean is $2nap$ and the additive variance (ignoring the contribution from gametic-phase disequilibrium) is $\sigma_A^2 = 2na^2p(1 - p)$. For σ_A^2 to remain bounded as the number of loci increase, a must be of order $n^{-1/2}$. The change in mean due to a single generation of selection is easily found to be $\Delta\mu = 2na\Delta p$. Assuming the frequency of Q changes by the same amount at each locus, $\Delta p = \Delta\mu/(2na)$. Since a is of order $n^{-1/2}$, Δp is of order $1/(n \cdot n^{-1/2}) = n^{-1/2}$, approaching zero as the number of loci becomes infinite. Thus the infinitesimal model allows for arbitrary changes in the mean with (essentially) no change in the allele frequencies at underlying loci.

What effect does this very small amount of allele frequency change have on the variance? Letting $p' = p + \Delta p$ denote the frequency after selection, the change in the additive genic variance is

$$\begin{aligned}\Delta\sigma_A^2 &= 2na^2p'(1 - p') - 2na^2p(1 - p) \\ &= 2na^2\Delta p(1 - 2p - \Delta p) \\ &\approx a(1 - 2p)\Delta\mu\end{aligned}$$

Since a is of order $n^{-1/2}$, the change in variance due to changes in allele frequencies is roughly $1/\sqrt{n}$ the change in mean. Thus, with a large number of loci, very large changes in the mean can occur without any significant change in the variance. In the limit of an infinite number of loci, there is no change in the genic variance ($\Delta\sigma_A^2 = 0$), while arbitrary changes in the mean can occur.

Linearity of parent-offspring regressions under the infinitesimal model

Under the infinitesimal model, genotypic values (G) are normally (or multivariate normally if G is a vector) distributed before selection (Bulmer 1971, 1976). Assuming environmental values (E) are

also normal, then so is the phenotype z (as $z = G + E$) and the joint distribution of phenotypic and genotypic values is multivariate normal. In this case, the regression of offspring phenotypic value z_o on parental phenotypes (z_f and z_m for the father and mother's values) is linear and homoscedastic.

2.2. Changes in Variances

As mentioned, under the infinitesimal model (in an infinite population) there is essentially no change in the genetic variances caused by allele frequency change. Changes in allele frequencies, however, are not the only route by which selection can change the variance (and other moments) of the genotypic distribution. Selection also creates associations (covariances) between alleles at different loci through the generation of gametic-phase (or linkage) disequilibrium, and such covariances can have a significant effect on the genetic variance. Disequilibrium can also change higher-order moments of the genotypic distribution as well, driving it away from normality and hence potentially causing parent-offspring regressions to deviate from linearity.

The additive variance under disequilibrium

The additive genetic variance σ_A^2 in the presence of linkage disequilibrium can be written as

$$\sigma_A^2 = \sigma_a^2 + d \quad (9)$$

where σ_a^2 is the additive genetic variance in the absence of disequilibrium and d the disequilibrium contribution. To formally define σ_a^2 and d , let $a_1^{(k)}$ and $a_2^{(k)}$ denote average effects of the two alleles at locus k from a random individual. Since σ_A^2 is the variance of the sum of average effects over all loci,

$$\sigma^2 \left(\sum_{k=1}^n (a_1^{(k)} + a_2^{(k)}) \right) = 2 \sum_{k=1}^n \sigma^2 (a^{(k)}) + 4 \sum_{k < j}^n \sigma (a^{(j)}, a^{(k)}) \quad (10a)$$

$$= 2 \sum_{k=1}^n C_{kk} + 4 \sum_{k < j}^n C_{jk} \quad (10b)$$

where n is the number of loci and $C_{jk} = \sigma(a^{(j)}, a^{(k)})$ is the covariance between allelic effects at locus j and k . Thus $\sigma_a^2 = 2 \sum C_{kk}$ is the additive variance in the absence of gametic-phase disequilibrium

and the disequilibrium contribution $d = 4 \sum_{j < k} C_{kj}$ is the covariance between allelic effects at different loci. The component of the additive genetic variance that is unaltered by changes in gametic-phase disequilibrium, σ_a^2 , is often referred to as the **additive genic variance** (or simply the **genic variance**) to distinguish it from the additive *genetic* variance σ_A^2 .

Under the infinitesimal model, selection does not change the C_{kk} (as this requires changes in the allele frequencies), and hence does not alter σ_a^2 . However, selection does generate correlations between loci ($C_{ik} \neq 0$), and this can result in significant changes in the overall additive variance σ_A^2 . Changes in the covariances C_{ij} between loci i and j (for $i \neq j$) are roughly of order n^{-2} (Bulmer 1980, Turelli and Barton 1990). Since there are n^2 terms contributing to d , the total disequilibrium is of order one ($n^2 \cdot n^{-2}$) and does not necessarily approach zero as the number of loci becomes infinite. The same reasoning holds for changes in the higher-order moments, which are caused by higher-order associations between groups of loci, and can also be non-trivial (Turelli and Barton 1990).

Dynamics of disequilibrium

Recall that the total genetic variance is the sum of the additive and dominance variances (plus any epistatic variances), so that $\sigma_G^2 = \sigma_A^2 + \sigma_D^2$. Under the infinitesimal model, disequilibrium changes the additive genetic, but not the dominance, variance (Walsh and Lynch 2003). Hence, the phenotypic variance in generation t of selection is

$$\sigma_z^2(t) = \sigma_E^2 + \sigma_D^2 + \sigma_A^2(t) = \sigma_z^2 + d(t) \quad (11a)$$

where $\sigma_z^2 = \sigma_z^2(0)$ is the phenotypic variance before selection in the initial (unselected) base population. The resulting heritability in generation t becomes

$$h^2(t) = \frac{\sigma_A^2(t)}{\sigma_z^2(t)} = \frac{\sigma_a^2 + d(t)}{\sigma_z^2 + d(t)} \quad (11b)$$

Assuming that the parent-offspring regression remains linear, the selection response in generation t becomes $R(t) = h^2(t) S(t)$. Assuming unlinked loci, Bulmer (1971) showed that the change in

disequilibrium is given by

$$d(t+1) = \frac{d(t)}{2} + \frac{h^4(t)}{2} \left(\sigma_{z^*}^2(t) - \sigma_{z(t)}^2 \right) \quad (12)$$

where $\sigma_{z^*}^2(t) - \sigma_{z(t)}^2$ is the within-generation change in the phenotypic variance. The first term represents the removal of disequilibrium by recombination, while the second is the generation of disequilibrium by selection. Note that (12) is the variance analog of the breeders' equation, relating the between $[d(t+1) - d(t)]$ and within-generation $[\sigma_{z^*}^2(t) - \sigma_{z(t)}^2]$ changes in the variance.

Starting with an unselected base population (where $d(0) = 0$), iterating (12) gives the disequilibrium (and hence the heritability, phenotypic variance, and response) in any desired generation. Under directional selection (selection only on the mean), most of the disequilibrium is generated in the first three to five generations, after which d is very close to its equilibrium value \tilde{d} . Once selection stops, the current disequilibrium is halved each generation, with the additive variation rapidly returning to its value before selection. This occurs because under the infinitesimal model, all changes in variances are due to disequilibrium, which decays via recombination in the absence of selection.

An important point regarding (12) is that there can be no change in the mean ($S = 0$) and yet selection can still act on the variance. Stabilizing selection, which removes extreme individuals and reduces the variance, is widely thought to be widespread in nature. The within-generation reduction in variance generates negative disequilibrium, which in turn reduces the additive variance. Under the infinitesimal model, once stabilizing selection stops, the variance returns (increases) to its initial value after a few generations. Likewise, under disruptive selection, individuals of intermediate value are selected against, increasing the variance. This generates positive d , increasing the additive variance. Again, when selection is stopped, the variance decays back to its initial (pre-selection) value.

Iteration of (12) allows any form of selection to be analyzed. For modeling purposes, it is often

assumed that the within-generation change in the phenotypic variance can be written as

$$\sigma_{z^*}^2(t) = (1 - \kappa) \sigma_{z(t)}^2 \quad (13)$$

where κ is a constant that does not change over time. Noting that $\sigma_{z(t)}^2 = \sigma_A^2(t)/h^2(t)$ and substituting (13) into (12) recovers the result of Bulmer (1974),

$$d(t+1) = \frac{d(t)}{2} - \frac{\kappa}{2} h^2(t) \sigma_A^2(t) = \frac{d(t)}{2} - \frac{\kappa}{2} \frac{[\sigma_a^2 + d(t)]^2}{\sigma_z^2 + d(t)} \quad (14)$$

Again, simple iteration allows one to compute the variance in any generation.

Equilibrium variances

The equilibrium variances, and hence the asymptotic rate of response under directional selection or the equilibrium variance under stabilizing or disruptive selection, are easily obtained. At equilibrium, (14) implies

$$\tilde{d} = -\kappa \tilde{h}^2 \tilde{\sigma}_A^2 = -\kappa \frac{(\sigma_a^2 + \tilde{d})^2}{\sigma_z^2 + \tilde{d}} \quad (15a)$$

Solving for \tilde{d} gives the equilibrium additive genetic variance as

$$\tilde{\sigma}_A^2 = \sigma_z^2 \theta, \quad \text{where} \quad \theta = \frac{2h^2 - 1 + \sqrt{1 + 4h^2(1 - h^2)\kappa}}{2(1 + \kappa)} \quad (15b)$$

and the resulting heritability at equilibrium is

$$\tilde{h}^2 = \frac{\tilde{\sigma}_A^2}{\tilde{\sigma}_z^2} = \frac{\tilde{\sigma}_A^2}{\sigma_z^2 + (\tilde{\sigma}_A^2 - \sigma_A^2)} = \frac{\theta}{1 + \theta - h^2} \quad (15c)$$

The simple picture that emerges from directional selection under the infinitesimal model is that while the heritability may decrease slightly from its initial value (due to generation of linkage disequilibrium), the response to selection continues without limit. Biological reality, of course, places limits on the character values that can be obtained by selection response. For example, selection may move the mean phenotype to a region on the fitness surface where stabilizing selection dominates (see Section 4 below). Likewise, selection on one character may be opposed by selective constraints

imposed by other characters also under selection, and a limit can be reached despite significant additive variance and a nonzero S in the character being followed (Section 6).

2.3. The Roles of Drift and Mutation Under the Infinitesimal Model

While selection does not change allele frequencies under the infinitesimal model, genetic drift (due to finite population size) and mutation can play very important roles in shaping the genetic variance. Inclusion of these two forces allows the infinitesimal model to be modified into a much more biologically realistic description of selection response.

Drift

Assuming no dominance or epistasis, with drift (assuming an effective population size of N_e) the expected genic variation σ_a^2 declines each generation from its initial value $\sigma_a^2(0)$, with the genic variance in generation t being

$$\sigma_a^2(t) = \sigma_a^2(0) \left(1 - \frac{1}{2N_e}\right)^t \quad (16)$$

If dominance (or epistasis) is present, the additive variation can actually increase (at least while the level of inbreeding is moderate) before it ultimately declines to zero (Walsh and Lynch 2003). Under the infinitesimal model, selection is so weak on any given underlying locus that the allele frequency dynamics are entirely governed by drift, and the change in the genic variance is given by (16).

Keightley and Hill (1987) show that, under drift, the change in the amount of disequilibrium is given by

$$\Delta d(t) = -\frac{d(t)}{2} \left(1 + \frac{1}{N_e}\right) - \frac{1}{2} \left(1 - \frac{1}{N_e}\right) \kappa h^2(t) \sigma_A^2(t) \quad (17)$$

Joint iteration of (16) and (17) allows for modeling of response with a finite population size. During the course of selection, genetic drift removes the initial variation, driving the genic (and hence additive genetic) variance to zero. With directional selection, a limit is reached, as the response stops as the heritability approaches zero. Robertson (1960) showed that the total change in the

mean in this case is approximately $2N_e R(1)$, with $R(1)$ the response in the first generation. It takes approximately $1.4N_e$ generations for half this total response to occur.

Mutation

Balancing the removal of genetic variation by genetic drift is the introduction of new variation by mutation. The mutation variance σ_m^2 is defined as the per-generation contribution by mutation to the additive variance, with the average value of σ_m^2/σ_E^2 typically around 0.005 (Lynch and Walsh 1998). When both mutation and drift occur, the additive variance eventually attains an equilibrium value of

$$\tilde{\sigma}_A^2 = 2N_e \sigma_m^2 \quad (18)$$

Assuming the infinitesimal model, completely additive loci, a base-population additive variance of $\sigma_A^2(0)$ and ignoring any effects of gametic-phase disequilibrium, the expected additive genetic variance at generation t is given by

$$\sigma_A^2(t) \simeq 2N_e \sigma_m^2 + [\sigma_A^2(0) - 2N_e \sigma_m^2] \exp(-t/2N_e) \quad (19a)$$

Setting $\sigma_A^2(0) = 0$ gives the additive variance contributed entirely from mutation as

$$\sigma_{A,m}^2(t) \simeq 2N_e \sigma_m^2 [1 - \exp(-t/2N_e)] \quad (19b)$$

Hence, the rate of response at generation t from mutational input is

$$R_m(t) = i \frac{\sigma_{A,m}^2(t)}{\sigma_z} \simeq 2N_e i \frac{\sigma_m^2}{\sigma_z} [1 - \exp(-t/2N_e)] \quad (19c)$$

with $i = S/\sigma_z$ (the selection intensity, which is discussed below). For $t \gg 2N_e$, the per-generation response approaches an asymptotic limit of

$$\tilde{R}_m = 2N_e i \frac{\sigma_m^2}{\sigma_z} = i \frac{\tilde{\sigma}_A^2}{\sigma_z} \quad (19d)$$

A full treatment allowing for mutation, drift, and disequilibrium under the infinitesimal model is given by simply adding a σ_m^2 term to (16) and then jointly iterating (16) and (17).

3. FITNESS

Predicting the selection response under the infinitesimal model requires knowledge of both the change in mean and variance after an episode of selection. In an artificial selection program, the breeder or experimentalist can not only measure these components, but can also largely set their values. In nature, the currency of selection is fitness, and the change in the phenotypic distribution is computed by first weighting individuals by their fitness values. Discussion of selection response in nature thus starts by trying to assign fitness values to particular character states. This is the linking step that allows use of the above machinery for prediction of selection response.

3.1. Individual Fitness

Loosely stated, the **lifetime** (or **total**) **fitness** of an individual is the number of descendants it leaves at the start of the next generation. When measuring the total fitness of an individual, care must be taken not to cross generations or to overlook any stage of the life cycle in which selection acts. To accommodate these concerns, lifetime fitness is defined as the total number of zygotes (newly fertilized gametes) that an individual produces. Measuring total fitness from any other starting point in the life cycle (e.g., from adults in one generation to adults in the subsequent generation) can result in a very distorted picture of true fitness of particular phenotypes (Prout 1965, 1969). If generations are crossed, measures of selection on a particular parental phenotype in reality are averages over both parental and offspring phenotypes, which may differ considerably.

Systems for measuring lifetime fitness have been especially well developed for laboratory populations of *Drosophila* (reviewed by Sved 1989). Measurements of lifetime fitness in field situations are more difficult and (not surprisingly) are rarely made. Attention instead is usually focused on

particular episodes of selection or particular phases of the life cycle. Fitness components for each episode of selection are defined to be multiplicative. For example, lifetime fitness can be partitioned as (probability of surviving to reproductive age)·(number of mates)·(number of zygotes per mating). Number of mates is a measure of **sexual selection**, while the viability and fertility components measure natural selection. A commonly measured fitness component is **reproductive success**, the number of offspring per adult, which confounds natural (fertility) and sexual selection (in males, the number of matings per adult). Clutton-Brock (1988) reviews estimates of reproductive success from natural populations.

Fitness components can themselves be further decomposed. For example, fertility in plants might be decomposed as (seeds per plant) = (number of stems per plant)·(number of inflorescences per stem)·(average number of seed capsules per inflorescence)·(average number of seeds per capsule). Such a decomposition allows the investigator to ask questions of the form: do plants differ in number of seeds mainly because some plants have more stems, or more flowers per stem, or are there tradeoffs between these?

Estimates of fitness can be obtained from either **longitudinal** or **cross-sectional** studies. A longitudinal study follows a cohort of individuals over time, while a cross-sectional study examines individuals at a single point in time. Cross-sectional studies typically generate only two fitness classes (e.g., dead versus living, mating versus unmated), and their analysis involves a considerable number of assumptions (Lande and Arnold 1983, Arnold and Wade 1984b). While longitudinal studies are preferred, they usually require far more work and may be impossible to carry out in many field situations. Age-structured populations pose further complications in that proper fitness measures require knowledge of the population's demography, see Charlesworth (1994), Lande (1982), Lenski and Service (1982), and Travis and Henrich (1986) for details.

3.2. Episodes of Selection

As mentioned, individuals are often measured over more than one episode of selection. Imagine

that a cohort of n individuals (indexed by $1 \leq r \leq n$) is followed through several episodes. Let $W_j(r)$ be the fitness measure for the j th episode of selection for the r th individual. For example, for viability W_j is either zero (dead) or one (alive) at the census period. Let \overline{W}_j denote the mean fitness of a random individual following the j th episode of selection. Relative fitness components $w_j(r) = W_j(r)/\overline{W}_j$ will turn out to be especially useful. At the start of the study, the frequency of each individual is $1/n$, giving for the first (observed) episode of selection

$$\overline{W}_1 = \frac{1}{n} \sum_{r=1}^n W_1(r) \quad (20a)$$

Caution is in order at this point as *considerable selection may have already occurred prior to the life cycle stages being examined*. Following the first episode of selection, the new fitness-weighted frequency of the r th individual is $w_1(r)/n$, implying

$$\overline{W}_2 = \sum_{r=1}^n W_2(r) \cdot w_1(r) \cdot \left(\frac{1}{n}\right) \quad (20b)$$

In general, for the j th episode of selection,

$$\overline{W}_j = \sum_{r=1}^n W_j(r) \cdot w_{j-1}(r) \cdot w_{j-2}(r) \cdots w_1(r) \cdot \left(\frac{1}{n}\right) \quad (20c)$$

Note that if $W_j(r) = 0$, further fitness components for r are unmeasured.

Letting $p_j(r)$ be the fitness-weighted frequency of individual r after j episodes of selection, it follows that $p_0(r) = 1/n$ and

$$p_j(r) = w_j(r) \cdot p_{j-1}(r) = \frac{1}{n} \prod_{i=1}^j w_i(r) \quad (21a)$$

Thus, (20c) can also be expressed as $\overline{W}_j = \sum W_j(r) \cdot p_{j-1}(r)$. Using these weights allows fitness-weighted moments to be calculated, e.g., the mean of a particular character following the j th episode is computed as

$$\mu_{z(j)} = \sum z(r) \cdot p_j(r) \quad (21b)$$

where $z(r)$ is the value of the character of individual r .

The directional selection differential S is computed as the difference between fitness-weighted means before and after an episode of selection, with the differential S_j for the j th episode given by

$$S_j = \mu_{z(j)} - \mu_{z(j-1)} \quad (22a)$$

Selection differentials are additive over episodes, so that the total differential S following k episodes of selection is

$$S = \mu_{z(k)} - \mu_{z(0)} = (\mu_{z(k)} - \mu_{z(k-1)}) + \cdots + (\mu_{z(1)} - \mu_{z(0)}) = S_k + \cdots + S_1 \quad (22b)$$

3.3. The Robertson-Price Identity, $S = \sigma(w, z)$

As first noted by Robertson (1966), and greatly elaborated on by Price (1970, 1972a), the directional selection differential equals the covariance of phenotype and relative fitness,

$$S = \sigma(w, z) \quad (23)$$

This identity is quite useful for obtaining the selection differential in complex selection schemes and (as is detailed below) forms the basis for a number of useful expressions relating selection and fitness.

To obtain the Robertson-Price identity, let μ_s be the fitness-weighted mean after selection and μ the mean before selection,

$$\begin{aligned} S = \mu_s - \mu &= \sum_{r=1}^n z(r) w(r) \cdot p(r) - \mu \\ &= E[zw] - 1 \cdot E[z] \\ &= E[zw] - E[w] \cdot E[z] \\ &= \sigma(w, z) \end{aligned}$$

where we have used the fact that (by construction) $E[w] = 1$.

3.4. I, The Opportunity for Selection

How does one compare the amount of selection acting on different episodes and/or different populations? At first thought, one might consider using the standardized selection differential,

$$i = \frac{S}{\sigma_z} \quad (24)$$

which is just the directional selection differential scaled in terms of character standard deviations (i is often called the **selection intensity** and is widely used in breeding). The drawback with i as a measure of *overall* selection on individuals is that it is *character specific*. Hence, i is appropriate if we are interested in comparing the strength of selection on a particular *character*, but inappropriate if we wish to compare the overall strength of selection on *individuals*. Two populations may have the same i value for a given character, but if that character is tightly correlated with fitness in one population and only weakly correlated in the other, selection is much stronger in the latter population. Further, considerable selection can occur without changing the mean (e.g., stabilizing selection). Standardized (quadratic) differentials also exist for the variance (section 4.3), but the problem of character-specificity still remains.

A much cleaner measure (independent of the characters under selection) is I , the **opportunity for selection**, defined as the variance in *relative* fitness:

$$I = \sigma_w^2 = \frac{\sigma_W^2}{\bar{W}^2} \quad (25)$$

This measure was introduced by Crow (1958, reviewed in 1989), who referred to it as the **Index of Total Selection**. Crow noted that if fitness is perfectly heritable (e.g., $h^2(\text{fitness}) = 1$), then $I = \Delta\bar{w}$, the scaled change in fitness. Following Arnold and Wade (1984a,b), I is referred to as the opportunity for selection, as any change in the distribution of fitness caused by selection represents an opportunity for within-generation change. A key feature of I is that it bounds the maximal selection intensity i for *any* character. From the Robertson-Price identity (23), the correlation between any character z and relative fitness (which is bounded in absolute value by one) is

$$|\rho_{z,w}| = \frac{|\sigma_{z,w}|}{\sigma_z \sigma_w} = \frac{|S|}{\sigma_z \sqrt{I}} = \frac{|i|}{\sqrt{I}} \leq 1$$

implying

$$|i| \leq \sqrt{I} \tag{26}$$

Thus, the most that the mean of any character can be shifted within a generation is \sqrt{I} phenotypic standard deviations.

=====> **Figure 1 here** < =====

The usefulness of I as a bound of i depends on the correlation between relative fitness and the character being considered. Figure 1 shows scatterplots of relative fitness versus two characters (z_1 and z_2) measured in the same set of individuals. The marginal distributions of fitness are identical for both characters (since the same set of individuals was measured for both traits), and hence both traits have the same opportunity for selection. The association between relative fitness and trait z_1 is fairly strong, while there is no relationship between relative fitness and z_2 , so that z_1 realizes much, while z_2 realizes none, of the opportunity for change.

3.5. Some Caveats in Using the Opportunity for Selection

Despite its usefulness, there are some subtle issues in the interpretation of I . To begin with, even though I appears to remove scaling effects due to different types of fitnesses, for estimates of I to be truly comparable, they must be based on consistent measures of fitness (Tria 1985). Second, if the variance in fitness is not independent of \bar{W} , comparisons of I values between populations are compromised. Such a lack of independence occurs in cross-sectional studies that measure sexual selection by simply counting the number of mating pairs. If the time scale is such that only single matings are observed, the fitness of any individual is either 1 (mating) or 0 (not mating). The resulting fitness of randomly-drawn individuals is binomially distributed with mean p (the mean copulatory success for the sex being considered) and variance $p(1 - p)$, hence

$$I = \frac{p(1 - p)}{p^2} \simeq \frac{1}{p} \quad \text{if } p \ll 1 \tag{27}$$

The mean and variance in individual fitness are not independent, and the opportunity for selection

depends entirely on mean population fitness. As the time window for observing mating pairs decreases, fewer matings are seen and p decreases, increasing I . Thus the opportunity for selection is often inflated if the observation period is short relative to the total mating period. Likewise, if one is only comparing viability selection (alive or dead after an episode of selection), (27) also holds.

Another example of the lack of independence between \overline{W} and σ_W^2 was given by Downhower et al. (1987), who assumed that the number of mates for any given male follows a Poisson distribution. In this case, the variance in number of mates equals the mean number of mates, giving

$$I = \frac{\overline{W}}{\overline{W}^2} = \overline{W}^{-1} \quad (28)$$

where \overline{W} is the mean number of mates per male. Thus, differences in I between populations do not necessarily indicate *biological* differences in male mating ability. For example, in a population of 100 males, if only 5 females mate, average male mating success is $\overline{W} = 0.05$, while if 50 females mate, $\overline{W} = 0.5$. For this example, differences in I come solely from variation in the number of mating females, not biological differences between males in their ability to acquire mates. Downhower et al. conclude from this example that comparing I values with the Poisson prediction ($I = 1/\overline{W}$), or some other appropriate random distribution, may help clarify the interpretation of I . For this case, values of I less than the Poisson prediction indicate a more uniform distribution of fitness than expected if mate choice is random, while values in excess of this expectation indicate disproportionately high fitness among a limited set of individuals.

The Poisson mating example further points out that random variation (differences in individual fitness not attributable to intrinsic differences between individuals) reduces the correlation between phenotypic value and relative fitness. For this reason, measures of selection based entirely upon variance in mating success have been criticized (Banks and Thompson 1985, Sutherland 1985a,b, Koenig and Albano 1986). Although carefully controlled studies can reduce the error variance induced by chance (e.g., Houck et al. 1985), accounting for inflation of the opportunity for selection by random effects remains a problem.

4. FITNESS SURFACES

4.1. Individual and Mean Fitness Surfaces

$W(z)$, the expected fitness of an individual with phenotype z , describes a **fitness surface** (or **fitness function**), relating fitness and character value. The *relative* fitness surface $w(z) = W(z)/\bar{W}$ is often more convenient than $W(z)$, and we use these two somewhat interchangeably. The nature of selection on a character in a particular population is determined by the local geometry of the individual fitness surface over that part of the surface spanned by the population (Figure 2). If fitness is strictly increasing (or decreasing) over some range of phenotypes, a population having its mean value in this interval experiences **directional selection**. If $W(z)$ contains a local maximum, a population with members within that interval experiences **stabilizing selection**. If the population is distributed around a local minimum, **disruptive selection** occurs. As is illustrated in Figure 2D, when the local geometry of the fitness surface is complicated (e.g., multimodal) the simplicity of description offered by these three types of selection breaks down, as the population can experience all three simultaneously.

=====> Figure 2 here <=====

$W(z)$ may vary with genotypic and environmental backgrounds. In some situations (e.g., predators with search images, sexual selection, dominance hierarchies, truncation selection) the fitness of a phenotype depends on the frequency of other phenotypes in the population. In this case, fitnesses are said to be **frequency-dependent**.

Mean population fitness \bar{W} is also a fitness surface, describing the expected fitness of the population as a function of the distribution $p(z)$ of phenotypes in that population,

$$\bar{W} = \int W(z) p(z) dz \quad (29)$$

Mean fitness can be thought of as a function of the parameters of the phenotypic distribution. For example, if z is normally distributed, mean fitness is a function of the mean μ_z and variance σ_z^2 for

that population.

To stress the distinction between the $W(z)$ and \bar{W} fitness surfaces, the former is referred to as the **individual fitness surface**, the latter as the **mean fitness surface**. Knowing the individual fitness surface allows one to compute the mean fitness surface for any specified $p(z)$ but the converse is not true. The importance of the mean fitness surface is that it provides one way of describing how the population changes under selection — under the infinitesimal model, the derivative of \bar{W} with respect to μ_z describes the change in mean (31b).

4.2. Measures of Selection on the Mean

A general way of detecting selection on a character is to compare the (fitness-weighted) phenotypic distribution before and after some episode of selection. Growth or other ontogenetic changes, immigration, and environmental changes can also change the phenotypic distribution, and great care must be taken to account for these factors. Another critical problem in detecting selection on a character is that selection on phenotypically correlated characters can also change the distribution. Keeping this important caveat in mind, we first examine measures of selection on a single character, as these form the basis for our discussion in Section 5 on measuring selection on multiple characters. Typically, selection on a character is measured by changes in the mean and variance, rather than changes in the entire distribution. While often only the mean is examined, considerable selection can occur without any significant change in the mean (for example, stabilizing selection).

Two measures of within-generation change in phenotypic mean have been previously introduced: the directional selection differential S and the selection intensity i . A third measure is the directional selection gradient

$$\beta = \frac{S}{\sigma_z^2} \quad (30)$$

As detailed in Section 4.6, β is the slope of the linear regression of fitness on phenotype. These three measures (i, S, β) are interchangeable for selection acting on a single character, but their multivariate extensions have very different behaviors (Section 5). In particular, the multivariate extension of β

is the measure of choice, as it measures the amount of direct selection on a character, while S (and hence i) confounds direct selection with indirect effects due to selection on phenotypically correlated traits (48).

Recalling that $h^2 = \sigma_A^2/\sigma_z^2$, the response is often written by breeders as

$$R = h^2 \sigma_z i = \sigma_A h i \quad (31a)$$

The response can also be expressed in terms of the directional selection gradient,

$$R = \sigma_A^2 \beta \quad (31b)$$

4.3. Measures of Selection on the Variance

Similar measures can be defined to quantify the change in the phenotypic variance. At first glance this change seems best described by $\sigma_{z^*}^2 - \sigma_z^2$ where $\sigma_{z^*}^2$ is the phenotypic variance following selection. The problem with this measure is that directional selection reduces the variance. Lande and Arnold (1983) showed that

$$\sigma_{z^*}^2 - \sigma_z^2 = \sigma [w, (z - \mu_z)^2] - S^2 \quad (32)$$

implying that directional selection decreases the phenotypic variance by S^2 . With this in mind, Lande and Arnold suggest a corrected measure, the **stabilizing selection differential**

$$C = \sigma_{z^*}^2 - \sigma_z^2 + S^2 \quad (33)$$

that describes selection acting directly on the variance. As we will see below, the term stabilizing selection differential may be slightly misleading, so following Phillips and Arnold (1989) we refer to C as the **quadratic selection differential**. Correction for the effects of directional selection is important, as claims of stabilizing selection based on a reduction in variance following selection can be due entirely to reduction in variance caused by directional selection. Similarly, disruptive selection can be masked by directional selection. Provided that selection does not act on characters phenotypically correlated with the one under study, C provides information on the nature of

selection on the variance. Positive C indicates selection to increase the variance (as would occur with disruptive selection), while negative C indicates selection to reduce the variance (as would occur with stabilizing selection). As we discuss shortly, $C < 0$ ($C > 0$) is *consistent* with stabilizing (disruptive) selection, but not *sufficient*. A further complication in interpreting C is that if the phenotypic distribution is skewed, selection on the variance changes the mean. This causes a non-zero value of S that in turn inflates C (Figure 3).

=====> **Figure 3 here** < =====

Analogous to S equaling the covariance between z and relative fitness, (32) implies that C is the covariance between relative fitness and the squared deviation of a character from its mean,

$$C = \sigma [w, (z - \mu)^2] \quad (34)$$

As was the case with S , the opportunity for selection I bounds the maximum possible within-generation change in variance (Arnold 1986). Expressing C as a covariance and using the standard definition of a correlation gives $C = \rho_{w, (z-\mu)^2} \sigma_w \sigma[(z - \mu)^2]$. Since $\rho_{w, (z-\mu)^2}^2 \leq 1$, we have

$$C^2 \leq \sigma_w^2 \sigma^2[(z - \mu)^2] = I \cdot (\mu_{4,z} - \sigma_z^4)$$

Thus,

$$|C| \leq \sqrt{I (\mu_{4,z} - \sigma_z^4)} \quad (35a)$$

If z is normally distributed, the fourth central moment $\mu_{4,z} = 3\sigma_z^4$, giving

$$|C| \leq \sigma_z^2 \sqrt{2I} \quad (35b)$$

The quadratic analog of β , the **quadratic (or stabilizing) selection gradient** γ , was suggested by Lande and Arnold (1983),

$$\gamma = \frac{\sigma [w, (z - \mu)^2]}{\sigma_z^4} = \frac{C}{\sigma_z^4} \quad (36)$$

As with β , γ is the measure of choice when dealing with multiple characters because it accounts for the effects of (measured) phenotypically correlated characters (Section 5.5).

4.4. Gradients and the Geometry of Fitness Surfaces

A conceptual advantage of β and γ is that they describe the average local geometry of the fitness surface when phenotypes are normally distributed. When z is normal and individual fitnesses are not frequency-dependent, β can be expressed in terms of the geometry of the *mean* fitness surface,

$$\beta = \frac{\partial \ln \bar{W}}{\partial \mu_z} = \frac{1}{\bar{W}} \frac{\partial \bar{W}}{\partial \mu_z} \quad (37a)$$

so that β is proportional to the slope of the \bar{W} surface with respect to population mean. β can also be expressed as a function of the *individual* fitness surface. Lande and Arnold (1983) showed, provided z is normally distributed, that

$$\beta = \int \frac{\partial w(z)}{\partial z} p(z) dz \quad (37b)$$

implying that β is the average slope of the individual fitness surface, the average being taken over the population being studied. Likewise, if z is normal,

$$\gamma = \int \frac{\partial^2 w(z)}{\partial z^2} p(z) dz \quad (37c)$$

which is the average curvature of the individual fitness surface (Lande and Arnold 1983). Thus, β and γ provide a measure of the geometry of the individual fitness surface averaged over the population being considered.

A final advantage of β and γ is that they appear as the measures of phenotypic selection in equations describing selection response. Recall (31b) that under the assumptions leading to the breeders' equation, $R = \Delta\mu = \sigma_A^2\beta$. Similarly, for predicting changes in additive genetic variance under the infinitesimal model, the expected change in variance from a single generation of selection is

$$\Delta\sigma_A^2 = \frac{\sigma_A^4}{2} (\gamma - \beta^2) \quad (38)$$

which decomposes the change in variance into changes due to selection on the variance (γ) and changes due to directional selection (β^2).

While the distinction between differentials and gradients seems trivial in the univariate case (only a scale difference), when considering multivariate traits, gradients have the extremely important feature of removing the effects of phenotypic correlations.

4.5. Estimating the Individual Fitness Surface

We can decompose the fitness W of an individual with character value z into the sum of its expected fitness $W(z)$ plus a residual deviation e ,

$$W = W(z) + e$$

The residual variance in fitness for a given z , $\sigma_e^2(z)$, measures the variance in fitness among individuals with phenotypic value z . Estimation of the individual fitness surface is thus a generalized regression problem, the goal being to choose a candidate function for $W(z)$ that minimizes the average residual variance $E_z[\sigma_e^2(z)]$. Since the total variance in fitness σ_W^2 equals the sum of the within-group (phenotype) and between-group variance in fitness,

$$\frac{\sigma_W^2 - E_z[\sigma_e^2(z)]}{\sigma_W^2}$$

is the fraction of individual fitness variation accounted for by a particular estimate of $W(z)$, and provides a measure for comparing different estimates. In the limiting case where fitness is independent of z (and any characters phenotypically correlated with z), $W(z) = \bar{W}$, so that the between-phenotypic variance is zero while $\sigma_e^2(z) = \sigma_W^2$.

There are at least two sources of error contributing to e . First, there can be errors in measuring the actual fitness of an individual (these are almost always ignored). Second, the *actual* fitness of a particular individual can deviate considerably from the *expected value* for its phenotype due to chance effects and selection on other characters besides those being considered. Generally, these residual deviations are heteroscedastic, with σ_e^2 often depending on the value of $W(z)$ (Mitchell-Olds and Shaw 1987, Schluter 1988). For example, suppose fitness is measured by survival to a

particular age. While $W(z) = p_z$ is the probability of survival for an individual with character value z , the fitness for a particular individual is either 0 (does not survive) or 1 (survives), giving $\sigma_e^2(z) = p_z(1 - p_z)$. Unless p_z is constant over z , the residuals are heteroscedastic. Note in this case that even after removing the effects attributable to differences in phenotypes, there still is substantial variance in individual fitness.

Inferences about the individual fitness surface are limited by the range of phenotypes in the population. Unless this range is very large, only a small region of the fitness surface can be estimated with any precision. Estimates of the fitness surface at the tails of the current phenotypic distribution are extremely imprecise, yet potentially very informative, suggesting what selection pressures populations at the margin of the observed range of phenotypes may experience. A further complication is that the fitness surface changes with the environment, so that year to year estimates can differ (e.g., Kalisz 1986) and cannot be lumped together to increase sample size.

4.6. Linear and Quadratic Approximations of $W(z)$

The individual fitness surface $W(z)$ can be very complex and a wide variety of functions may be chosen to approximate it. The simplest and most straightforward approach is to use a low-order polynomial (typically linear or quadratic).

Consider first the simple linear regression of *relative* fitness w as a function of phenotypic value z . Since the directional selection gradient

$$\beta = \frac{S}{\sigma_z^2} = \frac{\sigma(w, z)}{\sigma_z^2} \quad (39a)$$

it follows from regression theory that β is the slope of the least-squares linear regression of relative fitness on z ,

$$w = a + \beta z + e \quad (39b)$$

Hence the best linear predictor of relative fitness is $w(z) = a + \beta z$. Since the regression passes

through the expected values of w and z (1 and μ_z , respectively), this can be written as

$$w = 1 + \beta(z - \mu_z) + e \quad (39c)$$

giving $w(z) = 1 + \beta(z - \mu_z)$. Assuming the fitness function is well described by a linear regression, β is the expected change in relative fitness given a unit change in z . The fraction of variance in individual fitness accounted for by this regression is

$$r_{z,w}^2 = \frac{\text{Cov}^2(z, w)}{\text{Var}(z) \cdot \text{Var}(w)} = \hat{\beta}^2 \frac{\text{Var}(z)}{\hat{I}} \quad (40)$$

If the fitness surface shows curvature, as might be expected if there is stabilizing selection or disruptive selection, a **quadratic regression** is more appropriate,

$$w = a + b_1 z + b_2 (z - \mu_z)^2 + e \quad (41a)$$

Since the regression passes through the mean of all variables,

$$w = 1 + b_1(z - \mu_z) + b_2 [(z - \mu_z)^2 - \sigma_z^2] + e \quad (41b)$$

The regression coefficients b_1 and b_2 nicely summarize the local geometry of the fitness surface.

Evaluating the derivative of (41) at $z = \mu_z$ gives

$$\left. \frac{\partial w(z)}{\partial z} \right|_{z=\mu_z} = b_1 \quad \text{and} \quad \left. \frac{\partial^2 w(z)}{\partial z^2} \right|_{z=\mu_z} = 2b_2 \quad (42)$$

Hence b_1 is the slope and $2b_2$ the second derivative (curvature) of the best quadratic fitness surface around the population mean. $b_2 > 0$ indicates that the best-fitting quadratic of the individual fitness surface has an upward curvature, while $b_2 < 0$ implies the curvature is downward. Lande and Arnold (1983) suggest that $b_2 > 0$ indicates disruptive selection, while $b_2 < 0$ indicates stabilizing selection. Their reasoning follows from elementary geometry in that a *necessary* condition for a local minimum is that a function curves upward in some interval, while a necessary condition for a local maximum is that the function curves downward. Mitchell-Olds and Shaw (1987) and Schluter

(1988) argue that this condition is not *sufficient*. Stabilizing selection is generally defined as the presence of a local maximum in $w(z)$ and disruptive selection by the presence of a local minimum, while b_2 indicates curvature, rather than the presence of local extrema.

Recalling the covariances given by (23) and (34), the least-squares values for b_1 and b_2 become

$$b_1 = \frac{(\mu_{4,z} - \sigma_z^4) \cdot S - \mu_{3,z} \cdot C}{\sigma_z^2 \cdot (\mu_{4,z} - \sigma_z^4) - \mu_{3,z}^2} \quad (43a)$$

$$b_2 = \frac{\sigma_z^2 \cdot C - \mu_{3,z} \cdot S}{\sigma_z^2 \cdot (\mu_{4,z} - \sigma_z^4) - \mu_{3,z}^2} \quad (43b)$$

Provided z is normally distributed before selection, the skew $\mu_{3,z} = 0$ and the kurtosis $\mu_{4,z} - \sigma_z^4 = 2\sigma_z^4$. In this case, $b_1 = \beta$ and $b_2 = \gamma/2$, giving the univariate version of the **Pearson-Lande-Arnold regression**,

$$w = 1 + \beta(z - \mu_z) + \frac{\gamma}{2} \left((z - \mu_z)^2 - \sigma_z^2 \right) + e \quad (44)$$

developed by Lande and Arnold (1983), motivated by Pearson (1903). The Pearson-Lande-Arnold regression provides a connection between selection differentials (directional and stabilizing) and quadratic approximations of the individual fitness surface.

An important point from (43a) is that if skew is present ($\mu_{3,z} \neq 0$), $b_1 \neq \beta$ and the slope term in the linear regression (the best *linear* fit) of $w(z)$ differs from the linear slope term in the quadratic regression (the best *quadratic* fit) of $w(z)$. This arises because the presence of skew generates a covariance between z and $(z - \mu_z)^2$. The biological significance of this can be seen by reconsidering Figure 3, wherein the presence of skew in the phenotypic distribution results in a change in the mean of a population under strict stabilizing selection (as measured by the population mean being at the optimum of the individual fitness surface). Skew generates a correlation between z and $(z - \mu_z)^2$ so that selection acting only $(z - \mu_z)^2$ generates a correlated change in z . From the Robertson-Price identity (23), the within-generation change in mean equals the covariance between phenotypic value and relative fitness. Since covariances measure the amount of *linear* association between variables, in describing the change in mean, the correct measure is the slope of the best *linear* fit of the individual fitness surface. If skew is present, using the linear slope b_1 from the quadratic regression to describe

the change in mean is incorrect, as this quadratic regression removes the effects on relative fitness from a linear change in z due to the correlation between z and $(z - \mu_z)^2$.

5. MEASURING MULTIVARIATE SELECTION

Typically, investigation of the effects of natural selection involves measuring a suite of characters for each individual. In this setting, the phenotype of an individual is a vector $\mathbf{z} = (z_1, z_2, \dots, z_n)^T$ of the n measured character values. The major aim of a multivariate character analysis is to partition the total change in a trait into the effect from direct selection and the change due selection on phenotypically correlated characters. Indeed, an unstated assumption of a univariate analysis is that there is no indirect selection on the trait induced by direct selection on unmeasured correlated characters. This assumption is also required for a multivariate analysis, as one cannot correct for the effects of any *unmeasured* selected characters phenotypically correlated to one (or more) of the traits under consideration.

Our development is based on the multiple regression approach of Lande and Arnold (1983). Similar approaches based on path analysis have also been suggested e.g., (Maddox and Antonovics 1983, Mitchell-Olds 1987, Crespi and Bookstein 1988). We denote the mean vector and covariance matrix of \mathbf{z} before selection by $\boldsymbol{\mu}$ and \mathbf{P} , and by $\boldsymbol{\mu}^*$ and \mathbf{P}^* after selection (but before reproduction).

5.1 Changes in the Mean Vector: The Directional Selection Differential \mathbf{S}

The multivariate extension of the directional selection differential is the vector

$$\mathbf{S} = \boldsymbol{\mu}^* - \boldsymbol{\mu} \tag{45a}$$

whose i th element is S_i , the differential for character z_i . As with the univariate case, the Robertson-Price identity (23) holds, so that the elements of \mathbf{S} represent the covariance between character value and relative fitness, $S_i = \sigma(z_i, w)$. This immediately implies (26) that the opportunity for selection

I bounds the range of S_i ,

$$\frac{|S_i|}{\sigma_{z_i}} \leq \sqrt{I} \quad (45b)$$

=====> **Figure 4 here** < =====

As is illustrated in Figure 4, \mathbf{S} confounds the direct effects of selection on a character with the indirect effects due to selection on phenotypically correlated characters. Suppose that character 1 is under direct selection to increase in value while character 2 is not directly selected. As Figure 4 shows, if z_1 and z_2 are uncorrelated, there is no within-generation change in μ_2 (the mean of z_2). However, if z_1 and z_2 are positively correlated, individuals with large values of z_1 also tend to have large values of z_2 , resulting in a within-generation increase in μ_2 . Conversely, if z_1 and z_2 are negatively correlated, selection to increase z_1 results in a within-generation decrease in μ_2 . Hence, a character not under selection can still experience a within-generation change in its phenotypic distribution due to selection on a phenotypically correlated character (indirect selection). Fortunately, the **directional selection gradient** $\beta = \mathbf{P}^{-1}\mathbf{S}$ accounts for indirect selection due to phenotypic correlations, providing a less biased picture of the nature of directional selection acting on \mathbf{z} .

5.2. The Directional Selection Gradient β

From regression theory, the vector of partial regression coefficients for predicting the value of y given a vector of observations \mathbf{z} is $\mathbf{P}^{-1}\boldsymbol{\sigma}(\mathbf{z}, y)$ where \mathbf{P} is the covariance matrix of \mathbf{z} , and $\boldsymbol{\sigma}(\mathbf{z}, y)$ is the vector of covariances between the elements of \mathbf{z} and the variable y . Since $\mathbf{S} = \boldsymbol{\sigma}(\mathbf{z}, w)$, it immediately follows that

$$\mathbf{P}^{-1}\boldsymbol{\sigma}(\mathbf{z}, w) = \mathbf{P}^{-1}\mathbf{S} = \beta \quad (46)$$

is the vector of partial regression for the best linear regression of relative fitness w on phenotypic value \mathbf{z} , viz.,

$$w(\mathbf{z}) = 1 + \sum_{j=1}^n \beta_j(z_j - \mu_j) = 1 + \beta^T(\mathbf{z} - \boldsymbol{\mu}) \quad (47)$$

Since β_j is a partial regression coefficient, it represents the change generated in relative fitness by changing z_j while holding all other character values in \mathbf{z} constant — a one unit increase in z_j (holding all other characters constant) increases the expected relative fitness by β_j . Provided all characters under selection are measured (i.e., we can exclude the possibility of unmeasured characters influencing fitness that are phenotypically correlated with \mathbf{z}), a character under no directional selection has $\beta_j = 0$ — when all other characters are held constant, the best linear regression predicts no change in expected fitness as we change the value of z_j . Thus, β accounts for the effects of phenotypic correlations only among the *measured* characters. Provided we measure all characters under selection and there are no confounding environmental or genetic effects, β provides an unbiased measure of the amount of directional selection acting on each character.

Since $\mathbf{S} = \mathbf{P}\beta$, we have

$$S_i = \sum_{j=1}^n \beta_j P_{ij} = \beta_i P_{ii} + \sum_{j \neq i}^n \beta_j P_{ij} \quad (48)$$

illustrating that the directional selection differential confounds direct selection on that character (β_i) with indirect contributions due to selection on phenotypically correlated characters (β_j). Equation (48) implies that if two characters are phenotypically uncorrelated ($P_{ij} = 0$), selection on one has no within-generation effect on the phenotypic mean of the other.

5.3. Directional Gradients, Fitness Surface Geometry and Selection Response

When phenotypes are multivariate normal (MVN), β provides a convenient descriptor of the geometry of both the individual and mean population fitness surfaces. Lande (1976, 1979) showed that

$$\beta = \nabla_{\boldsymbol{\mu}}[\ln \bar{W}(\boldsymbol{\mu})] = \bar{W}^{-1} \cdot \nabla_{\boldsymbol{\mu}}[\bar{W}(\boldsymbol{\mu})] \quad (49)$$

which holds provided fitnesses are frequency-independent. Here $\nabla_{\boldsymbol{\mu}}$ is the **gradient** (or **gradient**

vector) with respect to the vector of population means,

$$\nabla_{\boldsymbol{\mu}}[f] = \frac{\partial f}{\partial \boldsymbol{\mu}} = \begin{pmatrix} \frac{\partial f}{\partial \mu_1} \\ \vdots \\ \frac{\partial f}{\partial \mu_n} \end{pmatrix}$$

The gradient at point $\boldsymbol{\mu}_o$ corresponds to a vector indicating the direction of steepest ascent of the function at that point (the multivariate slope of f at the point $\boldsymbol{\mu}_o$). Thus $\boldsymbol{\beta}$ is the gradient of mean population fitness with respect to the mean vector $\boldsymbol{\mu}$. Since $\boldsymbol{\beta}$ gives the direction of steepest increase in the mean population fitness surface, mean population fitness increases most rapidly when the change in the vector of means $\Delta\boldsymbol{\mu} = \boldsymbol{\beta}$. If fitnesses are frequency-dependent (individual fitnesses change as the population mean changes), then for $\mathbf{z} \sim \text{MVN}$,

$$\boldsymbol{\beta} = \nabla_{\boldsymbol{\mu}}[\ln \bar{W}(\boldsymbol{\mu})] + \int \nabla_{\boldsymbol{\mu}}[w(\mathbf{z})] \phi(\mathbf{z}) d\mathbf{z} \quad (50)$$

where the second term accounts for the effects of frequency-dependence and ϕ is the MVN density function (Lande 1976). Here $\boldsymbol{\beta}$ does not point in the direction of steepest increase in \bar{W} unless the second integral is zero. If we instead consider the individual fitness surface, we can alternatively express $\boldsymbol{\beta}$ as the gradient of individual fitnesses averaged over the population distribution of phenotypes,

$$\boldsymbol{\beta} = \int \nabla_{\mathbf{z}}[w(\mathbf{z})] \phi(\mathbf{z}) d\mathbf{z} \quad (51)$$

which holds provided $\mathbf{z} \sim \text{MVN}$ (Lande and Arnold 1983). Equation (51) holds regardless of whether fitness is frequency dependent or frequency independent.

Kingsolver et al. (2001) and Hoekstra et al. (2001) review estimates of $\boldsymbol{\beta}$ (the univariate elements of $\boldsymbol{\beta}$) from selection in natural populations. They observed that the distribution for $|\boldsymbol{\beta}|$ is close to exponential, with estimated selection gradients tending to be small (they found the median value for $|\boldsymbol{\beta}|$ was 0.15). Interestingly, they found that the selection gradients for sexual selection (mate choice) tend to be larger than those for viability selection (survivorship). Most of the difference was due to a much higher frequency of values of $|\boldsymbol{\beta}| < 0.1$ for viability selection. They further

noted that estimated β values are highly correlated with their estimated selection intensities, i . Since i is a univariate measure that incorporates both direct and indirect effects of selection, while β removes any indirect effects from *measured* characters, they conclude that indirect selection is generally modest at best. The caveat with this conclusion is that correlated character under direct selection, but not included in the study, can falsely create an impression of direct selection on the actual traits measured.

5.4. Changes in the Covariance Matrix: The Quadratic Selection Differential \mathbf{C}

Motivated by the univariate case (34), define the multivariate **quadratic selection differential** to be a square ($n \times n$) matrix \mathbf{C} whose elements are the covariances between all pairs of quadratic deviations $(z_i - \mu_{z_i})(z_j - \mu_{z_j})$ and relative fitness w , viz.,

$$C_{ij} = \sigma[w, (z_i - \mu_{z_i})(z_j - \mu_{z_j})] \quad (52a)$$

Lande and Arnold (1983) showed that

$$\mathbf{C} = \sigma[w, (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T] = \mathbf{P}^* - \mathbf{P} + \mathbf{S}\mathbf{S}^T \quad (52b)$$

If no quadratic selection is acting, the covariance between each quadratic deviation and fitness is zero and $\mathbf{C} = \mathbf{0}$. In this case (52b) gives

$$P_{ij}^* - P_{ij} = -S_i S_j \quad (53)$$

demonstrating that the $S_i S_j$ term corrects C_{ij} for the change in covariance caused by directional selection alone.

As was the case for \mathbf{S} , the fact that C_{ij} is a covariance immediately allows us to bound its range using the opportunity for selection. Since $\sigma^2(x, y) \leq \sigma^2(x) \sigma^2(y)$,

$$C_{ij}^2 \leq I \sigma^2[(z_i - \mu_{z_i})(z_j - \mu_{z_j})] \quad (54a)$$

When z_i and z_j are bivariate normal, then (Kendall and Stuart 1983),

$$\sigma^2[(z_i - \mu_{z_i})(z_j - \mu_{z_j})] = P_{ij}^2 + P_{ii} P_{jj} = P_{ij}^2(1 + \rho_{ij}^{-2}) \quad (54b)$$

where ρ_{ij} is the phenotypic covariance between z_i and z_j . Hence, for Gaussian-distributed phenotypes,

$$\left| \frac{C_{ij}}{P_{ij}} \right| \leq \sqrt{I} \sqrt{1 + \rho_{ij}^{-2}} \quad (55)$$

which is a variant of the original bound based on I suggested by Arnold (1986).

5.5. The Quadratic Selection Gradient γ

Like \mathbf{S} , \mathbf{C} confounds the effects of direct selection with selection on phenotypically correlated characters. However, as was true for \mathbf{S} , these indirect effects can also be removed by a regression. Consider the best quadratic regression of relative fitness as a function of multivariate phenotypic value,

$$w(\mathbf{z}) = a + \sum_{j=1}^n b_j z_j + \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n \gamma_{jk} (z_j - \mu_j)(z_k - \mu_k) \quad (56a)$$

$$= a + \mathbf{b}^T \mathbf{z} + \frac{1}{2} (\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\gamma} (\mathbf{z} - \boldsymbol{\mu}) \quad (56b)$$

Using multiple regression theory, Lande and Arnold (1983) showed that the matrix $\boldsymbol{\gamma}$ of quadratic partial regression coefficients is given by

$$\boldsymbol{\gamma} = \mathbf{P}^{-1} \boldsymbol{\sigma}[w, (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T] \mathbf{P}^{-1} = \mathbf{P}^{-1} \mathbf{C} \mathbf{P}^{-1} \quad (57)$$

This is the **quadratic selection gradient** and (like $\boldsymbol{\beta}$) removes the effects of phenotypic correlations, providing a more accurate picture of how selection is operating on the multivariate phenotype.

The vector \mathbf{b} of linear coefficients for the quadratic regression need not equal the vector $\boldsymbol{\beta}$ of partial regression coefficients obtained by assuming only a linear regression. As was true for the univariate case (43) when skew is present in at least one of the components of the phenotypic

distribution, \mathbf{b} is a function of both \mathbf{S} and \mathbf{C} , while β is only a function of \mathbf{S} . When skew is absent, $\mathbf{b} = \beta$, recovering the multivariate version of the Pearson-Lande-Arnold regression,

$$w(\mathbf{z}) = a + \beta^T \mathbf{z} + \frac{1}{2} (\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\gamma} (\mathbf{z} - \boldsymbol{\mu}) \quad (58)$$

Since the γ_{ij} are partial regression coefficients, they predict the change in expected fitness caused by changing the associated quadratic deviation while holding all other variables constant. Increasing $(z_j - \mu_j)(z_k - \mu_k)$ by one unit in such a way as to hold all other variables and pairwise combinations of characters constant, relative fitness is expected to increase by γ_{jk} for $j \neq k$ and by $\gamma_{jj}/2$ if $j = k$ (the difference arises because $\gamma_{jk} = \gamma_{kj}$, so that γ_{jk} appears twice in the regression unless $j = k$). The coefficients of $\boldsymbol{\gamma}$ thus describe the nature of selection on quadratic deviations from the mean for both single characters and pairwise combinations of characters. $\gamma_{ii} < 0$ implies fitness is expected to decrease as z_i moves away (in either direction) from its mean. As was discussed previously, this is a necessary, *but not sufficient*, condition for stabilizing selection on character i . Similarly, $\gamma_{ii} > 0$ implies fitness is expected to increase as z_i moves away from its mean, again a necessary, but not sufficient, condition for disruptive selection. Kingsolver et al. (2001) examined the distribution of estimated γ_{ii} values reported in the literature. Surprisingly, they found this distribution to be symmetric about zero, implying the disruptive selection ($\gamma_{ii} > 0$) is reported as often as stabilizing selection ($\gamma_{ii} < 0$), which is quite contrary to the widely-held view that stabilizing selection is perhaps the most common mode of selection in natural populations. Further, most of the estimated γ_{ii} values are small, with the median value for $|\gamma|$ being 0.10. Thus, the widely held notion of strong stabilizing selection being common in nature is (presently) not supported by the published data.

Turning to combinations of characters, non-zero values of γ_{jk} ($j \neq k$) suggest the presence of **correlation selection** — $\gamma_{jk} > 0$ suggests selection for a positive correlation between characters j and k , while $\gamma_{jk} < 0$ suggests selection for a negative correlation. Although it seems straightforward to infer the overall nature of selection by looking at these various pairwise combinations, this can

give a very misleading picture about the geometry of the fitness surface (Figure 5). We discuss this problem and its solution in Section 5.8.

Finally, we can see the effects of phenotypic correlations on the quadratic selection differential. Solving for \mathbf{C} by post- and pre-multiplying γ by \mathbf{P} gives $\mathbf{C} = \mathbf{P} \gamma \mathbf{P}$, or

$$C_{ij} = \sum_{k=1}^n \sum_{\ell=1}^n \gamma_{k\ell} P_{ik} P_{\ell j} \quad (59)$$

showing that within-generation changes in phenotypic covariance, as measured by \mathbf{C} , are influenced by quadratic selection on phenotypically-correlated characters.

5.6. Quadratic Gradients, Fitness Surface Geometry and Selection Response

When phenotypes are multivariate-normally distributed, γ provides a measure of the average curvature of the individual fitness surface, as

$$\gamma = \int \mathbf{H}_{\mathbf{z}}[w(\mathbf{z})] \phi(\mathbf{z}) d\mathbf{z} \quad (60a)$$

where $\mathbf{H}_{\mathbf{z}}[f]$ denotes the Hessian of the function f with respect to \mathbf{z} and is a multivariate measure of the quadratic curvature of a function (recall that the ij th element of the Hessian is $\partial^2 f / \partial z_i \partial z_j$). This result, due to Lande and Arnold (1983), holds for both frequency-dependent and frequency-independent fitnesses. When fitnesses are frequency-independent (again provided $\mathbf{z} \sim \text{MVN}$), γ provides a description of the curvature of the **log mean population** fitness surface, with

$$\mathbf{H}_{\boldsymbol{\mu}}[\ln \bar{W}(\boldsymbol{\mu})] = \gamma - \boldsymbol{\beta} \boldsymbol{\beta}^T \quad (60b)$$

This result is due to Lande (cited in Phillips and Arnold 1989) and points out that there are two sources for curvature in the mean fitness surface: $-\boldsymbol{\beta} \boldsymbol{\beta}^T$ from directional selection and γ from quadratic selection.

The careful reader will have undoubtedly noted a number of similarities between direction and quadratic gradients. Table 1 summarizes these and other analogous features of measuring changes in means and covariances.

=====> Table 1 here < =====

5.7. Estimation, Hypothesis Testing and Confidence Intervals

Even if we can assume that a best-fit quadratic is a reasonable approximation of the individual fitness surface, we are still faced with a number of statistical problems. Unless we test for, and confirm, multivariate normality, β and γ must be estimated from separate regressions — β from the best linear regression, γ from the best quadratic regression. In either case, there are a large number of parameters to estimate — γ has $n(n + 1)/2$ terms and β has n terms, for a total $n(n + 3)/2$. With 5, 10, and 25 characters, this corresponds to 20, 65 and 350 parameters. The number of observations should greatly exceed $n(n + 3)/2$ in order to estimate these parameters with any precision.

A second problem is multicollinearity — if some of the characters being measured are highly correlated with each other, the phenotypic covariance matrix \mathbf{P} can be nearly singular, so that small errors in estimating \mathbf{P} result in large differences in \mathbf{P}^{-1} , which in turn gives a very large sampling variances for the estimates of β and γ . One possibility is to use principal components to extract a subset of the characters (measured as PCs, linear combinations of the characters) that explains most of the phenotypic variance of \mathbf{P} . Fitness regressions using the first few PCs as the characters can then be computed. This approach also reduces the problem of the number of parameters to estimate, but risks the real possibility of removing the most important characters. Further, PCs are often difficult to interpret biologically. While the first PC for morphological characters generally corresponds to a general measure of size, the others are typically much more problematic. Finally, using PCs can spread the effects of selection on one character over several PCs, further complicating interpretation.

Another statistical concern with fitness regressions is that residuals are generally not normal — viability data typically have binomially-distributed residuals, while count data are often Poisson. Hence, many of the standard approaches for constructing confidence intervals on regression coefficients are not appropriate. Mitchell-Olds and Shaw (1987) suggest using the delete-one jack-

knife method for approximating confidence intervals in this case. Randomization tests and cross-validation procedures are also other approaches. Multivariate tests of the presence of a single mode in the fitness surface are discussed by Mitchell-Olds and Shaw (1987).

5.8. Geometric Interpretation of the Quadratic Fitness Regression

In spite of their apparent simplicity, multivariate quadratic fitness regressions have a rather rich geometric structure. Scaling characters so that they have mean zero, the general quadratic fitness regression can be written as

$$w(\mathbf{z}) = a + \sum_{i=1}^n b_i z_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \gamma_{ij} z_i z_j = a + \mathbf{b}^T \mathbf{z} + \frac{1}{2} \mathbf{z}^T \boldsymbol{\gamma} \mathbf{z} \quad (61)$$

If $\mathbf{z} \sim \text{MVN}$, then $\mathbf{b} = \boldsymbol{\beta}$ (the vector of coefficients of the best *linear* fit). Note that if we regard (61) as a second-order Taylor series approximation of $w(z)$, \mathbf{b} and $\boldsymbol{\gamma}$ can be interpreted as the gradient and hessian of individual fitness evaluated at the population mean (here $\boldsymbol{\mu} = \mathbf{0}$ by construction). The nature of curvature of (61) is determined by the matrix $\boldsymbol{\gamma}$. Even though a quadratic is the simplest curved surface, its geometry can still be very difficult to visualize.

We start our exploration of this geometry by considering the gradient of this best-fit quadratic fitness surface, which is

$$\nabla_{\mathbf{z}}[w(\mathbf{z})] = \mathbf{b} + \boldsymbol{\gamma} \mathbf{z} \quad (62)$$

Thus, at the point \mathbf{z} the direction of steepest ascent on the fitness surface (the direction in which to move in phenotype space to maximally increase individual fitness) is given by the vector $\mathbf{b} + \boldsymbol{\gamma} \mathbf{z}$ (when $\boldsymbol{\mu} = \mathbf{0}$). If the true individual fitness surface is indeed a quadratic, then the average gradient of individual fitness taken over the distribution of phenotypes is

$$\int \nabla_{\mathbf{z}}[w(\mathbf{z})] p(\mathbf{z}) d\mathbf{z} = \mathbf{b} \int p(\mathbf{z}) d\mathbf{z} + \boldsymbol{\gamma} \int \mathbf{z} p(\mathbf{z}) d\mathbf{z} = \mathbf{b} \quad (63)$$

as the last integral is $\boldsymbol{\mu}$ (which is zero by construction). Hence, if the true fitness function is quadratic, the average gradient of individual fitness is given by \mathbf{b} , independent of the distribution of \mathbf{z} .

Solving $\nabla_{\mathbf{z}}[w(\mathbf{z})] = \mathbf{0}$, a point \mathbf{z}_o is a candidate for a local extremum (stationary point) if $\gamma \mathbf{z}_o = -\mathbf{b}$. When γ is nonsingular,

$$\mathbf{z}_o = -\gamma^{-1}\mathbf{b} \quad (64a)$$

is the unique stationary point of this quadratic surface. Substituting into (61), the expected individual fitness at this point is

$$w_o = a + \frac{1}{2} \mathbf{b}^T \mathbf{z}_o \quad (64b)$$

as obtained by Phillips and Arnold (1989). Since $\partial^2 w(\mathbf{z})/\partial z_i \partial z_j = \gamma_{ij}$, the hessian of $w(\mathbf{z})$ is just γ . Thus \mathbf{z}_o is a local minimum if γ is positive-definite (all eigenvalues are positive), a local maximum if γ is negative-definite (all eigenvalues are negative), or a saddle point if the eigenvalues differ in sign. If γ is singular (has at least one zero eigenvalue) then there is no unique stationary point. An example of this is seen in Figure 5B, where there is a ridge (rather than a single point) of phenotypic values having the highest fitness value. The consequence of a zero eigenvalue is that the fitness surface has no curvature along the axis defined by the associated eigenvector. If γ has k zero eigenvalues, then the fitness surface has no curvature along k dimensions. Ignoring fitness change along these dimensions, the remaining fitness space has only a single stationary point, which is given by (64a) for γ and \mathbf{b} reduced to the $n - k$ dimensional subsurface showing curvature. While the curvature is completely determined by γ , it is easy to be misled about the actual nature of the fitness surface if one simply tries to infer it by inspection of γ , as the Figure 5 illustrates.

=====> **Figure 5 here** < =====

Canonical Form of the Quadratic Fitness Surface

As Figure 5 shows, visualizing the individual fitness surface (even for two characters) is not trivial and can easily be downright misleading. The problem is that the cross-product terms (γ_{ij} for $i \neq j$) make the quadratic form difficult to interpret geometrically. Removing these terms by a change of variables so that the axes of new variables coincide with the axes of symmetry of the quadratic form

(its canonical axes) greatly facilitates visualization of the fitness surface. Motivated by this, Phillips and Arnold (1989) suggest using two slightly different versions of the canonical transformation of γ to clarify the geometric structure of the best fitting quadratic fitness surface. If we consider the matrix \mathbf{U} whose columns are the eigenvectors of γ , the transformation $\mathbf{y} = \mathbf{U}^T \mathbf{z}$ (hence $\mathbf{z} = \mathbf{U} \mathbf{y}$ since $\mathbf{U}^{-1} = \mathbf{U}^T$ as \mathbf{U} is orthonormal, and $\mathbf{U}^T \gamma \mathbf{U} = \mathbf{\Lambda}$, a diagonal matrix whose elements correspond to the eigenvalues of γ) removes all cross-product terms in the quadratic form,

$$\begin{aligned}
w(\mathbf{z}) &= a + \mathbf{b}^T \mathbf{U} \mathbf{y} + \frac{1}{2} (\mathbf{U} \mathbf{y})^T \gamma (\mathbf{U} \mathbf{y}) \\
&= a + \mathbf{b}^T \mathbf{U} \mathbf{y} + \frac{1}{2} \mathbf{y}^T (\mathbf{U}^T \gamma \mathbf{U}) \mathbf{y} \\
&= a + \mathbf{b}^T \mathbf{U} \mathbf{y} + \frac{1}{2} \mathbf{y}^T \mathbf{\Lambda} \mathbf{y} \\
&= a + \sum_{i=1}^n \theta_i y_i + \frac{1}{2} \sum_{i=1}^n \lambda_i y_i^2
\end{aligned} \tag{65}$$

where $\theta_i = \mathbf{b}^T \mathbf{e}_i$, $y_i = \mathbf{e}_i^T \mathbf{z}$, with λ_i and \mathbf{e}_i the eigenvalues and associated unit eigenvectors of γ . Alternatively, if a stationary point \mathbf{z}_0 exists (e.g., γ is nonsingular), the change of variables $\mathbf{y} = \mathbf{U}^T (\mathbf{z} - \mathbf{z}_0)$ further removes all linear terms (Box and Draper 1987), so that

$$w(\mathbf{z}) = w_o + \frac{1}{2} \mathbf{y}^T \mathbf{\Lambda} \mathbf{y} = w_o + \frac{1}{2} \sum_{i=1}^n \lambda_i y_i^2 \tag{66}$$

where $y_i = \mathbf{e}_i^T (\mathbf{z} - \mathbf{z}_0)$ and w_o is given by (64b). Equation (65) is called the **A canonical form** and (66) the **B canonical form** (Box and Draper 1987). Both forms represent a rotation of the original axis to the new set of axes (the canonical axes of γ) that align them with axes of symmetry of the quadratic surface. The B canonical form further shifts the origin to the stationary point \mathbf{z}_o . Since the contribution to individual fitness from $\mathbf{b}^T \mathbf{z}$ is a hyperplane, its effect is to tilt the fitness surface. The B canonical form “levels” this tilting, allowing us to focus entirely on the curvature (quadratic) aspects of the fitness surface.

The orientation of the quadratic surface is determined by the eigenvectors of γ while the eigenvalues of γ determine the nature and amount of curvature of the surface along each canonical axis. Along the axis defined by y_i , the individual fitness function has positive curvature (is concave) if

$\lambda_i > 0$, has negative curvature (is convex) if $\lambda_i < 0$, or has no curvature (is a plane) if $\lambda_i = 0$. The amount of curvature is indicated by the magnitude of λ_i , the larger $|\lambda_i|$ the more extreme the curvature.

Returning to Figure 5, we see that the axes of symmetry of the quadratic surface are the canonical axes of γ . For $\gamma_{12} = 0.25$ (Figure 5A), $\lambda_1 = -2.06$ and $\lambda_2 = -0.94$ so that the fitness surface is convex along each canonical axis, with more extreme curvature along the y_1 axis. When $\gamma_{12} = \sqrt{2}$ (Figure 5B), one eigenvalue is zero while the other is -3 , so that the surface shows no curvature along one axis (it is a plane), but is strongly convex along the other. Finally, when $\gamma_{12} = 4$ (Figure 5C), the two eigenvalues differ in sign, being -5.53 and 2.53 . This generates a saddle point with the surface being concave along one axis and convex along with other, with the convex curvature being more extreme.

If $\lambda_i = 0$, the fitness surface along y_i has no curvature, so that the fitness surface is a ridge along this axis. If $\theta_i = \mathbf{b}^T \mathbf{e}_i > 0$ this is a rising ridge (fitness increases as y_i increases), it is a falling ridge (fitness decrease as y_i increase) if $\theta_i < 0$, and is flat if $\theta_i = 0$. Again returning to Figure 5, the effect of \mathbf{b} is to tilt the fitness surface. Denoting values on the axis running along the ridge by y_1 , if $\theta_1 > 0$ the ridge rises so that fitness increases as y_1 increases. Even if γ is not singular, it may be nearly so, with some of the eigenvalues being very close to zero. In this case, the fitness surface shows little curvature along the axes given by the eigenvectors associated with these near-zero eigenvalues. The fitness change along a particular axis (here given by \mathbf{e}_i) is $\theta_i y_i + (\lambda_i/2) y_i^2$. If $|\theta_i| \gg |\lambda_i|$, the curvature of the fitness surface along this axis is dominated by the effects of linear (as opposed to quadratic) selection. Phillips and Arnold (1989) present a nice discussion of several other issues relating to the visualization of multivariate fitness surfaces, while Box and Draper (1987) review the statistical foundations of this approach.

5.9. Unmeasured Characters and Other Biological Caveats

Even if we are willing to assume that the best-fitting quadratic regression is a reasonable approxi-

mation of the individual fitness surface, there are still a number of important biological caveats to keep in mind. For example, the fitness surface can change in both time and space, often over short spatial/temporal scales (e.g., Kalisz 1986, Stewart and Schoen 1987, Scheiner 1989, Jordan 1991), so that one estimate of the fitness surface may be quite different from another estimation at a different time and/or location. Hence, considerable care must be used before pooling data from different times and/or sites to improve the precision of estimates. When the data are such that selection gradients can be estimated separately for different times or areas, space/time by gradient interactions can be tested for in a straightforward fashion (e.g., Mitchell-Olds and Bergelson 1990).

Population structure can also influence fitness surface estimation in other ways. If the population being examined has overlapping generations, fitness data must be adjusted to reflect this (e.g., Stratton 1992). Likewise, if members in the population differ in their amount of inbreeding, measured characters and fitness may show a spurious correlation if both are affected by inbreeding depression (Willis 1996).

Perhaps the most severe caveat for the regression approach of estimating $w(z)$ is unmeasured characters — estimates of the amount of direct selection acting on a character are biased if that character is phenotypically correlated to unmeasured characters also under selection (Lande and Arnold 1983, Mitchell-Olds and Shaw 1987). Adding one or more of these unmeasured characters to the regression can change initial estimates of β and γ . Conversely, selection acting on unmeasured characters that are phenotypically *uncorrelated* with those being measured has no effect on estimates of β and γ .

6. MULTIPLE TRAIT SELECTION

6.1. Short-Term Changes in Means: The Multivariate Breeders' Equation

By combining (8) and (46), we can express the multivariate breeders equation as

$$\mathbf{R} = \mathbf{G}\mathbf{P}^{-1}\mathbf{S} = \mathbf{G}\boldsymbol{\beta} \quad (67)$$

Presented in this form, this nicely demonstrates the relationship between the two main complications with selection on multiple characters: the *within-generation* change due to *phenotypic* correlations and the *between-generation* change (response to selection) due to *additive genetic* correlations. Cheverud (1984) makes the important point that although it is often assumed a set of phenotypically correlated traits responses to selection in a coordinated fashion, this is not necessarily the case. Since $\boldsymbol{\beta}$ removes all the effects of phenotypic correlations and $\mathbf{R} = \mathbf{G}\boldsymbol{\beta}$, phenotypic characters will only response as a group if they are all under direct selection or if they are *genetically* correlated.

6.2. The effects of genetic correlations: direct and correlated responses

While the use of $\boldsymbol{\beta}$ removes any further evolutionary effect of phenotypic correlations, additive genetic correlations strongly influence the response to selection. If n characters are under selection, and g_{ij} is the additive genetic covariance between traits i and j , then the response in character i to a single generation of selection is

$$R_i = \Delta\mu_i = \sum_{j=1}^n g_{ij} \beta_j = g_{ii} \beta_i + \sum_{j \neq i} g_{ij} \beta_j \quad (68)$$

so that response has a component due to direct selection on that character ($g_{ii} \beta_i$) plus an addition component due to selection on all other genetically correlated characters.

If direct selection only occurs on character i , a genetically correlated character also shows a **correlated response to selection**. For example, for trait j , its response when only trait i is under selection is

$$R_j = g_{ij} \beta_i \quad (69a)$$

Thus, the ratio of the expected change in two characters when only one is under direct selection is

$$\frac{R_j}{R_i} = \frac{g_{ij}}{g_{ii}} = \rho_{ij}^g \sqrt{\frac{g_{jj}}{g_{ii}}} \quad (69b)$$

If both characters have the same additive genetic variance ($g_{ii} = g_{jj}$), the ratio of response simply reduced to ρ_{ij}^g , the correlation between additive genetic values.

6.3. Evolutionary constraints imposed by genetic correlations

One immediate consequence of (67) is that a character under selection does not necessarily change in the direction most favored by natural selection. For example, fitness may maximally increase if μ_2 decreases, so that $\beta_2 < 0$. However, if the sum of correlated responses on z_2 is sufficiently positive, then μ_2 will increase. Thus, a character may be dragged off its optimal fitness trajectory by a correlated response generated by stronger selection on other traits. As these other characters approach their equilibrium values, their β values approach zero, reducing their indirect effects on other characters, eventually allowing β_2 to dominate the dynamics (and hence ultimately μ_2 decreases).

Recall that the direction the mean vector must be changed to give the greatest change in mean fitness is β (as β is the gradient of \bar{W} with respect to the mean vector, see Equation 49). However, generally $\mathbf{R} \neq \beta$, and thus the effect of the additive-genetic covariance matrix \mathbf{G} is to constrain the selection response from its optimal value. The mean vector changes in the direction most favored by selection if and only if

$$\mathbf{G}\beta = \lambda\beta \tag{70}$$

which only occurs when β is an eigenvector (λ) of \mathbf{G} . Note that even if \mathbf{G} is a diagonal matrix (there is no correlation between the additive genetic values of the characters under selection) Equation (70) is usually not satisfied. In fact, only when we can write $\mathbf{G} = \sigma_A^2 \mathbf{I}$ is (70) satisfied for arbitrary β . Thus, only when all characters have the same additive genetic variance and there is no additive genetic covariance between characters is the response to selection in the directional most favored by natural selection. Differences in the amounts of additive genetic variances between characters, in addition to non-zero additive-genetic covariances, also impose constraints on character evolution.

6.4. Inferring the Nature of Previous Selection

Providing the breeders' equation holds over several generations, the cumulative response to m generations of selection is just

$$R^{(m)} = \sum_{t=1}^m R(t) = \sum_{t=1}^m \mathbf{G}(t)\boldsymbol{\beta}(t) \quad (71a)$$

In particular, if the genetic covariance matrix remains constant, then

$$R^{(m)} = \mathbf{G} \left(\sum_{t=1}^m \boldsymbol{\beta}(t) \right) \quad (71b)$$

Hence, if we observe a total response to selection of \mathbf{R}_{total} , then we can estimate the cumulative selection differential as

$$\boldsymbol{\beta}_{total} = \sum \boldsymbol{\beta}(t) = \mathbf{G}^{-1}\mathbf{R}_{total} \quad (71c)$$

For example, if one observes a difference in the mean multivariate phenotype between two diverged populations, then provided we can assume the genetic covariance matrix remains roughly constant, the history of past selection leading to the divergence (assuming it is entirely due to selection) can be inferred from (71c).

6.5. Changes in \mathbf{G} under the Infinitesimal Model

Within-generation Change in \mathbf{G}

If the vector of breeding values \mathbf{g} is multivariate-normal with covariance matrix \mathbf{G} before an episode of selection, then it can be shown (e.g., Lande and Arnold 1983) that the covariance matrix \mathbf{G}^* of breeding values after selection (but before reproduction) satisfies

$$\mathbf{G}^* - \mathbf{G} = \mathbf{G}\mathbf{P}^{-1}(\mathbf{P}^* - \mathbf{P})\mathbf{P}^{-1}\mathbf{G} \quad (72a)$$

we show below (75) that

$$\mathbf{G}^* - \mathbf{G} = \mathbf{G}(\boldsymbol{\gamma} - \boldsymbol{\beta}\boldsymbol{\beta}^T)\mathbf{G} \quad (72b)$$

Thus, when the breeders' equation holds, γ and β are sufficient to describe how phenotypic selection changes (within a generation) the additive-genetic covariance matrix. Equation (72b) shows that the within-generation change in \mathbf{G} has a component from directional selection and a second due from quadratic selection, as

$$\mathbf{G}^* - \mathbf{G} = -\mathbf{G}\beta\beta^T\mathbf{G} + \mathbf{G}\gamma\mathbf{G} = -\mathbf{R}\mathbf{R}^T + \mathbf{G}\gamma\mathbf{G} \quad (72c)$$

In terms of the change in covariance for two particular characters, this can be factored as

$$\begin{aligned} G_{ij}^* - G_{ij} &= -\left(\sum_{k=1}^n \beta_k G_{ik}\right) \left(\sum_{k=1}^n \beta_k G_{jk}\right) + \sum_{k=1}^n \sum_{\ell=1}^n \gamma_{k\ell} G_{ik} G_{\ell j} \\ &= -\Delta\mu_i \cdot \Delta\mu_j + \sum_{k=1}^n \sum_{\ell=1}^n \gamma_{k\ell} G_{ik} G_{\ell j} \end{aligned} \quad (73a)$$

Thus the within-generation change in the additive genetic variance of character i is given by

$$G_{ii}^* - G_{ii} = -(\Delta\mu_i)^2 + \sum_{k=1}^n \sum_{\ell=1}^n \gamma_{k\ell} G_{ik} G_{i\ell} \quad (73b)$$

The Response (Between-generation change) in \mathbf{G}

While (72) and (73) describe how the covariance structure of breeding values changes within a generation, the between-generation change must consider how recombination alters any gametic-phase disequilibrium generated by selection. Recall our treatment of the change in the additive variance in the univariate case (Section 2.2), wherein all changes in variances occur due to gametic-phase disequilibrium d_t changing the additive variance. In this case, $\sigma_A^2(t) = \sigma_A^2(0) + d_t$, and hence $\sigma_z^2(t) = \sigma_z^2(0) + d_t$. Moving to multiple characters, disequilibrium is now measured by the matrix $\mathbf{D}_t = \mathbf{G}_t - \mathbf{G}_0$. Thus D_{ij} is the change in the additive genetic covariance between characters i and j induced by disequilibrium. Likewise, the phenotypic covariance matrix is given by $\mathbf{P}_t = \mathbf{P}_0 + \mathbf{D}_t$. Tallis (1987, Tallis and Leppard 1988b) extend Bulmer's result to multiple characters, showing that

$$\Delta\mathbf{D}_t = \frac{1}{2} (\mathbf{G}_t \mathbf{P}_t^{-1} (\mathbf{P}_t^* - \mathbf{P}_t) \mathbf{P}_t^{-1} \mathbf{G}_t - \mathbf{D}_t) \quad (74)$$

The first term represents the within-generation change in \mathbf{G} , the second the decay in \mathbf{D} due to recombination.

We can also express (74) in terms of the quadratic and directional selection gradients, γ and β .

Recalling the definitions of \mathbf{C} , β , and γ , we have

$$\mathbf{P}^* - \mathbf{P} = \mathbf{C} - \mathbf{s}\mathbf{s}^T$$

Hence

$$\begin{aligned} \mathbf{P}^{-1}(\mathbf{P}^* - \mathbf{P})\mathbf{P}^{-1} &= \mathbf{P}^{-1}(\mathbf{C} - \mathbf{s}\mathbf{s}^T)\mathbf{P}^{-1} \\ &= \mathbf{P}^{-1}\mathbf{C}\mathbf{P}^{-1} - (\mathbf{P}^{-1}\mathbf{s})(\mathbf{P}^{-1}\mathbf{s})^T \\ &= \gamma - \beta\beta^T \end{aligned} \tag{75}$$

Substituting into (74) gives

$$\Delta\mathbf{D}_t = \frac{1}{2} \left(\mathbf{G}_t(\gamma_t - \beta_t\beta_t^T)\mathbf{G}_t - \mathbf{D}_t \right) \tag{76a}$$

Finally, recalling (72b), we can also express this as

$$\Delta\mathbf{D}_t = \frac{1}{2} (\mathbf{G}_t^* - \mathbf{G}_t) - \frac{\mathbf{D}_t}{2} \tag{76b}$$

Equation (76b) shows that only half of the disequilibrium deviation in \mathbf{G} from the previous generation (\mathbf{D}_t) is passed onto the next generation, as is half the within-generation change induced by selection in that generation. This is expected since the infinitesimal mode assumes unlinked loci and hence each generation recombination removes half of the present disequilibrium. The disequilibrium component converges to

$$\hat{\mathbf{D}} = \hat{\mathbf{G}} \left(\hat{\gamma} - \hat{\beta}\hat{\beta}^T \right) \hat{\mathbf{G}} = (\mathbf{G}_0 + \hat{\mathbf{D}}) \left(\hat{\gamma} - \hat{\beta}\hat{\beta}^T \right) (\mathbf{G}_0 + \hat{\mathbf{D}}) \tag{77}$$

Where we have placed carets over γ and β to remind the reader that these depend on \mathbf{P} , and hence on \mathbf{D} .

6.6. Effects of Drift and Mutation.

Drift

If \mathbf{G}_t changes strictly by drift, the multivariate analog to (16) is

$$\mathbf{G}_t = \left(1 - \frac{1}{2N_e}\right)^t \mathbf{G}_0 \quad (78)$$

This assumes that only additive genetic variance is present. Hence, ignoring the effect of disequilibrium, under the infinitesimal model the expected change after t generations of selection and drift is

$$\begin{aligned} \boldsymbol{\mu}_t - \boldsymbol{\mu}_0 &= \sum_{k=0}^t \mathbf{G}_k \boldsymbol{\beta} = \mathbf{G}_0 \boldsymbol{\beta} \sum_{k=0}^t \left(1 - \frac{1}{2N_e}\right)^k \\ &\simeq 2N_e (1 - e^{-t/2N_e}) \mathbf{G}_0 \boldsymbol{\beta} \end{aligned} \quad (79)$$

Mutation

Let \mathbf{U} be the matrix of per-generation input from mutations to the additive genetic variances and covariance. The expected equilibrium covariance matrix under drift and mutation becomes

$$\widehat{\mathbf{G}} = 2N_e \mathbf{U} \quad (80)$$

Under the assumptions of the infinitesimal model (selection is sufficiently weak that the dynamics of changes in allele frequencies are completely driven by drift and mutation), the additive genetic variance-covariance matrix at generation t is

$$\mathbf{G}_t = \widehat{\mathbf{G}} + \left(\mathbf{G}_0 - \widehat{\mathbf{G}}\right) e^{-t/2N_e} \quad (81)$$

giving the cumulative response to t generations of constant directional selection $\boldsymbol{\beta}$ as

$$\mathbf{R}_t = \boldsymbol{\mu}_t - \boldsymbol{\mu}_0 = \left[t \widehat{\mathbf{G}} + 2N_e (1 - e^{-t/2N_e}) \left(\mathbf{G}_0 - \widehat{\mathbf{G}}\right) \right] \boldsymbol{\beta} \quad (82a)$$

Note if there is no mutational input $\mathbf{U} = 0$, then the total response by generation t is given

$$\mathbf{R}_t = 2N_e (1 - e^{-t/2N_e}) \boldsymbol{\beta} \mathbf{G}_0 = 2N_e (1 - e^{-t/2N_e}) \mathbf{R}_0 \quad (82b)$$

giving, at the limit

$$\mathbf{R}_\infty = 2N_e \mathbf{R}_0 \quad (82c)$$

$2N_e$ times the initial response, as was found (Robertson 1960) for univariate selection.

7. PHENOTYPIC EVOLUTION MODELS

The above machinery for selection response has been widely used by evolutionary biologists to model the dynamics of complex traits under selection. Under the assumptions of the infinitesimal model, one only needs to specify the nature of selection, the starting mean and the initial covariance matrix. Further genetic details do not enter (unless one wishes to include mutation, when the additional composite parameter, σ_m^2 enters). Such analyses are often called **phenotypic evolution models** given that the genetic details can be handed with a few composite parameters. As with the entire field of evolutionary quantitative genetics, this is a rather large enterprise. We will concern ourselves here to two applications: testing whether an observed pattern seen in the fossil record is due to selection or drift, and examining the consequences of stabilizing selection on mean population fitness. The goal is to give the reader a feel for the flavor of such modeling. A much more detailed treatment can be found in Walsh and Lynch (2003).

7.1. Selection vs. Drift in the Fossil Record.

One of the more interesting applications of phenotypic selection models is in trying to make inferences about the forces that shaped morphological evolution on extinct organisms. Obviously, one can not perform the required crosses to estimate heritabilities on traits of interests. However, one might reasonably use the average heritability values for similar traits in (hopefully close) relatives of the extinct species. A variety of tests for whether an observed pattern of morphological change in the fossil record is consistent with drift (as opposed to having to invoke selection) have been

proposed, and we briefly discuss a few here. An excellent overview of the many pitfalls of this general approach is given by Reymont (1991). Despite these concerns, this approach offers a powerful method of analysis to gain insight on the evolution of species long extinct.

Lande's Test

Lande (1976) offered the first test for whether an observed amount of morphological divergence was consistent with simple random genetic drift. If the mean is changing strictly under genetic drift, then if the distribution of breeding and environmental values is bivariate normal, the resulting distribution for the mean at generation t , $\mu(t)$, is normal with mean $\mu(0)$ (the initial mean) and variance $t\sigma_A^2/N_e$. Letting the observed divergence in means be $\Delta\mu = \mu(t) - \mu(0)$, then the largest effective population size consistent (at the 5% level) with this divergence being due strictly to drift is

$$N_e^* = \frac{(1.96)^2 h^2 t}{(\Delta\mu/\sigma_z)^2} = 3.84 \frac{th^2}{(\Delta\mu/\sigma_z)^2} \quad (83)$$

This follows since the probability that a unit normal exceeds 1.96 is just 5%, and we have written $\sigma_A^2 = h^2\sigma_z^2$. Thus, provided that the heritability remains roughly constant, one can apply (83) to see if the maximal effective population size required for this amount of drift is exceeded by our reasonable estimate of the actual population size. If this is the case, then we can rule out drift as the sole agent for the observed divergence.

The Turelli-Gillespie-Lande and Bookstein Tests

Turelli et al. (1988) noted that Lande's test has the hidden assumption that the additive genetic variance is independent of the population size. Recall from (18) that the equilibrium additive genetic variance under mutation and drift is $\sigma_A^2 = 2N_e\sigma_m^2$, where σ_m^2 is the polygenic mutation rate. Hence, after a sufficient number of generations, the additive variance should scale with N_e (if no

other evolutionary forces are acting), in which case the drift variance becomes

$$\sigma^2 [\mu(t)] = t \frac{\sigma_A^2}{N_e} = t \frac{2N_e \sigma_m^2}{N_e} = 2t \sigma_m^2 \quad (84)$$

The variance expression given by (84) transforms Lande's test from a critical maximal effective population size to one based on a critical minimum mutational variance. In particular, (83) is replaced by

$$\sigma_m^2 = \frac{(\Delta\mu)^2}{2t(1.96)^2} = 0.130 \frac{(\Delta\mu)^2}{t} \quad (85)$$

Drift is rejected at the 5% of the suspected value of σ_m^2 is less than this value.

Turelli et al. also make the important point that the test for drift is really a two-sided test. While the above tests (83 and 85) ask whether the observed divergence is too large under drift (as might be expected under directional selection), we can also ask if the observed divergence is *too small* under drift as would be expected if stabilizing selection constrains the change on the mean. Noting that for a unit normal random variable U that $\Pr(|U| < 0.03) = 0.025$ and $\Pr(|U| > 2.24) = 0.025$ suggests the following two-sided test: If

$$\sigma_m^2 < \frac{(\Delta\mu)^2}{2t(2.24)^2} = 0.0858 \frac{(\Delta\mu)^2}{t} \quad (86a)$$

then the rate of evolution is too fast to be explained by drift, while if

$$\sigma_m^2 > \frac{(\Delta\mu)^2}{2t(0.03)^2} = 509 \frac{(\Delta\mu)^2}{t} \quad (86b)$$

the rate of divergence is too slow for drift.

If one has access to a sample of generation means (as opposed to a starting and end points), then Bookstein (1989) notes that a more powerful test can be obtained (from the theory of random walks) by considering $\Delta\mu^*$, the largest absolute deviation from the starting mean anywhere along the series of t generation means. Here, the rate of evolution is too fast for drift if

$$\sigma_m^2 < \frac{(\Delta\mu^*)^2}{2t(2.50)^2} = 0.080 \frac{(\Delta\mu^*)^2}{t} \quad (87a)$$

while it is too slow for drift when

$$\sigma_m^2 > \frac{(\Delta\mu^*)^2}{2t(0.56)^2} = 1.59 \frac{(\Delta\mu^*)^2}{t} \quad (87a)$$

Note by comparison to (86a) and (86b) that the tests for too fast a divergence are very similar, while Bookstein's test for too slow a divergence (a potential signature of stabilizing selection) is much less stringent than the Turelli-Gillespie-Lande test.

7.2. Stabilizing Selection.

The Non-optimal fitness function.

A widely-used fitness function in the evolutionary genetics literature is the **non-optimal** or **normalizing** fitness function;

$$W(z) = \exp\left(-\frac{z^2}{2\omega^2}\right) \quad (88)$$

This is proportional to a normal density function with mean zero and variance ω^2 , and implies that fitness decreases with increasing values of z^2 , with the maximal fitness occurring for an individual with trait value $z = 0$. Typically, one scales the trait so that the optimal value occurs as zero, but more generally we can replace z^2 with $(z - \mu_o)^2$, where the optimal fitness now occurs at μ_o . Besides being a reasonable model for stabilizing selection, the main reason for the popularity of (88) in the literature is that if the phenotypic distribution is normal (with mean μ and variance σ^2) before selection, then the distribution of trait values after selection is normal with mean and variance

$$\mu^* = \frac{\omega^2 \mu}{\sigma^2 + \omega^2} \quad \text{and} \quad \sigma_*^2 = \frac{\omega^2 \sigma^2}{\sigma^2 + \omega^2} \quad (89a)$$

The expected response to selection (measured by the change in mean) in generation t is

$$R(t) = h^2(t)S(t) = h^2(t) [\mu^*(t) - \mu(t)] = -h^2(t) \frac{\mu(t) \sigma^2(t)}{\sigma^2(t) + \omega^2} \quad (89b)$$

Selection thus drives the mean towards the optimal value of $\mu = 0$. When μ is non-zero, there is both directional selection (non-zero S) in addition to stabilizing selection

Turning to the dynamics of the variance, the change in the phenotypic variance (due to stabilizing selection generating disequilibrium) is given by (12),

$$\begin{aligned}
\Delta d(t) &= d(t+1) - d(t) \\
&= - \left(\frac{d(t)}{2} + \frac{h^2(t)}{2} [\sigma^2(t) - \sigma_*^2(t)] \right) \\
&= - \left(\frac{d(t)}{2} + \frac{h^2(t)\sigma^4(t)}{2(\sigma^2(t) + \omega^2)} \right)
\end{aligned} \tag{89b}$$

Recalling (11a, 11b) that $\sigma_A^2(t) = \sigma_a^2 + d(t)$ and $\sigma_z^2(t) = \sigma_z^2 + d(t)$, then at equilibrium ($\Delta d(t) = 0$), Equation (89b) reduces to a quadratic equation whose admissible solution is given by

$$\hat{\sigma}_z^2 = \sigma_z^2 \left(1 - \frac{2 + h^2 + \omega^2/\sigma_z^2 - \sqrt{8h^2 + (2 + h^2 + \omega^2/\sigma_z^2)^2}}{4} \right) \tag{89c}$$

Here σ_z^2 and h^2 are the values for the base population before selection. Figure 6 shows the reduction in the phenotypic variance at equilibrium. As expected, the reduction is more severe with either higher heritabilities and/or stronger selection (smaller values of ω^2/σ^2).

=====> **Figure 6 here** < =====

The cost of stabilizing selection.

Charlesworth (1984) examined the so-called cost of stabilizing selection by computing the selective load — the reduction in the mean fitness from the maximal value ($W = 1$ in this case). The load is given by $L = 1 - \bar{W}$ and measures the fraction of the population suffering selective mortality in a particular generation. The mean population fitness is obtained by integrating the fitness function over the distribution of phenotypes,

$$\bar{W} = \int p(z) W(z) dz = \sqrt{\frac{\omega^2}{\sigma^2 + \omega^2}} \exp \left[-\frac{\mu^2}{2(\sigma^2 + \omega^2)} \right] \tag{90a}$$

as found by Charlesworth. After selection has driven the phenotypic mean to zero, the load becomes

$$L = 1 - \sqrt{\frac{\omega^2}{\sigma^2 + \omega^2}} = 1 - \sqrt{\frac{\Lambda}{1 + \Lambda}}, \quad \text{where } \Lambda = \omega^2/\sigma^2 \tag{90b}$$

Charlesworth's expression for the load (90b) ignores the reduction in the phenotypic variance due to the generation of linkage disequilibrium (89c). A more correct expression for the load (when the population is at equilibrium) is

$$L = 1 - \sqrt{\frac{\omega^2}{\sigma_*^2 + \omega^2}} = 1 - \sqrt{\frac{\Lambda}{\sigma_*^2/\sigma^2 + \Lambda}} \quad (90b)$$

where σ_*^2 is the equilibrium phenotypic variance. Figure 7 plots the corrected load for various values of initial heritability and different strengths of selection ω^2/σ^2 . As expected, the selective loads are reduced when the disequilibrium is taken into account, as the negative disequilibrium generated by selection reduces the phenotypic variance, and this narrowing of the phenotypic variance increases the mean fitness. The reduction in the load is most dramatic for strong selection ($\omega^2 \ll \sigma^2$) and high heritabilities.

=====> **Figure 7 here** < =====

8. THEOREMS OF NATURAL SELECTION: FUNDAMENTAL AND OTHERWISE

Our final comments in this brief overview of evolutionary quantitative genetics concern the ultimate quantitative trait: individual fitness. What general statements, if any, can we make about the behavior of multilocus systems under selection? One oft-quoted is Fisher's **fundamental theorem of natural selection**, which states that "The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time" This simple statement from Fisher's 1930 (reprinted in 1958) book (which was dictated to his wife as he paced about their living room) has generated a tremendous amount of work, discussion, and sometimes heated arguments. Fisher claimed his result was exact, a true theorem. The common interpretation of Fisher's theorem, that the rate of increase in fitness equals the additive variance in fitness, has been referred to by Karlin as "neither fundamental nor a theorem" as it requires rather special conditions, especially when multiple loci influence fitness.

Since variances are non-negative, the classical interpretation of Fisher's theorem (namely,

$\Delta\bar{W} = \sigma_A^2(W)/\bar{W}$) implies that mean population fitness never decreases in a constant environment. As we discuss below, this interpretation of Fisher's theorem holds *exactly* only under restricted conditions, but is often a good approximate descriptor. However, an important corollary holds under very general conditions (Kimura 1965; Nagylaki 1976, 1977b, Ewens 1976, Ewens and Thompson 1977, Charlesworth 1987): in the absence of new variation from mutation or other sources such as migration, *selection is expected to eventually remove all additive genetic variation in fitness*. If the population is at equilibrium, the average excesses (in fitness) for each allele is zero as all segregating alleles have the same marginal fitness and hence $\sigma_A^2 = 0$ (Fisher 1941).

8.1. The Classical Interpretation of Fishers' Fundamental Theorem

We first review the "classical" interpretation and then discuss what Fisher actually seems to have meant.

To motivate Fisher's theorem for one locus, consider a diallelic (alleles **A** and **a**) locus with constant fitnesses under random mating. Let p denote the frequency of **A**, and denote the fitnesses for the three genotypes as W_{AA} , W_{Aa} , and W_{aa} . The mean population fitness is given by

$$\bar{W} = p^2W_{AA} + 2p(1-p)W_{Aa} + (1-p)^2W_{aa} \quad (91a)$$

and the change in allele frequency can be expressed in terms of Wright's formula as

$$\Delta p = \frac{p(1-p)}{2\bar{W}} \frac{d\bar{W}}{dp} = \frac{p(1-p)}{2} \frac{d \ln \bar{W}}{dp} \quad (91b)$$

Turning now to fitness, if the allele frequency change Δp is small, a first order Taylor series approximation gives

$$\bar{W}(p + \Delta p) = \bar{W}(p + \Delta p) + \bar{W}(p) \simeq \bar{W}(p) + \frac{\partial \bar{W}}{\partial p} \Delta p \quad (92a)$$

Applying (91b),

$$\Delta \bar{W} \simeq \frac{\partial \bar{W}}{\partial p} \Delta p = \frac{p(1-p)}{2\bar{W}} \left(\frac{\partial \bar{W}}{\partial p} \right)^2 \quad (92b)$$

From (91a),

$$\begin{aligned}\frac{\partial \bar{W}}{\partial p} &= 2pW_{AA} + 2(1-2p)W_{Aa} + 2(p-1)W_{aa} \\ &= 2[p(W_{AA} - W_{Aa}) + (1-p)(W_{Aa} - W_{aa})] = 2(\alpha_A - \alpha_a)\end{aligned}$$

where the last equality follows the average effects of alleles A and a on fitness,

$$\alpha_A = pW_{AA} + (1-p)W_{Aa} - \bar{W} \quad \text{and} \quad \alpha_a = pW_{AA} + (1-p)W_{Aa} - \bar{W}$$

The quantity $\alpha = \alpha_A - \alpha_a$ is the **average effect of an allelic substitution**, the difference in the average effects of these two alleles gives the mean effect on fitness from replacing a randomly-chosen a allele with an A allele. The additive genetic variance is related to α by $\sigma_A^2 = 2p(1-p)\alpha^2$, giving

$$\Delta \bar{W} \simeq \frac{p(1-p)(2\alpha)^2}{2\bar{W}} = \frac{\sigma_A^2(W)}{\bar{W}} \quad (93)$$

Hence, the change in mean population fitness is proportional to the additive genetic variance in fitness.

As an example of the implication of Fisher's theorem, consider a diallelic locus with overdominance in fitness, $W_{AA} = 1$ $W_{Aa} = 1 + s$ $W_{aa} = 1$, giving the additive and dominance variance in fitness as

$$\sigma_A^2(W) = 2p(1-p)s^2(1-2p)^2, \quad \text{and} \quad \sigma_D^2(W) = [2p(1-p)s]^2$$

As plotted in Figure 8, these variances change dramatically with p . The maximum genetic variance in fitness occurs at $p = 1/2$, but none of this variance is additive, and heritability in fitness is zero. It is easily shown that $\Delta p = 0$ when $p = 1/2$, and at this frequency $\sigma_A^2(W) = 0$, as the corollary of Fisher's theorem predicts. Thus, even though *total* genetic variation in fitness is maximized at $p = 1/2$, no change in \bar{W} occurs as the *additive* genetic variance in fitness is zero at this frequency.

=====> **Figure 8 here** < =====

Even if Fisher's theorem holds exactly, its implication for character evolution can often be misinterpreted as the following example illustrates. Suppose that locus A completely determines a character under stabilizing selection. Let the genotypes AA , Aa , and aa have discrete phenotypic

values of $z = -1, 0,$ and $1,$ respectively (so that this locus is strictly additive) and let the fitness function be $W(z) = 1 - sz^2$. If we assume no environmental variance, this generates very nearly the same fitnesses for each genotype as those given above, as the fitnesses can be normalized as $1 : (1 - s)^{-1} : 1,$ where $(1 - s)^{-1} \simeq 1 + s$ for small s . The additive genetic variance for the trait z is maximized at $p = 1/2,$ precisely the allele frequency at which the additive genetic variance in fitness $\sigma_A^2(W) = 0$. This stresses that Fisher's theorem concerns additive genetic variance in *fitness*, not in the *character*. In this example, the transformation of the phenotypic character value z to fitness W takes a character that is completely additive and introduces dominance when fitness is considered.

The above derivation of Fisher's theorem was only approximate. Under what conditions does the classical interpretation actually hold? While it is correct for a single locus with random mating, a single locus with no dominance under nonrandom mating (Kempthorne 1957), and multiple additive loci (no dominance or epistasis, Ewens 1969), it is generally compromised by nonrandom mating and departures from additivity (such as dominance or epistasis). Even when the theorem does not hold exactly, how good an approximation is it? Nagylaki (1976, 1977a,b, 1991, 1992, 1993) has examined ever more general models of fitnesses when selection is weak (the fitness of any genotype can be expressed as $1 + a s$ with s small and $|a| \ll 1$) and mating is random. Selection is further assumed to be much less than the recombination frequency c_{min} for the closest pair of loci ($s \ll c_{min}$). Under these fairly general conditions, Nagylaki shows that the evolution of mean fitness falls into three distinct stages. During the first $t < 2 \ln s / \ln(1 - c_{min})$ generations, the effects of any initial disequilibrium are moderated, first by reaching a point where the population evolves approximately as if it were in linkage equilibrium and then reaching a stage where the linkage disequilibria remain relatively constant. At this point, the change in mean fitness is

$$\Delta \bar{W} = \frac{\sigma_A^2(W)}{\bar{W}} + O(s^3) \quad (94)$$

where $O(s^3)$ means that terms on the order of s^3 have been ignored. Since additive variance is expected to be of order s^2 , Fisher's theorem is expected to hold to a good approximation during

this period. However, as gametic frequencies approach their equilibrium values, additive variance in fitness can be much less than order s^2 , in which case the error terms of order s^3 can be important. During the first and third phases, mean fitness can decrease, but the fundamental theorem holds during the central phase of evolution.

8.2. What Did Fisher Really Mean?

Fisher warned that his theorem “requires that the terms employed should be used strictly as defined”, and part of the problem stems from what Fisher meant by “fitness”. Price (1972b) and Ewens (1989, 1992) have argued that Fisher’s theorem is always true, because Fisher held a very narrow interpretation of the change in mean fitness (see also Edwards 1990, 1994; Frank 1995; Lessard and Castilloux 1995). Rather than considering the *total* rate of change in fitness, they argue that Fisher was instead concerned only with the *partial* rate of change, that due to changes in the contribution of individual alleles (specifically, changes in the average excesses/effects of these alleles). In particular, Ewens (1994) states that

“I believe that the often-made statement that the theorem concerns changes in mean fitness, assumes random-mating populations, is an approximation, and is not correct in the multi-locus setting, embodies four errors. The theorem relates the so-called partial increase in mean fitness, makes no assumption about random mating, is an exact statement containing no approximation, and finally is correct (as a theorem) no matter how many loci are involved.”

What exactly is meant by the partial increase in fitness? A nice discussion is given by Frank and Slatkin (1992), who point out that the change in mean fitness over a generation is also influenced by the change in the “environment”, E . Specifically,

$$\Delta \bar{W} = (\bar{W}' | E') - (\bar{W} | E) \quad (95a)$$

where the prime denotes the fitness/environment in the next generation, so that the change in mean fitness compounds both the change in fitness and the change in the environment. We can partition out these components by writing

$$\Delta\bar{W} = [(\bar{W}' | E) - (\bar{W} | E)] + [(\bar{W}' | E') - (\bar{W}' | E)] \quad (95b)$$

where the first term in brackets represents the change in mean fitness under the initial “environment” while the second represents the change in mean fitness due to changes in the environmental conditions. Fisher’s theorem relates solely to changes in the first component, $(\bar{W}' | E) - (\bar{W} | E)$, which he called the change in fitness due to natural selection. Price (1972b) notes that Fisher had a very broad interpretation of “environment”, referring to both physical and genetic backgrounds. In particular, Fisher

“regarded the natural selection effect on fitness as being limited to the additive or linear effects of changes in gene (allele) frequencies, while everything else — dominance, epistasis, population pressure, climate, and interactions with other species — he regarded as a matter of the environment”

Hence the change in fitness referred to by Fisher is solely that caused by changes in allele frequencies and nothing else. Changes in gamete frequencies caused by recombination and generation of disequilibrium beyond those that influence individual allele frequencies are consigned to the “environmental” category by Fisher. Given this, it is perhaps not too surprising that the strict interpretation of Fisher’s theorem holds under very general conditions as nonrandom mating and nonadditive effects generate linkage disequilibrium, changing gametic frequencies in ways not simply predictable from changes in allele frequencies.

Nagylaki (1993) suggests that the statement $\Delta\bar{W} = \sigma_A^2(W)/\bar{W}$ be referred to as the **asymptotic fundamental theorem of natural selection**, while Fisher’s more narrow (and correct) interpretation be referred to as the **Fisher-Price-Ewens theorem of natural selection**. This distinction seems quite reasonable given the considerable past history of confusion.

8.3. Heritabilities of Characters Correlated With Fitness

The corollary of Fisher's theorem, that selection drives the additive variance in fitness to zero, makes a general prediction. Characters strongly genetically correlated with fitness should thus show reduced h^2 relative to characters less correlated with fitness (Robertson 1955), reflecting the removal of additive variance by selection (which is partly countered by new mutational input). Because of the past action of natural selection, traits correlated with fitness are expected to have reduced levels of additive variance, while characters under less direct selection are expected to retain relatively greater amounts of additive variance. How well does this prediction hold up? Many authors have noticed that characters expected to be under selection (e.g., life-history characters, such as clutch size) tend, on average, to have lower heritabilities than morphological characters measured in the same population/species (reviewed by Robertson 1955, Roff and Mousseau 1987, Mousseau and Roff 1987, Charlesworth 1987). However, some notable exceptions are also apparent (Charlesworth 1987). The difficulty with these general surveys is knowing whether a character is, indeed, highly genetically correlated with lifetime fitness. Clutch size, for example, would seem to be highly correlated with total fitness, but if birds with large clutch sizes have poorer survivorship, the correlation with lifetime fitness may be weak. Negative genetic correlations between components of fitness allow significant additive variance in each component at equilibrium, even when additive variance in total fitness is zero (Robertson 1955, Rose 1982).

8.4. Robertson's Secondary Theorem of Natural Selection.

While the fundamental theorem predicts the rate of change of mean population fitness, we are often more interested in predicting the rate of change of a particular *character* under selection. Using a simple regression argument, Robertson (1966, 1968; also see Crow and Nagylaki 1976, Falconer

1985) suggested the **secondary theorem of natural selection**

$$R = \mu'_z - \mu_z = \sigma_A(z, w)$$

Namely, that *the rate of change in a character equals the additive genetic covariance between the character and fitness* (the covariance between the additive genetic value of the character and the additive genetic value of fitness). If this assertion holds, it provides a powerful connection between the fundamental theorem and the response of a character under selection. For example, it predicts that the character mean should not change when additive variance in *fitness* equals zero, regardless of the additive variance in the *character*.

To examine the conditions under which the secondary theorem holds, we first examine a single locus model in some detail before considering what can be said for increasingly general multi-locus models. Assuming random mating, a single generation of selection changes the frequency of allele \mathbf{A}_i from p_i to $p'_i = p_i(1 + s_i)$, where s_i is the average excess in relative fitness for allele \mathbf{A}_i . To map these changes in allele frequencies into changes in mean genotypic values, decompose the genotypic value of $\mathbf{A}_i\mathbf{A}_j$ as $G_{ij} = \mu_G + \alpha_i + \alpha_j + \delta_{ij}$, where α_i is the average effect of \mathbf{A}_i on character value and δ_{ij} the dominance deviation. Putting these together, the contribution from this locus to the change in mean after a generation of selection is

$$\begin{aligned} R &= \sum_{i,j} G_{ij} p'_i p'_j - \sum_{i,j} G_{ij} p_i p_j \\ &= \sum_{i,j} G_{ij} p_i(1 + s_i) p_j(1 + s_j) - \sum_{i,j} G_{ij} p_i p_j \\ &= \sum_{i,j} (\alpha_i + \alpha_j + \delta_{ij}) p_j p_i s_i + \sum_{i,j} (\alpha_i + \alpha_j + \delta_{ij}) p_i p_j s_i s_j \end{aligned} \quad (96)$$

The careful reader will note that we have already made an approximation by using G_{ij} instead of G'_{ij} in the very first sum. Since α_i and δ_{ij} are functions of the allele frequencies, they change as p_i changes, but we assume these are much smaller than the change in p_i itself (so we have assumed $\alpha'_i \simeq \alpha_i$ and $\delta'_{ij} \simeq \delta_{ij}$). A little more algebra (Nagylaki 1989, 1991) simplifies the two sums in (96) to

give the expected contribution to response as

$$R = \sum_j \alpha_j s_j p_j + \sum_{i,j} \delta_{ij} p_i s_i p_j s_j \quad (97)$$

The first sum is the expected product of the average effect α_j of an allele on character value times the average excess s_i of that allele on relative fitness. Since (by definition), $E[\alpha_j] = E[s_i] = 0$, the first sum is just the covariance between the average effect on the character with the average excess on fitness, in other words, the additive genetic covariance between *relative* fitness and character. Hence we can express (96) as

$$R = \sigma_A(z, w) + B = \frac{\sigma_A(z, W)}{\bar{W}} + B \quad (98)$$

If the character has no dominance (all $\delta_{ij} = 0$), the correction term B vanishes, recovering Robertson's original suggestion. If $\sigma_D^2(z)$ denotes the dominance variance in the trait value, Nagylaki (1989) shows that

$$|B| \leq \frac{\sigma_D(z) \cdot \sigma_A^2(w)}{2} \quad (99)$$

where $\sigma_A^2(w)$ is the additive variance in relative fitness. Assuming no epistasis, Nagylaki (1989, 1991) shows that (98) holds for n loci, with the error term being

$$B = \sum_{k=1}^n \sum_{i,j}^{n_k} \delta_{ij}(k) p_i(k) s_i(k) p_j(k) s_j(k) \quad (100)$$

where k indexes the loci, the k -th of which has n_k alleles. This also holds when linkage disequilibrium is present, but the dominance deviations and average excesses in fitness are expected to be different than those under linkage equilibrium. Nagylaki (1991) shows that (100) is bounded by

$$|B| \leq \left(\sum_{k=1}^n \sigma_{D(k)}(z) \right) \cdot \frac{\sigma_A^2(w)}{2} \quad (101a)$$

where $\sigma_{D(k)}^2(z)$ is the dominance variance in the character contributed by locus k . If all loci underlying the character are identical (the **exchangeable model**), this bound reduces to

$$|B| \leq \frac{\sigma_D(z) \cdot \sigma_A^2(w)}{2\sqrt{n}} \quad (101b)$$

Hence, for fixed amounts of genetic variances, as the number of loci increases, the error in the secondary theorem becomes increasingly smaller.

Finally, the most general statement on the validity of the secondary theorem is due to Nagylaki (1993) for weak selection and random mating, but arbitrary epistasis and linkage disequilibrium. Similar to the weak selection analysis of Fisher's theorem discussed above, after a sufficient time the change in mean fitness is given by

$$R = \sigma_A(z, w) + O(s^2) \quad (102)$$

As with the fundamental theorem, when gametic frequencies approach their equilibrium values, terms of s^2 can become significant and mean response can differ significantly from Robertson's prediction.

FINAL REMARKS AND ACKNOWLEDGMENTS

Quantitative genetics is a very broad enterprise, comprising a number of distinct subfields – plant breeding, animal breeding, human genetics, and evolutionary quantitative genetics. The focus of this review has been on some of the unique aspects of evolutionary quantitative genetics that are likely to be less well known, but nonetheless potential very useful, to practitioners of these other subfields. For example, methods for estimating the nature of natural selection on traits are certainly of interest to plant and animal breeders, as they try to account for natural selection in the face of artificial selection. Likewise, evolutionary quantitative geneticists are also starting to draw upon some of the machinery developed in other subfields (such as BLUP estimation of breeding values). This is an encouraging trend, as each subfield of quantitative genetics has developed much useful machinery that deserves to be much more widely known (e.g., Walsh 2002).

Significant portions of this review have appeared on-line in the various draft chapters posted on the webpage for Walsh and Lynch (2003).

Table 1. Analogous features of directional and quadratic differentials and gradients.

Changes in Means (Directional Selection)	Changes in Covariances (Quadratic Selection)
Differentials measure the covariance between relative fitness and phenotype	
$S_i = \sigma [w, z_i]$	$C_{ij} = \sigma [w, (z_i - \mu_i)(z_j - \mu_j)]$
The opportunity for selection bounds the differential	
$\frac{ S_i }{\sigma(z_i)} \leq \sqrt{I}$ for any distribution of \mathbf{z}	$\left \frac{C_{ij}}{P_{ij}} \right \leq \sqrt{I} \sqrt{1 + \rho_{ij}^{-2}}$ provided $\mathbf{z} \sim \text{MVN}$
Differentials confound direct and indirect selection	
$\mathbf{S} = \boldsymbol{\mu}^* - \boldsymbol{\mu} = \mathbf{P}\boldsymbol{\beta}$	$\mathbf{C} = \mathbf{P}^* - \mathbf{P} + \mathbf{S}\mathbf{S}^T = \mathbf{P}\boldsymbol{\gamma}\mathbf{P}$
Gradients measure the amount of direct selection	
$\boldsymbol{\beta} = \mathbf{P}^{-1}\mathbf{S}$	$\boldsymbol{\gamma} = \mathbf{P}^{-1}\mathbf{C}\mathbf{P}^{-1}$
Gradients describe the slope and curvature of the mean population fitness surface, provided $\mathbf{z} \sim \text{MVN}$ and fitnesses are frequency-independent	
$\beta_i = \frac{\partial \ln \bar{W}(\boldsymbol{\mu})}{\partial \mu_i}$	$\gamma_{ij} = \frac{\partial^2 \ln \bar{W}(\boldsymbol{\mu})}{\partial \mu_i \partial \mu_j} + \beta_i \beta_j$
Gradients describe the average curvature of the individual fitness surface, provided $\mathbf{z} \sim \text{MVN}$ (with $p(\mathbf{z})$ the density function),	
$\beta_i = \int \frac{\partial w(\mathbf{z})}{\partial z_i} p(\mathbf{z}) d\mathbf{z}$	$\gamma_{ij} = \int \frac{\partial^2 w(\mathbf{z})}{\partial z_i \partial z_j} p(\mathbf{z}) d\mathbf{z}$
Gradients appear as coefficients in fitness regressions	
$w(\mathbf{z}) = a + \boldsymbol{\beta}^T(\mathbf{z} - \boldsymbol{\mu})$ $\boldsymbol{\beta}$ = slope of best linear fit	$w(\mathbf{z}) = a + \mathbf{b}^T(\mathbf{z} - \boldsymbol{\mu}) + \frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\gamma}(\mathbf{z} - \boldsymbol{\mu})$ $\boldsymbol{\gamma}$ = the quadratic coefficient of the best quadratic fit. $\mathbf{b} = \boldsymbol{\beta}$ when $\mathbf{z} \sim \text{MVN}$
Gradients appear as coefficients in evolutionary equations when $(\mathbf{z}, \mathbf{g}) \sim \text{MVN}$	
$\Delta \boldsymbol{\mu} = \mathbf{G}\boldsymbol{\beta}$	$\mathbf{G}^* - \mathbf{G} = \mathbf{G}(\boldsymbol{\gamma} - \boldsymbol{\beta}\boldsymbol{\beta}^T)\mathbf{G}$

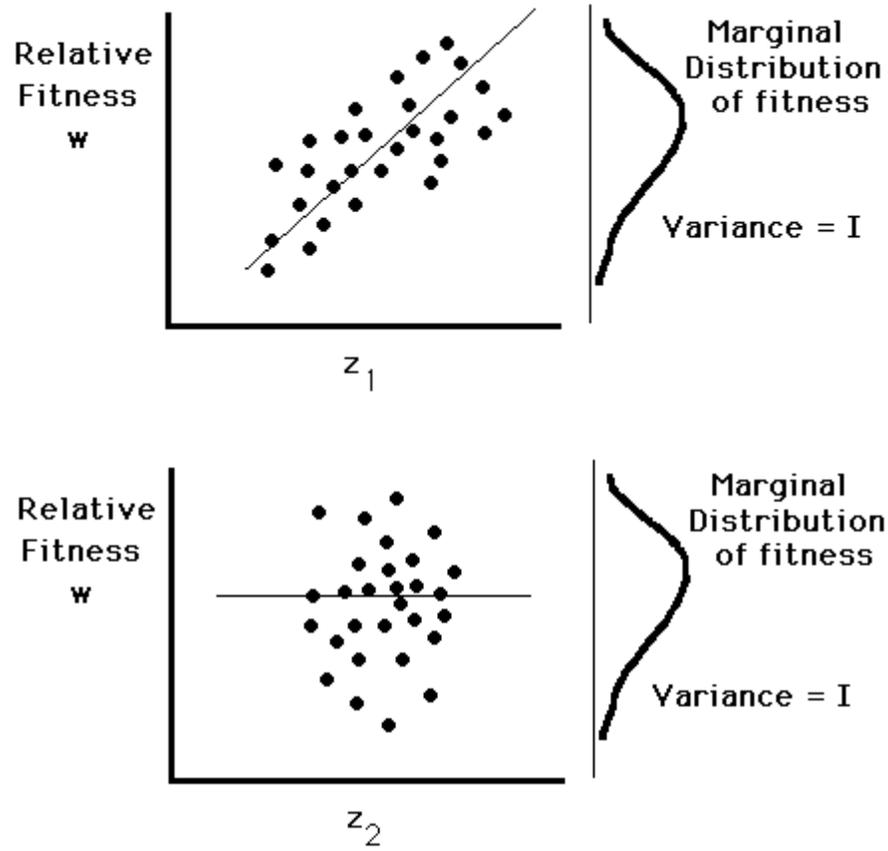


Figure 1. The fraction of the opportunity for selection I that is translated into a change in the mean depends on the correlation between relative fitness and the character. Characters z_1 and z_2 have the same marginal distribution of fitness, but only the regression of w on z_1 is significant.

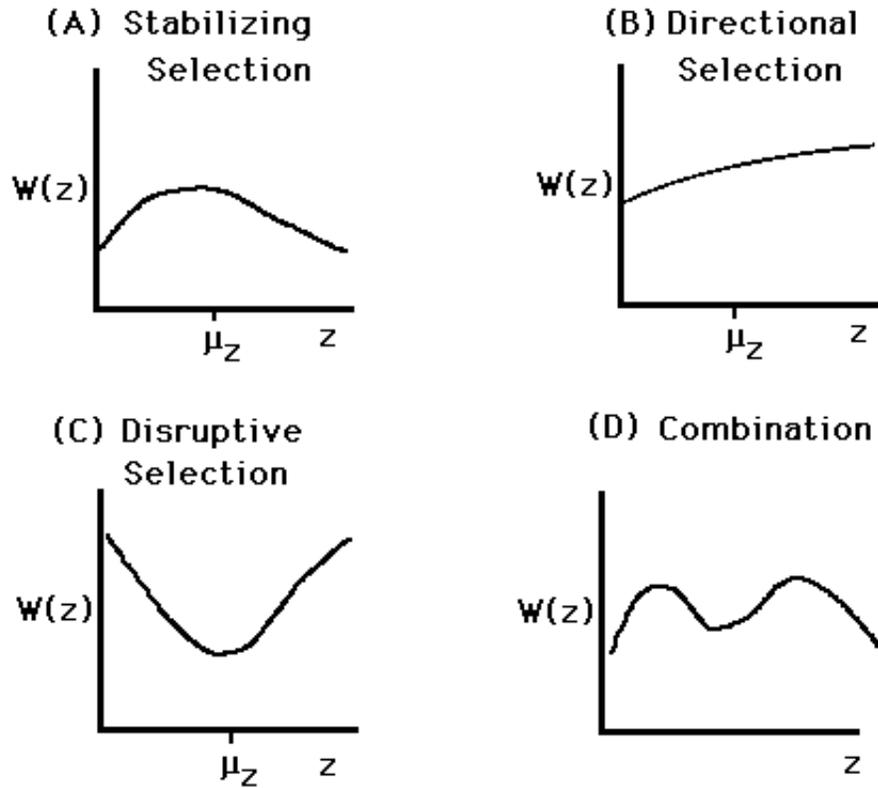


Figure 2. Selection is usually classified into three basic forms depending on the local geometry of $W(z)$: stabilizing (A), directional (B), and disruptive (C). As D illustrates, populations can simultaneously experience multiple forms of selection.

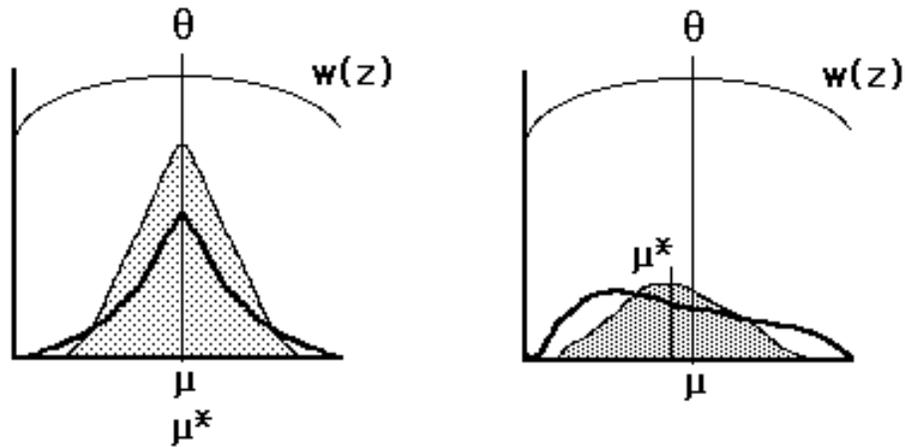


Figure 3. Even when a population is under strict stabilizing selection, the mean can change if the phenotypic distribution is skewed. A standard fitness function for stabilizing selection is $W(z) = 1 - b(\theta - z)^2$. O'Donald (1968) found that, even if the population mean is at the optimum value ($\mu_z = \theta$), S is nonnegative if the skew is nonzero ($\mu_{3,z} \neq 0$) as $S = -(b\mu_{3,z})/(1 - b\sigma_z^2)$. **Left:** If phenotypes are distributed symmetrically about the mean ($\mu_{3,z} = 0$), the distribution after selection (stippled) has the same mean when $\mu_z = \theta$. **Right:** If, however, the distribution is skewed, more of the distribution lies on one side of the mean than the other. Since the distribution is “unbalanced”, the longer tail experiences more selection than the shorter tail, changing the mean.

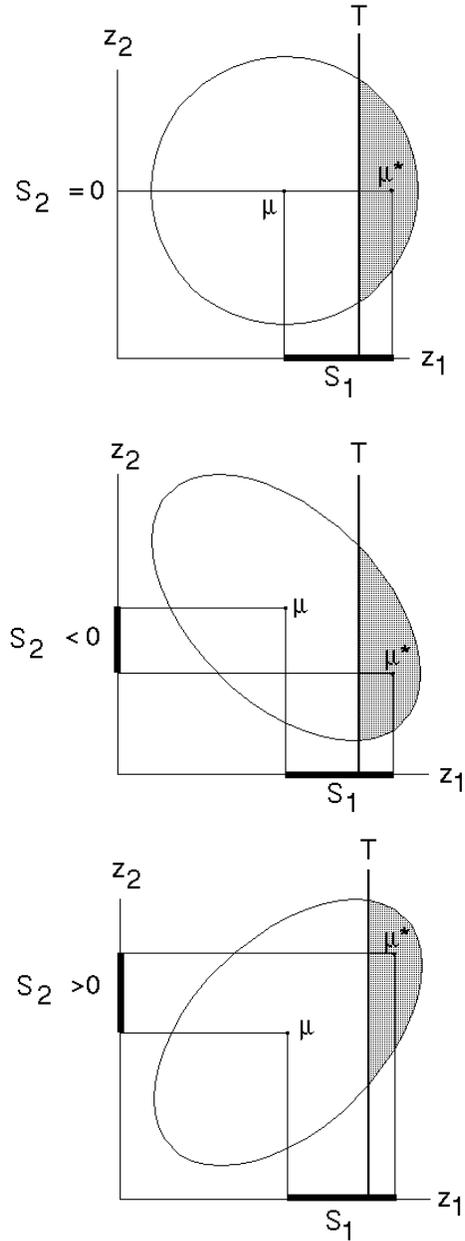


Figure 4. Selection on a character can result in a within-generation change in the mean of other phenotypically correlated characters not themselves under direct selection. Assume that character 1 is under simple truncation selection so that only individuals with $z_1 > T$ reproduce. Top: When z_1 and z_2 are uncorrelated, $S_2 = 0$. Middle: When z_1 and z_2 are negatively correlated, $S_2 < 0$. Bottom: When z_1 and z_2 are positively correlated, $S_2 > 0$.

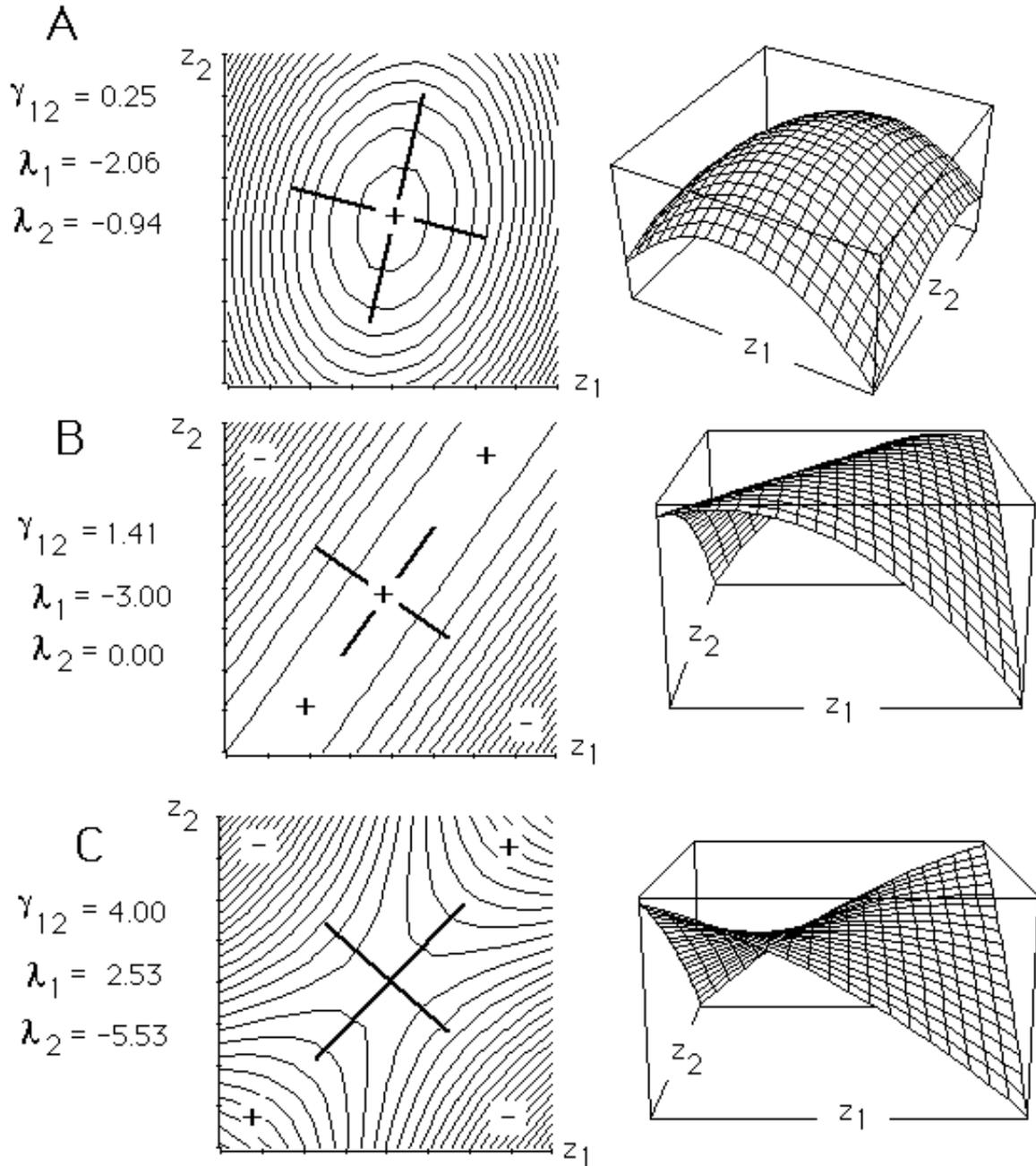


Figure 5. Three quadratic fitness surfaces, all of which have $\gamma_{11} = -2$ and $\gamma_{22} = -1$ and $\mathbf{b} = \mathbf{0}$. On the left are curves of equal fitness values, with peaks being represented by +, and valleys by -. Axes of symmetry of the surface (the canonical or principal axes of γ) are denoted by the thick lines. These axes correspond to the eigenvectors of γ . On the right are three dimensional plots of

individual fitness as a function of the phenotypic values of the characters z_1 and z_2 . **A:** $\gamma_{12} = 0.25$. This corresponds to stabilizing selection on both characters, with fitness falling off more rapidly (as indicated by the shorter distance between contour lines) along the z_1 axis than along the z_2 axis. **B:** $\gamma_{12} = \sqrt{2} \simeq 1.41$, in which case γ is singular. The resulting fitness surface shows a ridge in one direction with strong stabilizing selection in the other. **C:** When $\gamma_{12} = 4$, the fitness surface now shows a saddle point, with stabilizing selection along one of the canonical axes of the fitness surface and disruptive selection along the other.

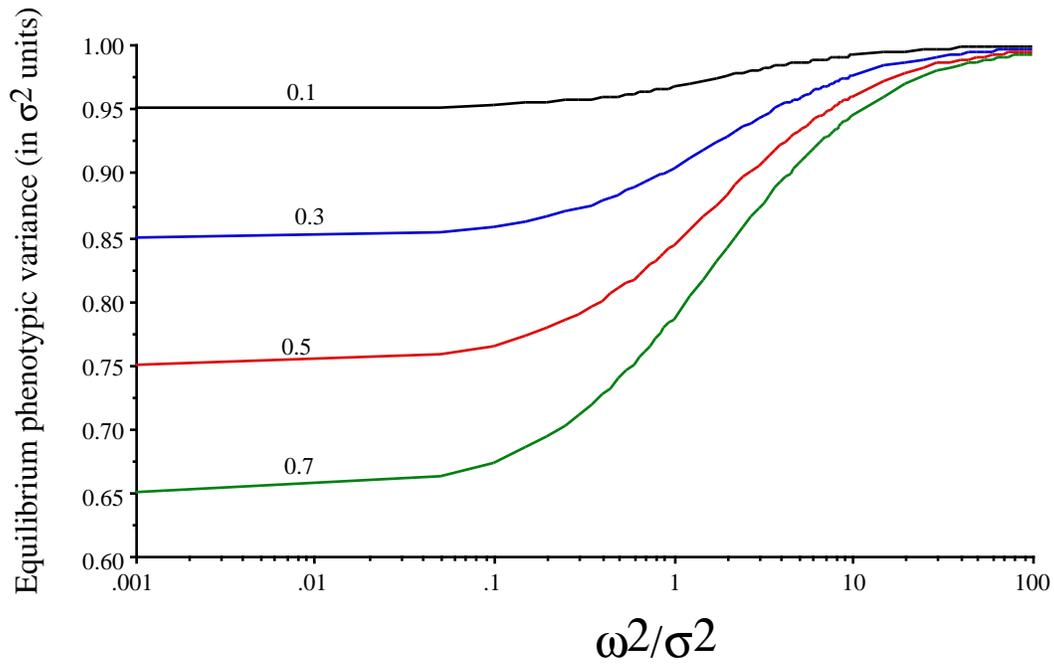


Figure 6. The equilibrium reduction in phenotypic variance under the infinitesimal model with non-optimal selection (88). The curves correspond to initial heritability values ranging from 0.1 to 0.7. Stabilizing selection creates negative disequilibrium which reduces the additive genetic (and hence phenotypic) variance below its initial value. The reduction in phenotypic variance is greatest when the initial heritability is large and/or selection is strong (the width of the selection function is narrow relative to the initial width of the phenotypic distribution, i.e., $\omega^2 \ll \sigma^2$).

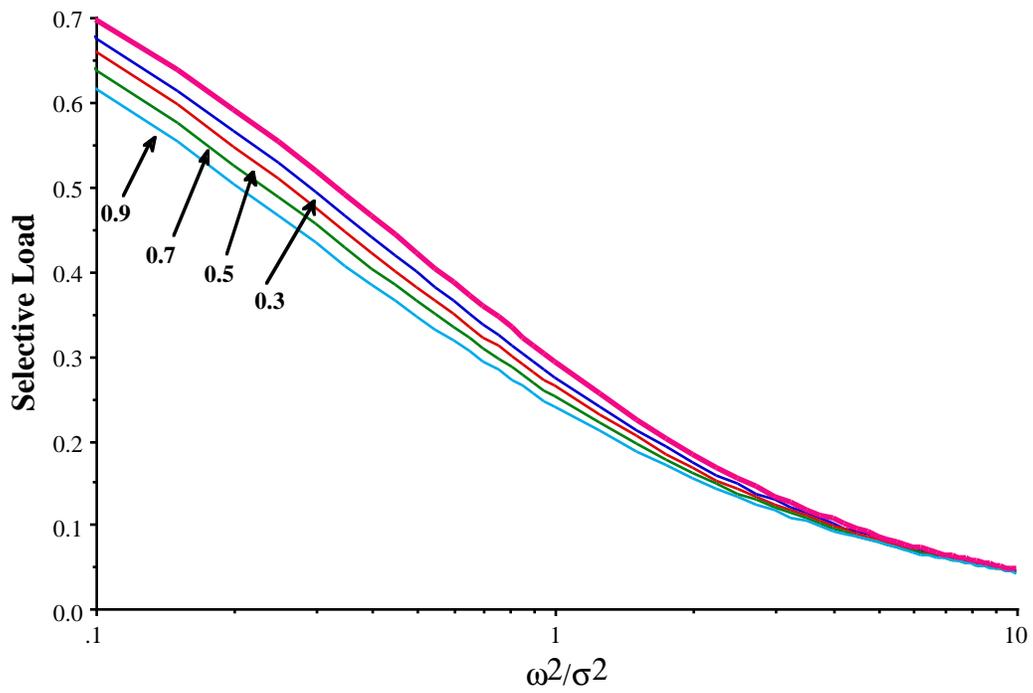


Figure 7. The selective load for an equilibrium population under non-optimal selection. The solid upper curve is Charlesworth's approximation, while the four lower curves correspond to initial heritabilities ranging from 0.3 to 0.9.

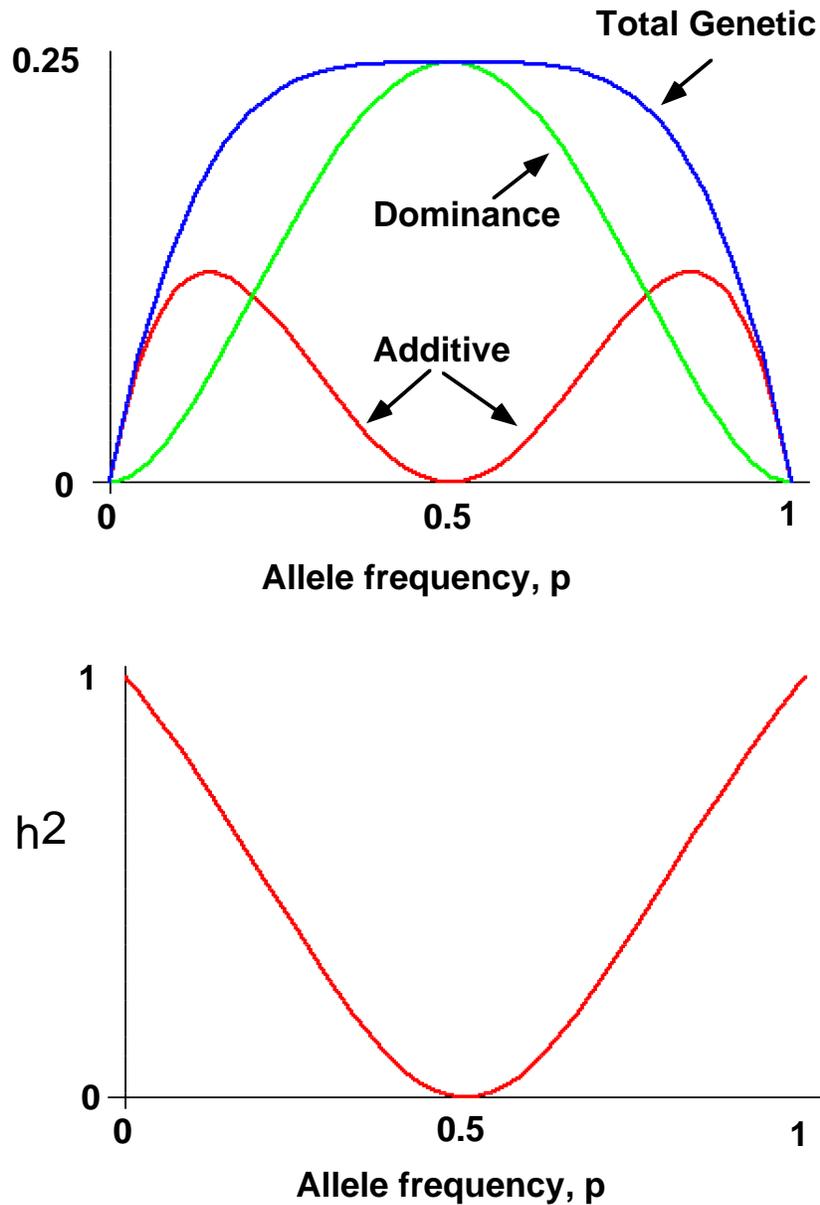


Figure 8. Genetic variances in a model with overdominance. Here, the heterozygote has fitness $1 + s$, while both homozygotes have fitness 1. As the top figure shows, the additive genetic variance is zero at allele frequency 0.5, but this is also the value where the total genetic variance is maximized. The bottom figure shows that the heritability goes to zero as the allele frequency approaches 0.5. Thus, even though the genetic variance in fitness has maximal value at allele frequency 0.5, there is no response to selection, as none of the genetic variance is additive .

REFERENCES

- Arnold, S. J. 1986. Limits on stabilizing, disruptive, and correlational selection set by the opportunity for selection. *Amer. Nat.* 128: 143–146.
- Arnold, S. J., and M. J. Wade. 1984a. On the measurement of natural and sexual selection: theory. *Evolution* 38: 709–719.
- Arnold, S. J., and M. J. Wade. 1984b. On the measurement of natural and sexual selection: applications. *Evolution* 38: 720–734.
- Banks, M. J., and D. J. Thompson. 1985. Lifetime mating success in the damselfly *Coenagrion puella*. *Anim. Behav.* 33: 1175–1183.
- Bookstein, F. L. 1989. Comment on a rate test. *Evolution* 43: 1569–1570.
- Box, G. E. P. and N. R. Draper. 1987. *Empirical model-building and response surfaces*. John Wiley and Sons, New York.
- Bulmer, M. G. 1971. The effect of selection on genetic variability. *Amer. Nat.* 105: 201–211.
- Bulmer, M. G. 1974. Linkage disequilibrium and genetic variability. *Genet. Res.* 23: 281–289.
- Bulmer, M. G. 1976. Regressions between relatives. *Genet. Res.* 28: 199–203.
- Bulmer, M. G. 1980. *The mathematical theory of quantitative genetics*. Oxford Univ. Press, NY.
- Bürger, R. 2000. *The mathematical theory of selection, recombination, and mutation*. Wiley, Chichester.
- Charlesworth, B. 1984. The cost of phenotypic evolution. *Paleobiology* 10: 319–327.
- Charlesworth, B. 1987. The heritability of fitness. In J. W. Bradbury and M. B. Andersson (eds.) *Sexual selection: testing the alternatives* pp. 21–40. Wiley, New York.
- Charlesworth, B. 1994. *Evolution in age-structured populations*. 2nd Ed. Cambridge Univ. Press, Cambridge, UK.

- Cheverud, J. M. 1984. Quantitative genetics and developmental constraints on evolution by selection. *J. Theor. Biol.* 110: 155–171.
- Clutton-Brock, T. H. (ed.) 1988. *Reproductive success: studies of individual variation in contrasting breeding systems*. University of Chicago Press, Chicago.
- Crow, J. F. 1958. Some possibilities for measuring selection intensities in man. *Hum. Biol.* 30: 1–13.
- Crow, J. F. 1989. Fitness variation in natural populations. In W. G. Hill, and T. F. C. Mackay, (eds.), *Evolution and animal breeding*, pp. 91–97. C. A. B. International, Wallingford.
- Crow, J. F., and T. Nagylaki. 1976. The rate of change of a character correlated with fitness. *Amer. Nat.* 110: 207–213.
- Crespi, B. J. and F. L. Bookstein. 1988. A path-analytic model for the measurement of selection on morphology. *Evolution* 43: 18–28.
- Downhower, J. F., L. S. Blumer, and L. Brown. 1987. Opportunity for selection: an appropriate measure for evaluating variation in the potential for selection? *Evolution* 41: 1395–1400.
- Edwards, A. W. F. 1990. Fisher, \bar{W} , and the fundamental theorem. *Theor. Pop. Biol.* 38: 276–284.
- Edwards, A. W. F. 1994. The fundamental theorem of natural selection. *Biol. Rev.* 69: 443–474.
- Ewens, W. J. 1969. A generalized fundamental theorem of natural selection. *Genetics* 63: 531–537.
- Ewens, W. J. 1976. Remarks on the evolutionary effect of natural selection. *Genetics* 83: 601–607.
- Ewens, W. J. 1989. An interpretation and proof of the fundamental theorem of natural selection. *Theo. Pop. Biol* 36: 167–180.
- Ewens, W. J. 1992. An optimizing principle of natural selection in evolutionary population genetics. *Theo. Pop. Biol* 42: 333–346.
- Ewens, W. J. 1994. The changing role of population genetics theory. In S. Levin (ed.), *Frontiers in*

Mathematical Biology, pp. 186-197. Springer NY.

Ewens, W. J. and G. Thomsom. 1977. Properties of equilibria in multi-locus genetic systems. *Genetics* 87: 807–819.

Falconer, D. S. 1985. A note on Fisher's 'average effect' and 'average excess'. *Genet. Res.* 46: 337–347.

Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to quantitative genetics*. 4th Ed. Longman Sci. and Tech., Harlow, UK.

Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Royal Soc. Edinburgh* 52: 399–433.

Fisher, R. A. 1941. Average excess and average effect of a gene substitution. *Ann. Eugen.* 11: 53–63.

Fisher, R. A. 1930. *The genetical theory of natural selection*. Clarendon Press, Oxford. (reprinted in 1958 by Dover Publications, New York).

Frank, S. A. 1995. George Price's contributions to evolutionary genetics. *J. Theor. Biol* 175: 373–388.

Frank, S. A., and M. Slatkin. 1992. Fisher's fundamental theorem of natural selection. *Trends Ecology and Evolution* 7: 92–95.

Hoekstra, H. E., J. M. Hoekstra, D. Berrigan, S. N. Vignieri, A. Hoang, C. E. Hill, P. Beerli, and J. G. Kingsolver. 2001. Strength and tempo of directional selection in the wild. *Proc. Natl. Acad. Sci.* 98: 9157–9160.

Houck, L. D. , S. J. Arnold, and R. A. Thisted. 1985. A statistical study of mate choice: Sexual selection in a plethodontid salamander (*Desmognathus ochrophaesus*). *Evolution* 39: 370–386.

Jordan, N. 1991. Multivariate analysis of selection in experimental populations derived from hybridization of two ecotypes of the annual plant *Diodia teres*. *Evolution* 45: 1760–1772.

Kalisz, S. 1986. Variable selection on the timing of germination in *Collinsia verna* (*Scrophulariaceae*). *Evolution* 40: 479–491.

- Keightley, P. D. and W. G. Hill. 1987. Directional selection and variation in finite populations. *Genetics* 117: 573–582.
- Kempthorne, O. 1957. *An introduction to genetic statistics*. Iowa State University Press, Ames.
- Kimura, M. 1965. Attainment of quasi-linkage equilibrium when gene frequencies are changing by natural selection. *Genetics* 52: 875–890.
- Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gilbert and P. Beerli. 2001. The strength of phenotypic selection in natural populations. *American Naturalist* 157: 245-261.
- Koenig, W. D., and S. S. Albano. 1986. On the measurement of sexual selection. *Amer. Nat.* 127: 403–409.
- Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30: 314–334.
- Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* 33: 402–416.
- Lande, R. 1982. A quantitative genetic theory of life history evolution. *Ecology* 63: 607–615.
- Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37: 1210–1226.
- Lenski, E. E. and P. M. Service. 1982. The statistical analysis of population growth rates calculated from schedules for survivorship and fecundity. *Ecology* 63: 655–662.
- Lessard, S., and A.-M. Castilloux. 1995. The fundamental theorem of natural selection in Ewens' sense (case of fertility selection). *Genetics* 141: 733–742.
- Lynch, M., and B. Walsh. 1998. *Genetics and analysis of quantitative traits*.. Sinauer Associates, Sunderland, MA.

- Maddox, G. D. and J. Antonovics. 1983. Experimental ecological genetics in *Plantago*: a structural equation approach to fitness components in *P. aristata* and *P. patagonica*. *Ecology* 64: 1092–1099.
- Mitchell-Olds, T. 1987. Analysis of local variation in plant size. *Ecology* 68: 82–87.
- Mitchell-Olds, T., and J. Bergelson. 1990. Statistical genetics of *Impatiens capensis*. II. Natural selection. *Genetics* 124: 417–421. [32]
- Mitchell-Olds, T. and R. G. Shaw. 1987. Regression analysis of natural selection: statistical inference and biological interpretation. *Evolution* 41: 1149–1161.
- Mousseau, T. A., and D. A. Roff. 1987. Natural selection and the heritability of fitness components. *Heredity* 59: 181–197.
- Nagylaki, T. 1976. The evolution of one- and two-locus systems. *Genetics* 83: 583–600.
- Nagylaki, T. 1977a. *Selection in one- and two-locus systems*. Springer-Verlag, Berlin.
- Nagylaki, T. 1977b. The evolution of one- and two-locus systems. II. *Genetics* 85: 347–354.
- Nagylaki, T. 1989. Rate of evolution of a character without epistasis. *Proc. Natl. Acad. Sci.* 86: 1910–1913.
- Nagylaki, T. 1991. Error bounds for the fundamental and secondary theorems of natural selection. *Proc. Natl. Acad. Sci.* 88: 2402–2406. [14, 15]
- Nagylaki, T. 1992. Rate of evolution of a quantitative character. *Proc. Natl. Acad. Sci.* 89: 8121–8124. [14]
- Nagylaki, T. 1993. The evolution of multilocus systems under weak selection. *Genetics* 134: 627–647. [14]
- O'Donald, P. 1968. Measuring the intensity of selection. *Nature* 220: 197–198.
- Pearson, K. 1903. Mathematical contributions to the theory of evolution. XI. On the influence of

natural selection on the variability and correlation of organs. *Phil. Trans. Roy. Soc. Lond. A* 200: 1–66.

Phillips, P. C. and S. J. Arnold. 1989. Visualizing multivariate selection. *Evolution* 43: 1209–1222.

Price, G. R. 1970. Selection and covariance. *Nature* 227: 520–521.

Price, G. R. 1972a. Extension of covariance selection mathematics. *Ann. Hum. Genet.* 35: 485–490.

Price, G. R. 1972b. Fisher's 'fundamental theorem' made clear. *Ann. Hum. Genet.* 36: 129–140.

Prout, T. 1965. The estimation of fitness from genotypic frequencies. *Evolution* 19: 546–551.

Prout, T. 1969. The estimation of fitness from population data. *Genetics* 63: 949–967.

Reyment, R. A. 1991. *Multidimensional palaeobiology*. Pergamon Press, Oxford.

Robertson, A. 1955. Selection in animals: synthesis. *Cold Spring Harbor Symp. Quant. Biol.* 20: 225–229.

Robertson, A. 1960. A theory of limits in artificial selection. *Proc. Royal Soc. B* 153: 234–249.

Robertson, A. 1966. A mathematical model of the culling process in dairy cattle. *Anim. Prod.* 8: 95–108.

Robertson, A. 1968. The spectrum of genetic variation. In R. C. Lewontin (ed.), *Population biology and evolution*, pp. 5–16. Syracuse University Press, Syracuse, N. Y.

Roff, D. A. 1997. *Evolutionary quantitative genetics*. Chapman and Hall, New York.

Roff, D. A., and T. A. Mousseau. 1987. Quantitative genetics and fitness: lessons from *Drosophila*. *Heredity* 58: 103–118.

Rose, M. 1982. Antagonistic pleiotropy, dominance and genetic variance. *Heredity* 48: 63–78.

Scheiner, S. M. 1989. Variable selection along a successional gradient. *Evolution* 43: 548–562.

Schluter, D. 1988. Estimating the form of natural selection on a quantitative trait. *Evolution* 42:

849–861.

Stewart, S. C. and D. J. Schoen. 1987. Patterns of phenotypic variability and fecundity selection in a natural population of *Impatiens pallida*. *Evolution* 41: 1290–1301.

Stratton, D. A. 1992. Life-cycle components of selection in *Erigeron annuus*: I. Phenotypic selection. *Evolution* 46: 92–106.

Sutherland, W. J. 1985a. Chance can produce a sex difference in variance in mating success and explain Bateman's data. *Anim. Behav.* 33: 1349–1352.

Sutherland, W. J. 1985b. Measures of sexual selection. *Oxford Surveys Evol. Bio.* 2: 90–101.

Sved, J. A. 1989. The measurement of fitness in *Drosophila*. In W. G. Hill, and T. F. C. Mackay, (eds.), *Evolution and animal breeding*, pp. 113–120. C. A. B. International, Wallingford.

Tallis, G. M. 1987. Ancestral covariance and the Bulmer effect. *Theor. Appl. Genet.* 73: 815–820. [33]

Tallis, G. M. and P. Leppard. 1988. The joint effects of selection and assortative mating on multiple polygenic characters. *Theor. Appl. Genet.* 75: 278–281.

Trail, P. W. 1985. The intensity of selection: intersexual and interspecific comparisons require consistent measures. *Amer. Nat.* 126: 434–439.

Travis, J. and S. Henrich. 1986. Some problems in estimating the intensity of selection through fertility differences in natural and experimental populations. *Evolution* 40: 786–790.

Turelli, M. and N. H. Barton. 1990. Dynamics of polygenic characters under selection. *Theor. Pop. Biol.* 38: 1–57.

Turelli, M., J. H. Gillespie, and R. Lande. 1988. Rate tests for selection on quantitative characters during macroevolution and microevolution. *Evolution* 42: 1085–1089.

Walsh, B. 2002. Quantitative genetics, genomics, and the future of plant breeding, in, M. S. Kang (ed), *Quantitative genetics, genomics, and plant breeding*, pp. 23 – 32. CAB International.

- Walsh, B. and M. Lynch. 2003. *Evolution and Selection of Quantitative Traits*. Sinauer Associates, Sunderland, MA. (in preparation). [Draft chapters can be found on the book website at http://nitro.biosci.arizona.edu/zbook/volume_2/vol2.html]
- Willis, J. H. 1996. Measures of phenotypic selection are biased by partial inbreeding. *Evolution* 50: 1501–1511.
- Yule, G. U. 1902. Mendel's laws and their probable relation to intra-racial heredity. *New Phytol.* 1: 193–207, 222–238.