

Chapter 1: Selection on a Single Trait

© Stevan J. Arnold 2016

Overview.- Phenotypic selection can be measured by its effects on trait distributions within a generation. Our fundamental approach contrasts trait distribution before and after selection. This contrast is less intuitive than the comparison of traits in survivors and nonsurvivors, but it has an important statistical advantage. The difference in trait means before and after selection is equivalent to the covariance between a trait and fitness. Such selection differentials have been measured in a wide variety of natural populations and show that trait means are usually shifted by less than a half a phenotypic standard deviation (mean about 0.6) and that the modal value is close to zero. A similar perspective on trait variances shows that they usually contract by 0-50 percent within a generation or expand by 0-25 percent as a consequence of selection, with a mode close to zero.

In this chapter we will focus on simple descriptive characterizations of selection. Our account goes only a little way beyond older treatments of selection on quantitative traits in which correlations with other traits are ignored and no attempt is made to relate selection coefficients to equations for evolutionary change (Cook 1971, Endler 1986). In later chapters we will correct for the effects of selection on correlated characters, deduce modes of selection and fitness functions from changes in trait distributions, and use our measures of selection to model evolution.

1.0 Traits and trait distributions.

Many important traits show continuous distributions within populations, rather than discrete polymorphisms. Such traits are represented by multiple values rather than a few, so that the resulting distribution is continuous, and it is often unimodal (Wright 1968). Although normal distributions are not universal, many traits approach such a distribution, or can be transformed so that they approach normality more or less closely (Wright 1968). In the following sections of this chapter, and in most of the chapters that follow we will assume normality of trait distributions. This assumption is less restrictive than it may appear. In many theoretical situations that follow the crucial assumptions are actually unimodality and symmetry rather than normality *per se*.

A few examples will illustrate the kinds of traits that are continuously distributed. The examples that follow were chosen because not only because their



statistical distributions are well known, but because they are the subjects of research from diverse points of view. Because this extensive backlog of information, we will use them as examples throughout this book.

Figure 1.0. Radiograph of a natricine snake showing vertebrae in the body and tail. The distinction between the two kinds of vertebrae is not arbitrary. Ribs are attached to body vertebrae, but not to tail vertebrae. These two vertebral numbers can be assessed, without recourse to radiography, by counting ventral and subcaudal scales. The electronic object is a radiotransmitter used to study thermoregulation in free-ranging females during pregnancy.

Vertebral numbers in snakes (Fig. 1.0) have been important characters in systematics since the time of Linnaeus because they often differentiate even closely related species, as well as higher taxa. Vertebral counts also serve as markers for the occupancy of different adaptive zones; as few as 100 in fossorial species, as many as 300 in arboreal species (Marx & Rabb 1972). In most snakes the vertebrae show a 1:1 correspondence with external scales, so counts can be made using those scales (ventral and subcaudal) without recourse to radiography (Alexander & Gans 1966, Voris 1975). Furthermore, the transition from body vertebrae (with ribs) to tail vertebrae (with ribs) is marked by the anal scale, so counts on both body regions can be made in any specimen without a broken tail. In most snakes, both counts are sexually-dimorphic, typically with more vertebrae in males. Counts from females in a single population of the garter snake *Thamnophis elegans* are shown in Fig. 1.1. Distributions of body and tail vertebrae are generally unimodal and closely approximate normal or lognormal distributions (Kerfoot & Kluge 1971), as in these examples.

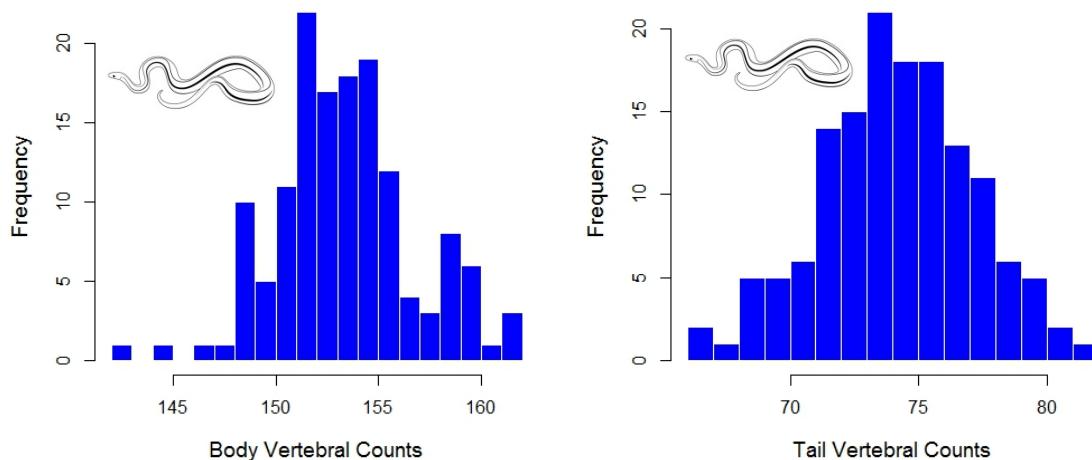


Figure 1.1 Distributions of body and tail vertebral numbers in 143 newborn garter snakes, *Thamnophis radix* (data from Arnold & Bennett 1984).

Counts of bristles on the thorax and abdomen of *Drosophila melanogaster* have been used in studies of inheritance and responses to deliberate selection since the 1940s (Mather 1941, 1942). Usually two kinds of counts are made: abdominal bristles (on the sternites located on the ventral surface of the abdomen) and sternopleural bristles (on the sternopleuron located laterally on the thorax, Fig. 1.2). The bristles are actually the moving parts of a mechanoreception system. When the bristles are moved they activate an electrical signal that is sent to the brain, keeping the fly aware of changes in its environment. Because the larger bristles (macrochaetae) on the sternopleuron are fewer in number and almost completely invariant, they are sometimes ignored and so that the count is based only on the smaller, more numerous bristles (microchatae) (Clayton et al. 1957b). Distributions of abdominal and sternopleural bristle numbers closely approach normal distributions (Fig. 1.3).

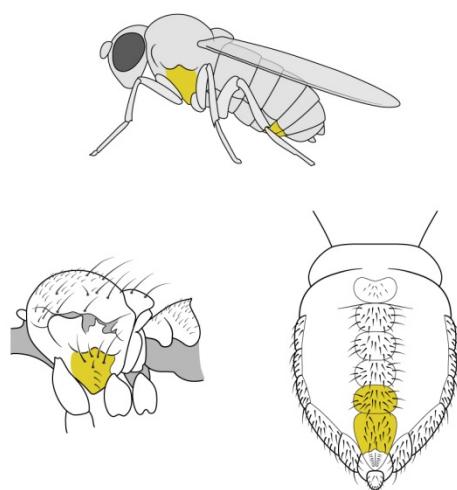


Figure 1.2. *Drosophila*, showing sites of important bristle counts. (Left). The sternopleuron (colored) on the thorax, showing eight sternopleural bristles (after Wheeler 1981). (Right). Abdominal tergites, showing abdominal bristles on the 4th and 5th tergites (colored).

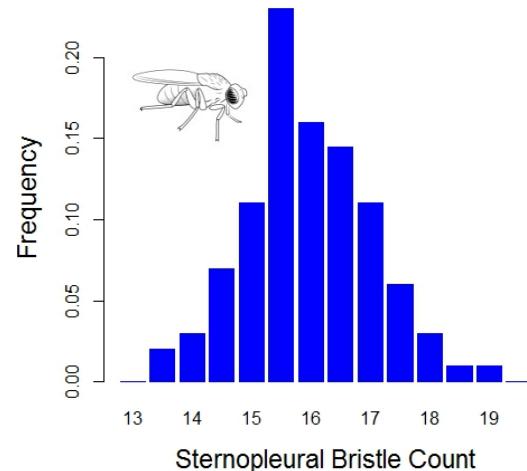
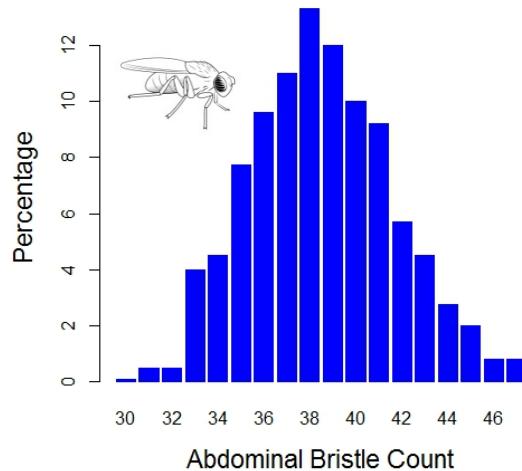


Figure 1.3. Histograms illustrating variation in *Drosophila melanogaster* bristle counts. (Left). Abdominal bristle number (Falconer & Mackay 1989). (Right). Sternopleuroal bristle number (i.e., sum of right and left sides). This histogram shows just the contributions to total variation from chromosome 2 (Mackay & Lyman 2005).

The dimensions of bird bills often reflect differences in food habits among species and so capture an essential feature of adaptive radiations (Schluter 2000). Diversification of bills is a pivotal feature of the adaptive radiation of the ground finches of the Galapagos, and for this reason many aspects of bills have been intensively studied (Lack 1947, Bowman 1961, Abbott et al. 1977, Grant 1986). The measurements are made on individuals that have reached adult size so that there are no ontogenetic complications. Distributions of bill depth, for example, often approximate a normal distribution (Fig. 1.4).

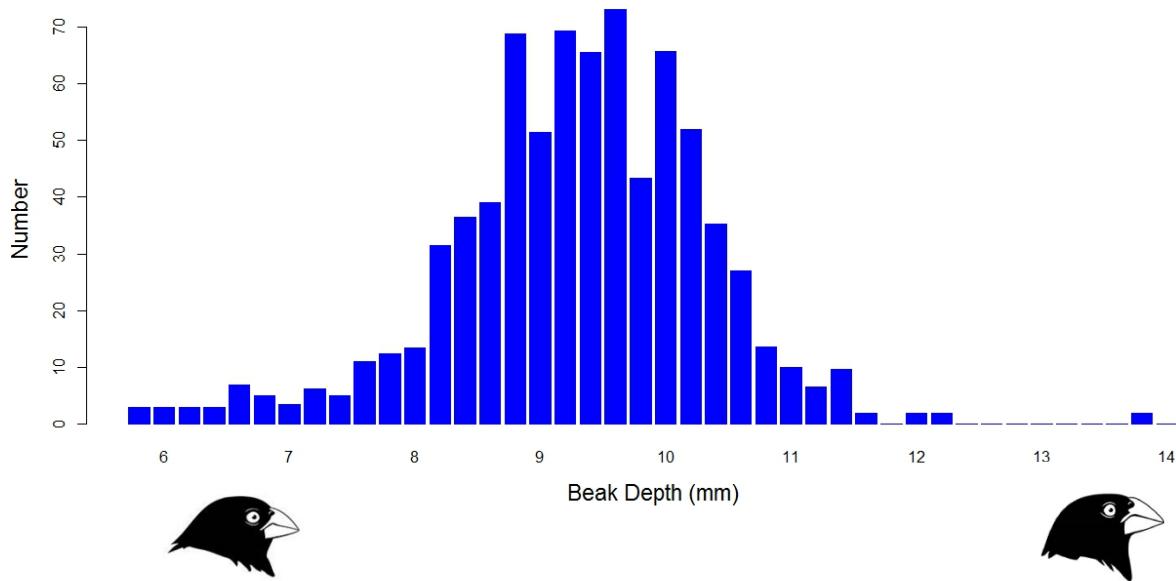


Figure 1.4. The distribution of beak depth of the Medium Ground Finch (*Geospiza fortis*) on Daphne Major, Galapagos Islands, in 1976 before a drought (n=751, from Grant 1986).

In this and the chapters that follow, we will assume that the trait does not change during ontogeny, as a result of age, growth or experience. Some traits are naturally of this kind. Vertebral numbers in snakes and other vertebrates, for example, are determined relatively early in development and do not change during the postnatal ontogeny. Likewise, bristle numbers do not change once the fly ecloses from its pupal stage. In other cases ontogeny-invariance can be achieved by defining age-specific traits (e.g., size at age three years), as in the beak dimensions of *Geospiza*. A general solution to the issue of traits that vary with age, size, experience, environment, etc. can be achieved by treating them as function-valued or infinite dimensional attributes (Kirkpatrick 1989, Gomulkiewicz & Kirkpatrick 1992, Kingsolver et al. 2001). In this approach, the phenotypic size of an individual is represented as a continuous function of age. The resulting theory closely follows the more simple theory for point-valued traits that is sketched here and in later chapters. In general, the main expressions remain the same except those involving traits values are transformed to continuous functions. In any case, the general point in trying to achieve size- and age-independence is that we want to define a phenotype that enables us to separate the effects of ontogeny from the effects of selection.

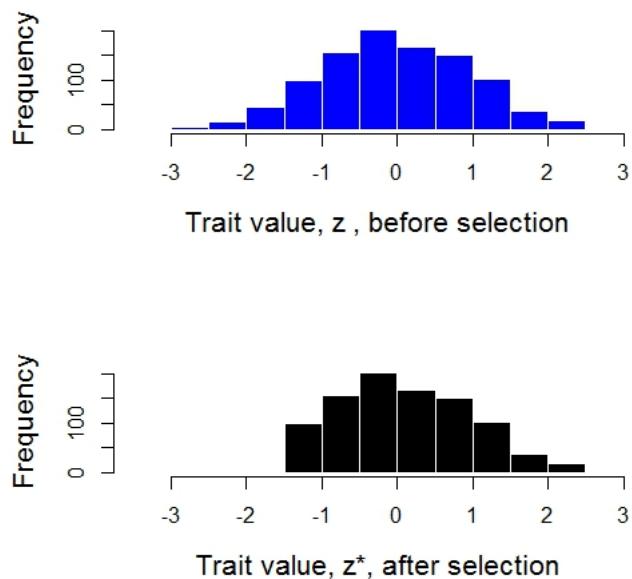
The choice of scale for a particular trait can be based on practical concerns. Homogeneity of variance among populations or higher taxa is often desirable for then the evolution of the trait mean can be divorced from concerns about the evolution of trait variance. The logarithmic scale is often useful in attaining this kind of invariance and has useful properties in its own right (Wright 1969). On the other hand, transforming a trait with the sole goal of making its distribution approach normality is seldom useful. Most statistical tests assume that the distribution of errors is normal (not the trait distribution itself) and, in any case, are robust to even appreciable departures from normality.

1.1 Selection changes the trait distribution.

We are concerned here not with the agents of selection but with the statistical effects of those agents on a particular trait, z . Those statistical effects are evolutionarily important even though they fail to capture the personality of selection. Imagine the statistical distribution of the trait in a population before selection has acted (Fig. 1.5). We will call the continuous version of that distribution $p(z)$, a distribution function that might take any of a variety of forms. Later we will assume that the function is a normal

distribution, but for the moment we will not make any assumptions about its form. Now imagine that as a consequence of selection some phenotypes increase in frequency, while others decrease (Fig. 1.5).

Figure 1.5. A single hypothetical trait in a sample of 1000 individuals is subjected to truncation selection. (above) The histogram of trait values for these individuals before selection is shown in blue (mean = -0.02, variance = 0.96). (below) Only individuals with trait values greater than -1.5 (n=921) survived selection. The trait distribution after selection is shown in black. Selection has shifted the trait mean and contracted its variance (mean = 0.11, variance = 0.75). $s = \bar{z}^* - \bar{z} = 0.13$; $(P^* - P)/P = -0.22$; $(P^* - P + s^2)/P = -0.21$.



We ascribe those changes in frequency to differences in fitness as a function of phenotype. The essence of selection is that all individuals with a particular phenotype, z , have an expected absolute fitness, which we will call $W(z)$. To determine average absolute fitness in the population we need to weight each value of fitness by its frequency, in other words,

$$\bar{W} = \int p(z)W(z)dz. \quad (1.00)$$

The differences in fitness are crucial in determining how the frequency of individuals with phenotype z will be changed from $p(z)$ before selection to $p^*(z) = w(z)p(z)$ after selection, where $w(z) = W(z)/\bar{W}$ is relative fitness of an individual with phenotype z (Fig. 1.5). Note that because mean relative fitness is

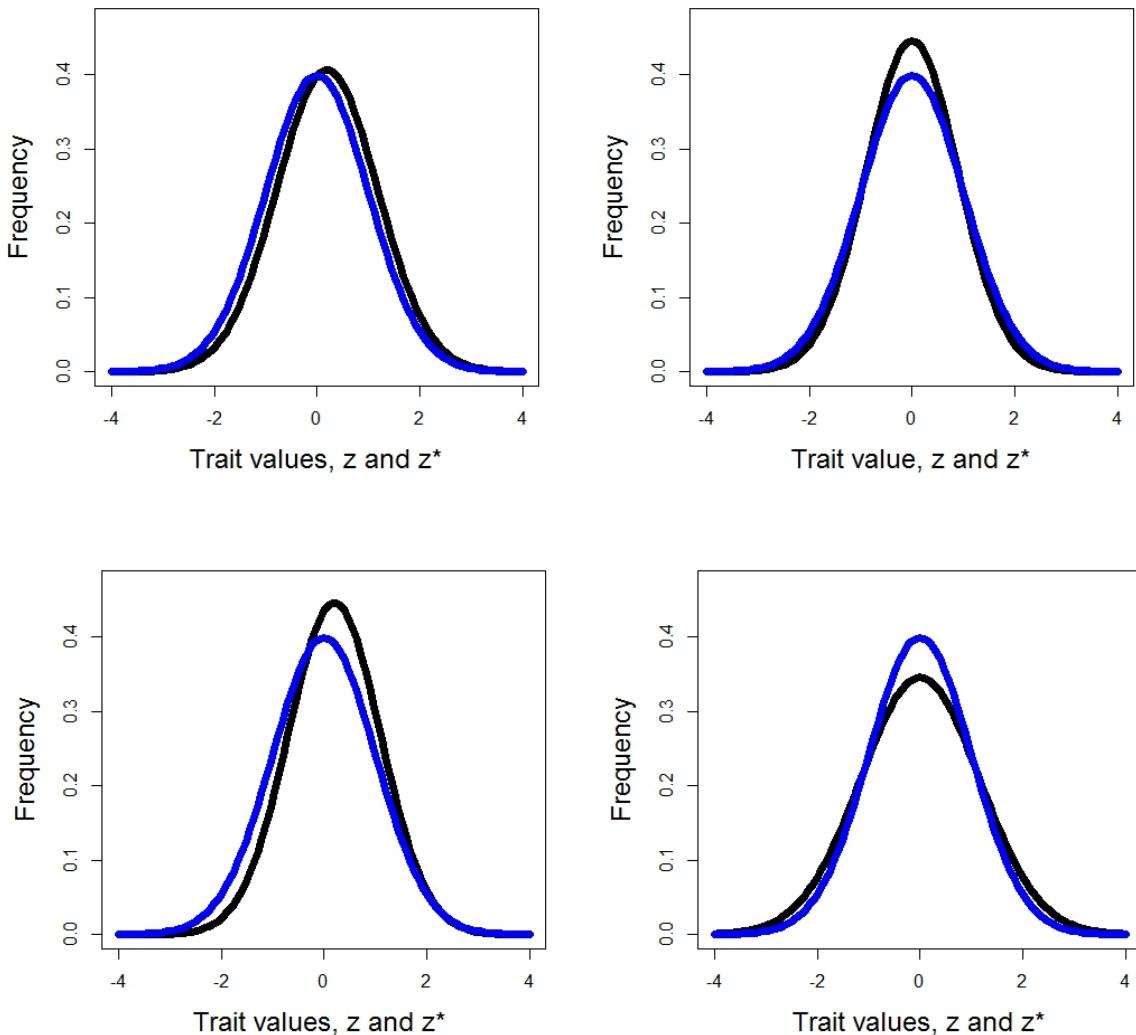


Figure 1.6 Hypothetical examples of selection acting on normally-distributed trait distributions. Trait distributions are shown in blue before selection and in black after selection. (upper left) An upward shift in mean with little change in variance; $\bar{z} = 0.00$, $P = 1.00$, $\bar{z}^* = 0.02$, $P^* = 0.96$ (upper right) a contraction in variance with no shift in mean; $\bar{z} = 0.00$, $P = 1.00$, $\bar{z}^* = 0.00$, $P^* = 0.80$. (lower left) An upward shift in mean with a contraction in variance; $\bar{z} = 0.00$, $P = 1.00$, $\bar{z}^* = 0.20$, $P^* = 0.80$. (lower right) An expansion of variance with no shift in mean; $\bar{z} = 0.00$, $P = 1.00$, $\bar{z}^* = 0.00$, $P^* = 1.33$.

$$w = \int p(z)w(z)dz = \int p(z)\frac{W(z)}{\bar{W}}dz = \int p(z)^*dz, \quad (1.01)$$

it equals 1. We will need the crucial function, $p(z)^*$, the frequency distribution after selection, to calculate various coefficients that can be used to characterize selection (Lande 1976).

To simplify particular theoretical results it will sometimes be useful to assume that the trait distribution before selection is normal. Under this assumption we have the following expression for $p(z)$,

$$p(z) = \frac{1}{\sqrt{2\pi P}} \exp\left\{-\frac{(z - \bar{z})^2}{2P}\right\}. \quad (1.02)$$

The $1/\sqrt{2\pi P}$ term is a normalization factor which insures that the trait probabilities sum to one.

1.2 Shift in the trait mean, the linear selection differential.

A fundamental question is to ask what does selection do to the mean of our trait distribution. The mean before selection is, using the standard definition of the mean,

$$\bar{z} = \int p(z)zdz. \quad (1.03)$$

Using that same, familiar definition, the mean after selection must be

$$\bar{z}^* = \int p(z)^*zdz. \quad (1.04)$$

A natural way to express the effect of selection on the mean is to take the difference between the mean after selection and mean before selection. The difference is taken in this order so that it will be positive, when the mean is shifted upwards. This difference is called the *directional selection differential*,

$$s = \bar{z}^* - \bar{z}. \quad (1.05)$$

It is useful to measure the shift in mean given by s in units of within-population phenotypic standard deviation. If we let the variance in the population for trait z be P before selection, then the trait standard deviation is $P^{1/2}$, and our *standardized directional selection differential* is $s/P^{1/2}$. A standardized selection differential of 2 means that the mean has been shifted upward by two phenotypic standard deviations (Lande & Arnold 1983).

1.3 The directional selection differential as a covariance.

The directional selection differential is an especially powerful descriptor of selection because it is a covariance as well as a shift in mean. Recall that the covariance between two variables, call them x and y , is defined as

$$\text{Cov}(x, y) = \int p(x, y)(x - \bar{x})(y - \bar{y})dxy , \quad (1.06)$$

where $p(x,y)$ is the frequency of observations with values x and y . With a little rearrangement we can express this same equation as

$$\text{Cov}(x, y) = \int p(x, y)xydxy - \bar{xy} . \quad (1.07)$$

Substituting $w(z)$ for x and z for y into these two expressions, and remembering that the mean of $w(z)$ is 1, we obtain

$$\text{Cov}(w, z) = \int p(z)[w(z)-1][z - \bar{z}]dz = \int p(z)* zdz - \bar{z} = \bar{z}^* - \bar{z} = s . \quad (1.08)$$

In other words, the directional selection differential, s , is the covariance between relative fitness and trait values (Robertson 1962). This second definition of the directional selection differential will be of special importance in later sections. An example, showing the equivalence of $\text{Cov}(x,z)$ and the directional

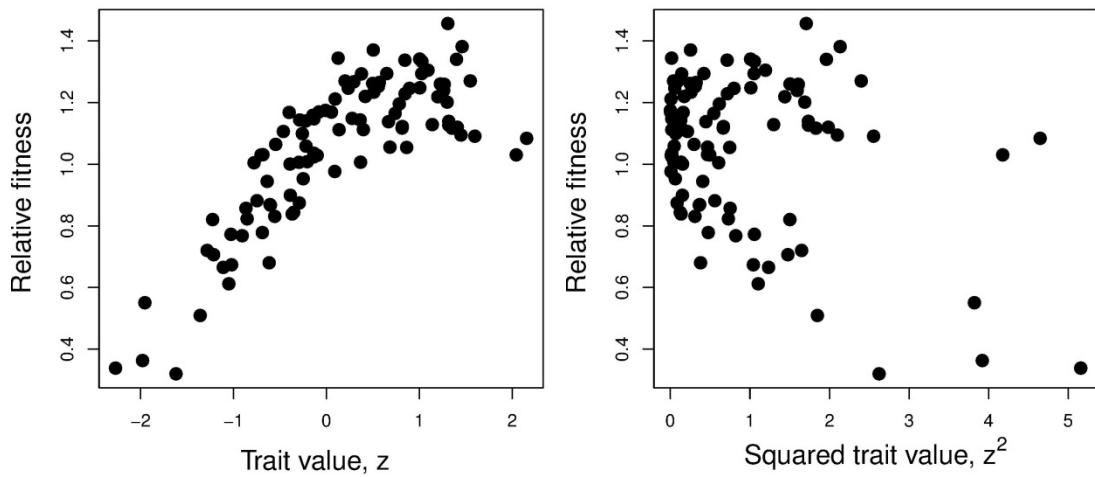


Figure 1.7. Scatterplots plots showing the equivalence between covariance and selection differentials. The sample of 100 individuals in these plots were drawn from the trait and fitness distributions, $p(z)$ and $w(z)$, used in Fig. 1.1c. (a) Relative fitness as a function of trait value, z ($\bar{z} = -0.13$, $P = 0.98$, $w = 0.97$, $\text{cov}(z,w)=0.21$, $r(z,w)=0.81$). (b) Relative fitness as a function of squared trait values, z^2 (mean $z^2 = 0.99$, $\text{var}(z^2)= 0.98$, $\text{cov}(z^2,w)=-0.17$, $r(z^2,w)=-0.46$). *pull in correct values from program file*

selection differential is provided in Fig. 1.7a, which shows relative fitness as a function of trait values. This sample was drawn from the trait and fitness distributions, $p(z)$ and

$w(z)$, used in Fig. 1.5c. As expected the covariance in Fig. 1.7a (0.21) is very close to the shift in mean in Fig. 1.1c (0.20)

1.4 Change in the trait variance, the nonlinear selection differential.

We can expect that selection might change the variance of a trait, just as it might shift the trait mean. Applying the same logic as before, the variance before selection is, using the standard expression for variance,

$$P = \int p(z)(z - \bar{z})^2 dz \quad (1.09)$$

and after selection it is

$$P^* = \int p^*(z)(z - \bar{z})^2 dz. \quad (1.10)$$

By analogy with our treatment of the mean, we might measure the absolute effect of selection on the variance as P^*-P and its proportional effect as $(P^*-P)/P$. For example, when this proportional measure is -0.5, the trait variance has been reduced by 50%.

A slightly more complicated measure of effects on variance will prove useful because it leads to a equivalence with a covariance. The essential point behind this more complicated measure is that the same mode of selection that shifts the trait mean will also contract its variance. This effect of directional selection on variance is especially easy to characterize if the trait is normally distributed before selection. In that case, directional selection that shifts the mean by an amount s will contract the trait's variance by an amount s^2 . Consequently, we can define a *nonlinear selection differential* so that it measures effects on variance from sources other than directional selection (e.g., from stabilizing and disruptive selection), viz.

$$C = P^* - P + s^2. \quad (1.11)$$

This same selection differential, not P^*-P , is equivalent to the covariance between relative fitness and squared deviations from the trait mean,

$$C = Cov[w, \tilde{z}^2], \quad (1.12)$$

where $\tilde{z} = z - \bar{z}$ (Lande & Arnold 1983). An example, showing the equivalence of $Cov(x,z)$ and the nonlinear selection differential is provided in Fig. 1.7b, which shows relative fitness as a function of squared trait values. As expected the covariance in Fig. 1.7b (-0.17) is very close to the corrected change in variance, $C = P^* - P + s^2$, observed in the parent distributions shown in Fig. 1.6c (-0.16), within the bounds of sampling error. As before, it is useful to standardize using the variance before selection to obtain a proportional measure of effects on variance, a *standardized nonlinear selection differential*, C/P . In the next section we will show that s and C reflect selection acting on correlated traits, as well as selection acting directly on the trait in question.

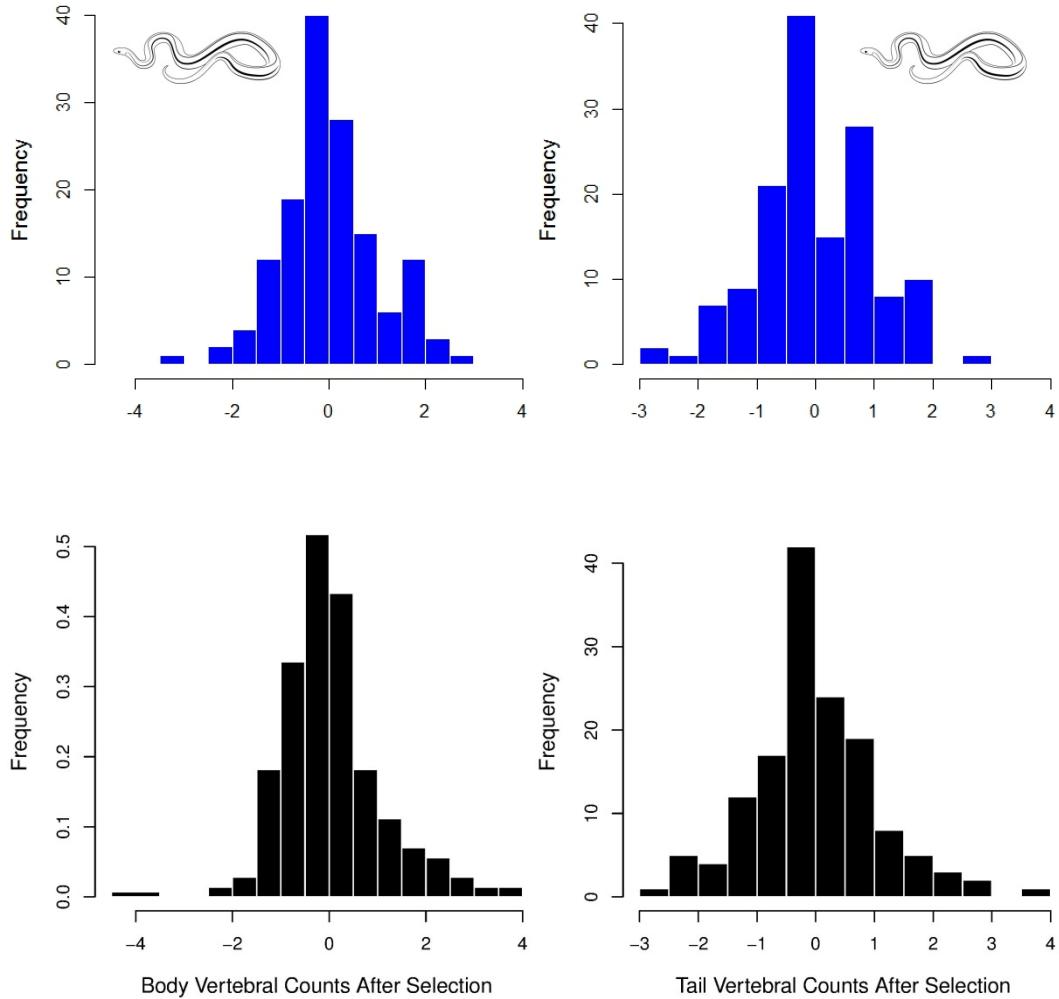


Figure 1.75. Samples of body and tail vertebral counts in the garter snake *T. radix* ($n=143$) before and after selection. The samples before selection are standardized to zero means. In the samples after selection, the vertebral count of each individual is weighted by relative crawling speed.

1.5 Estimates of univariate selection differentials

Statistics associated with traits distributions before and after selection allow us to visualize the impact of selection. The following example illustrates how various

univariate statistics contribute to our visualization of selection. In this example, crawling speed of newborn garter snakes (*Thamnophis radix*) was measured in the laboratory and related to counts of body and tail vertebrae (Arnold & Bennett 1988). Crawling speed is plausibly related to vertebral counts because larger vertebral counts promote the flexibility needed for snake locomotion (Fig. zz.x)

Figure zz.x Diagram of a crawling snake showing how the body pushes against points in the environment. Contractions of musculature in particular segments of the body (black) produce the pressure that moves the snake forward (from Jayne 1985 = PhD dissertation).

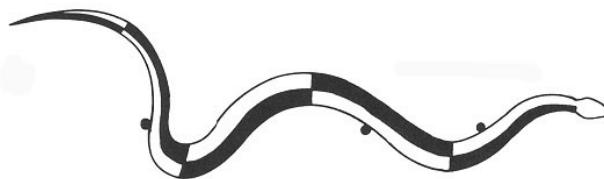


Table 1.1 Selection differentials describing effects of numbers of body and tail vertebrate on crawling speed in newborn garter snakes (*Thamnophis radix*) (n=143). Standardized measures are denoted with a '. Bootstrap estimates of 95% confidence limits are shown in parentheses.

	BODY		TAIL	
	mean	variance	mean	variance
Before selection	153.414	11.437	74.257	9.094
Before selection'	0.000	1.000	0.000	1.000
After selection'	0.031	1.122	0.004	1.142
		$P_1^* - P_1$		$P_2^* - P_2$
Change in variance		0.122 (-0.062, 0.342)		0.142 (-0.015, 0.327)
	s_1	C_1	s_2	C_2
Selection	0.031	0.123	0.004	0.142

As is typical in snakes, males had several more tail vertebrae than females and a few more body vertebrae, so the sex

differential	(-0.032, 0.096)	(-0.062, 0.328)	(- 0.058, 0.063)	(-0.015, 0.327)
Correction term, s^2		0.077		0.001

difference in means was added to the counts of females, and the combined sample of 143 neonates was analyzed. The trait sample before selection was standardized by subtracting the trait mean from each observation and dividing by the trait standard deviation, so that the means of each trait were zero and variances were equal to one. This standardization simplifies the interpretation of selection statistics. The sample after selection, $p(z)^*$, in this example is not a sample of survivors, but is instead the trait distribution weighted by relative crawling speed (Fig. 1.75). Because of the trait standardization, we can immediately see the changes in trait means and variances before we calculate the selection differentials (Table 1.1). We see that the body mean has been shifted upward by slight amount (3.1% of standard deviation in the body vertebral count before selection). The tail mean has also been shifted upward but to a smaller degree (0.4% of standard deviation in the body vertebral count before selection). The 95% confidence intervals for these and other statistics were estimated by bootstrapping over the sample of 143 individuals and are reported in Table 1.1. We see that the confidence intervals for s_1 and s_2 do not overlap zero, so those estimates are not significant at the 0.05 level. Turning to the trait variance, we crawling performance has caused the body variance to expand by 12.2% ($P_1^*-P_1$). The effect of directional selection on the trait variance will be to decrease the variance by 7.7% ($s_2=0.077$), so when we add that amount to $P_1^*-P_1$, we obtain a selection differential of 12.3%, which represents the effect of disruptive selection, after correcting for directional selection. The effect on tail variance is similar, except the expansion is slightly larger and the effect of directional selection is smaller. Bootstrapped confidence intervals indicate that neither the uncorrected changes in variance nor the nonlinear selection differentials are different from zero. One could easily conclude the slight changes in means and variances - and from our analysis - that the trait distributions have not been changed by selection. See shall see in the next chapter that this conclusion is premature, because we have not considered effects on trait covariance!

Ground finches (*Geospiza fortis*) present an especially well understood example of selection. Intensive mark-recapture studies on Dapne Major, a small island in the Galapagos, provided measurements of beak depth on 751 adult birds in 1976. Shortly after this field work, a severe drought markedly changed the availability of seed types on the island. When a sample was taken after the drought in 1978, only 90 birds survived. An inspection of the beak depth distribution before and after selection (Fig. 1.8) reveals that the average beak depth was shifted upwards and variance contracted.

Indeed, the mean beak depth distribution shifted upward by 60% of a phenotypic standard deviation ($s = 0.60$). Both tails of the distribution were trimmed by the drought condition, but especially the lower tail. Observation studies showed that birds with deep beaks were better able to open larger seeds that increased in frequency during the drought (Grant 1986).

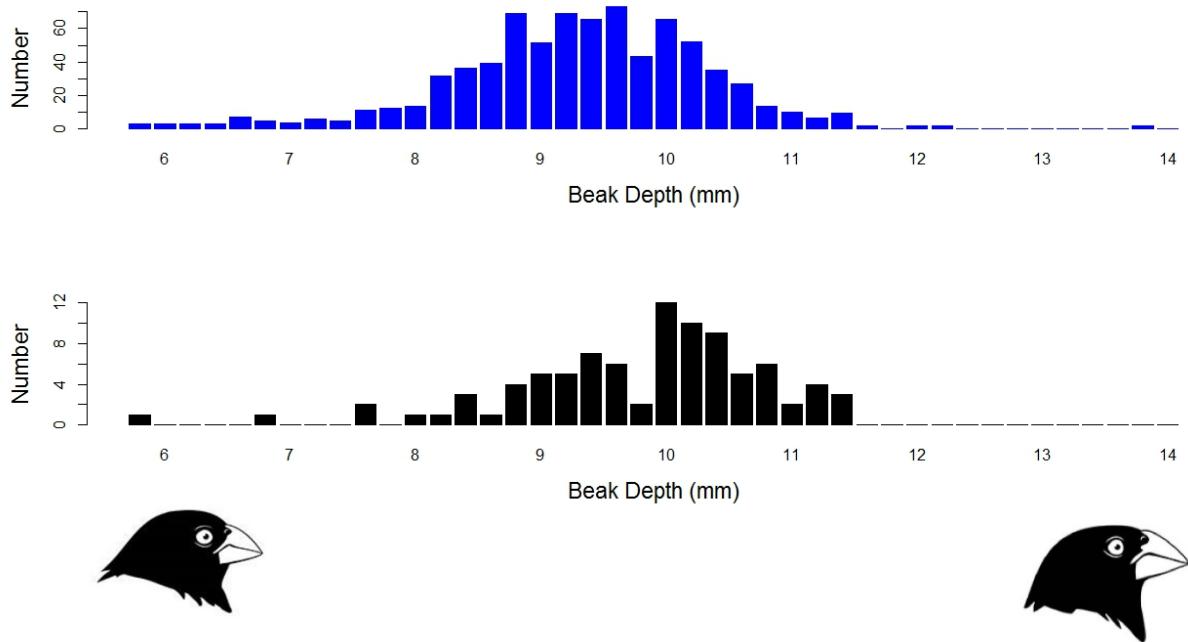


Figure 1.8 Distributions of beak depth measurements before and after selection on the Daphne Major population of *Geospiza fortis*. (above) The distribution in 1976, before selection ($n=751$). (below) The distribution in 1978, after a drought killed many birds ($n=90$). (From Boag & Grant 1984, Grant 1986).

1.6 Technical issues in estimating and interpreting selection differentials

Different kinds of data are used to infer selection and measure its impact. It is useful to recognize two broad categories of samples. In *longitudinal samples*, a set of individuals is followed through time. Phenotypes are measured before selection, as well as after selection, and a particular value for fitness can be assigned to each individual. Because fitness values are attached to individuals, the covariance forms for selection differentials (1.07, 1.11) can be used. The significance of fitness assignments is apparent when we consider the contrasting case of *cross-sectional samples*. In a cross-sectional study, one sample is taken before selection and another is taken after selection, but individuals are

not tracked through time. As a consequence, more assumptions must be made to interpret selection differentials. In particular, one must assume that the sample before selection is representative of the statistical population that gave rise to the sample after selection. Although this assumption is straight-forward in some cases, it can be tortuous if the study begins with a sample of survivors and the probable population before selection must be reconstructed (Blanckenhorn et al. (1999)). The difference in samples also affects the estimation of standard errors. In the case of longitudinal data, estimation is straight-forward. The two key selection differentials are covariances which can be converted to correlations with well-characterized sampling properties, assuming a normal distribution of errors or by using nonparametric correlations. In the case of cross-sectional data, however, one must use the difference formulas (1.04, 1.10) to estimate selection differentials, and standard errors must be estimated by a re-sampling procedure (e.g., boot-strapping or jack-knifing).

Studies of selection are nearly always based on particular periods or episodes rather than lifetimes of exposure. Because this restriction is universally recognized by investigators, it may not always be acknowledged in print. For example, studies of sexual selection often use mating success as a fitness currency. The selection that is measured is distinct but it is usually not summed up over a lifetime of episodes. Instead, a snapshot of selection is taken at a particular place and time (e.g., one mating season), ignoring differences in age and the possibility of age-specific differences in selection. A similar restricted focus is often taken in studies of viability selection. Such restrictions are so common that they become a common denominator in comparisons across studies of a particular kind. The restriction to selection snapshots will make a difference when we consider responses to selection across generations (Chapters 9-10), for then the focus will necessarily be on lifetime measures of fitness and selection.

The use of standing, natural variation to assess fitness differences is powerful when it succeeds, but the approach can fail if variation is limited. Measuring selection on floral morphology has been challenging for precisely this reason (Fenster et al. 2004). Despite abundant evidence that pollinators shape the morphologies of the flowers they visit, selection on specific floral traits has often proved difficult to detect.

Throughout this chapter we have been concerned with viewing selection from the standpoint of a single trait. The univariate measures of selection that we have considered (s , P^*-P , and C) are all useful, but they share a common ambiguity. Each of these indices reflects the effects of selection on correlated traits as well as on the trait in question. In the next chapter we will consider techniques for separating these two kinds of effects.

1.6 Surveys of selection differentials

Endler (1986) surveyed about 30 studies of about 24 species published between 1904 and 1985 that measured selection in natural or experimental populations exposed to selection in nature. Those studies encompassed a wide range of organisms (plants, invertebrates, vertebrates) and traits (mostly linear measurements but some counts). Endler's survey indicates that selection typically changes trait means and variances by rather small amounts. The modal values for standardized change in mean, $(\bar{z}^* - \bar{z})/\sqrt{P}$, and standardized change in variance, $(P^* - P)/P$, are very close to zero (Figs. 1.9a, b). Note that in Fig. 1.9a, the positive and negative changes in the mean are grouped together, so that $|(\bar{z}^* - \bar{z})/\sqrt{P}|$ is shown, since we are interested in the overall picture of how strong selection might be. In general, selection shifts the trait mean by less than half a phenotypic standard deviation (Fig. 1.9a). Likewise, selection generally causes a less than 50% change in trait variance (Fig. 1.9b). Contraction of variance is more common than expansion of variance; 68% of the values in Fig. 1.9b are negative. The distribution of changes in trait mean and variance are portrayed in Figs. 1.9 c-d, where for purposes of illustration the traits are assumed to be normally distributed before and after selection. Notice that the trait mean can be shifted by more than a standard deviation and the variance can change by more than 100%, but instances of such dramatic changes are relatively rare.

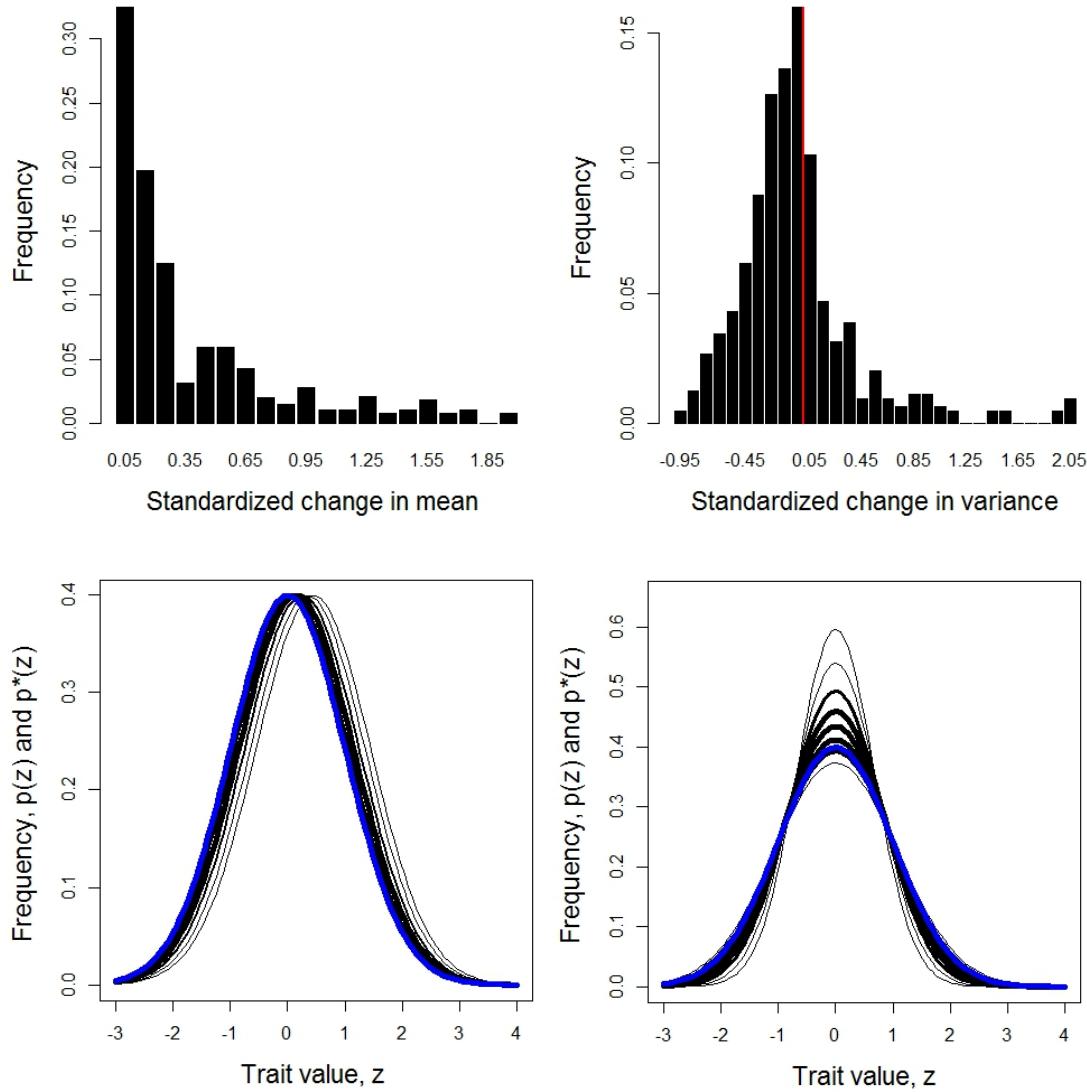
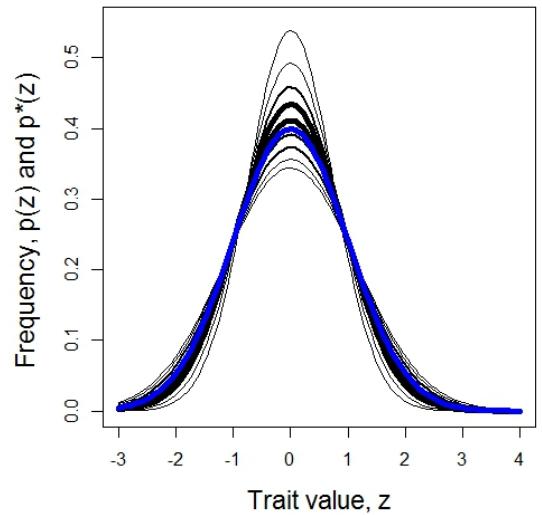
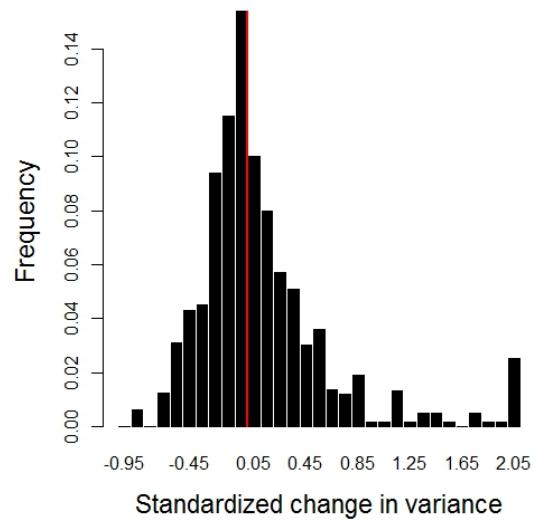


Figure 1.9. Distributions of standardized changes in means and variances. (Upper left) Distribution of estimates of the standardized change in mean, $(\bar{z}^* - \bar{z}) / \sqrt{P}$. From Endler (1986), $n=262$. (Upper right) Distribution of estimates of the standardized change in variance, $(P^* - P)/P$. From Endler (1986), $n=330$. The red vertical line shows the transition from negative to positive change in variance. (Lower left) Selection differential, s , illustrated as normal distributions after selection with mean shifted towards higher values. Line widths represent frequency in Endler's histogram; before selection (blue, with unit standard deviation and zero mean) and after selection (black). (Lower right) Standardized change in variance, $(P^* - P)/P$ illustrated as normal distributions after selection with variance contracted or expanded. Line widths represent bin frequencies in Endler's histogram.

The consequences of correcting the change in variance for the effect of directional selection is shown in Fig. 1.10. The overall effect is, as Endler (1986) noted, extremely slight. Since the modal value of directional selection is close to zero, it is not surprising that correcting for this generally weak selection usually makes a small contribution to change in trait variance. The correction does however have the effect of making corrected expansions of variance nearly as common as contractions; only 52% of the values are negative in Fig. 1.10.

Figure 1.10. Distribution of estimates of the standardized change in variance, corrected for the effects of directional selection, showing that the predominant effect of selection is to contract variance. The red vertical line shows the transition from negative to positive change in variance. (Above) Distribution of estimates of the standardized change in variance, $(P^*-P+s^2)/P$. From Endler (1986), $n=330$. (Below) Standardized change in variance, $(P^*-P+s^2)/P$, illustrated as normal distributions after selection with variance contracted or expanded. Line widths represent bin frequencies in Endler's histogram, before (blue, with unit standard deviation and zero mean) and after selection (black).



Chapter 2: Selection on Multiple Traits

© Stevan J. Arnold 2016

Overview.- The phenotypic effect of selection on multiple traits can be assessed by its effects on multivariate trait distributions. As in the case of a single trait, the fundamental approach is to compare the first and second moments of traits distributions before and after selection. Such a multivariate comparison of moments represents a major statistical improvement over trait-by-trait comparisons. By taking a multivariate approach we may be able to identify which traits are the actual targets of selection. Analysis of selection in natural systems reveals that the effects of selection on actual targets are often obscured by correlations between traits. In addition to identifying targets of selection, multivariate analysis also gives us a way to quantify functional interactions between traits. Measuring the strength of trait interactions is especially important because most traits belong to one or more functional complexes.

Animal and plant breeders may select on a single trait with the goal of improving their stocks. In the natural world, however, selection inevitably acts simultaneously on many traits. In this section we will introduce matrix algebra tools that will enable us to deal with this multivariate aspect of selection. In particular, we will move beyond the ambiguity of selection differentials. s and C are ambiguous because the shifts that they quantify could represent effects of selection acting on correlated traits as well as on the trait in question. Matrix algebra will help us to disentangle those direct and indirect effects and it will help us measure how strongly traits interact in functional complexes. The theoretical results that follow are from Lande & Arnold (1983), unless noted otherwise.

2.1 Selection changes the multivariate trait distribution.

Before we consider selection, we need to imagine the distribution of multiple traits before selection has happened. To visualize this distribution, picture a cloud of trait values in three-dimensional space. If more than three traits are involved, so that the cloud hangs in n-dimensional space, a standard convention is to depict those dimension two or three at a time. Some examples of actual two-dimensional trait distributions are shown in Fig. 2.0 {= examples of bivariate trait distributions, as in Fig. 1.0 drawing from show-cased examples, such as g-snake vertebral numbers ...}.

To consider how selection might affect such a bivariate distribution, it will be useful to consider the case of a hypothetical trait distribution that is subjected to truncation section. We will assume that the trait distribution is multivariate normal before selection, even though some of the results that follow do not depend on this

assumption. An example is provided in Fig. 2.1a, which shows a sample from normal distribution of just two hypothetical traits, z_1 and z_2 . We impose truncation selection, so that only individuals with $z_1 > -1$ and $z_2 > -1$ survive. The sample after selection is shown in Figure 2.1b. Calculation of selection measures (discussed below) confirm the impression that selection has shifted the bivariate mean and reduced dispersion in the sample.

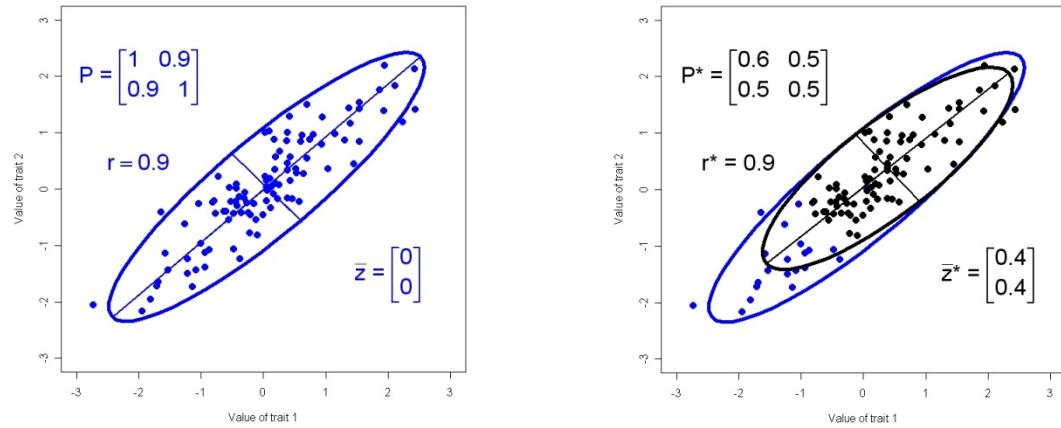


Figure 2.1. A two traits in a sample of 100 individuals are subjected to truncation selection. (a) A scatter plot trait values for these individuals before selection is shown in blue, along with the sample's 95% confidence ellipse. (b) Only individuals with trait values greater than -1.0 for both traits survived selection ($n=81$). The scatter plot of the sample after selection is shown in black along with its confidence ellipse. Selection has shifted the trait means upwards and contracted both variances, the covariance, and the correlation.

Multivariate selection can change the trait distribution in a variety of ways that we might not have expected from a simple univariate view (Fig. 1.1). In particular, bivariate selection can change trait covariances and correlations, as well as means and variances. Only contractions of variance and covariance are shown in Fig. 1.1, but expansions can occur as well. As we shall see in later section, the hypothetical selection regimes used to produce Fig. 2.2 were, for the purposes of illustration, stronger than we would expect in nature.

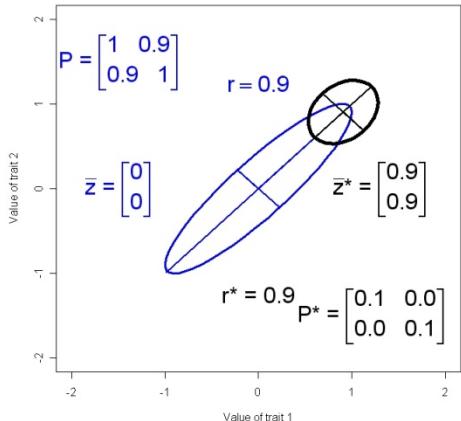
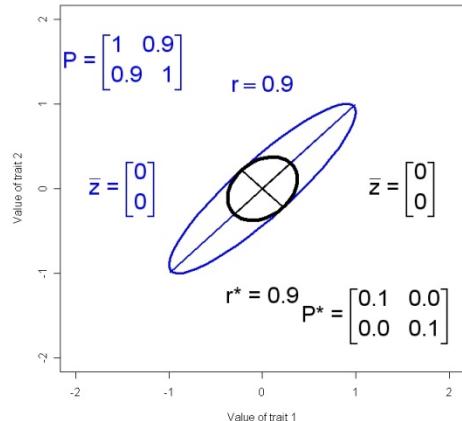
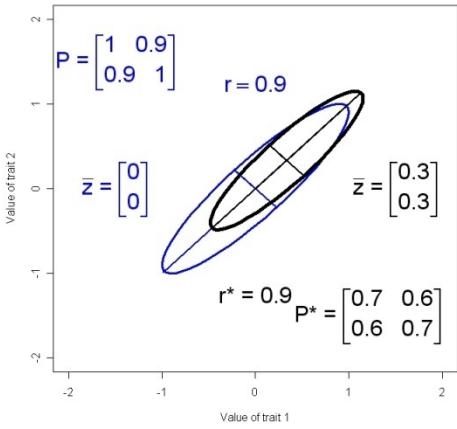
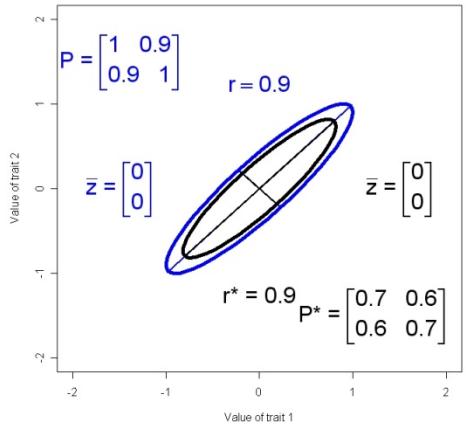
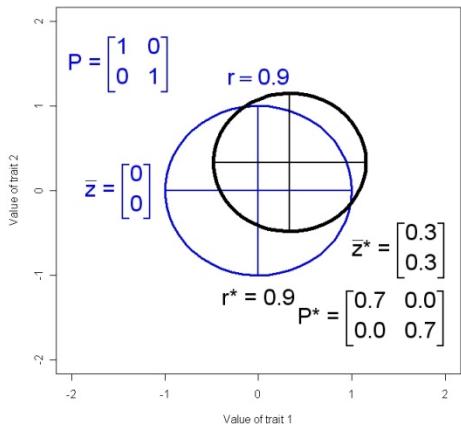
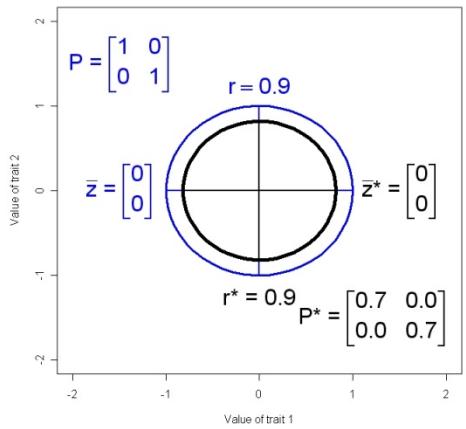
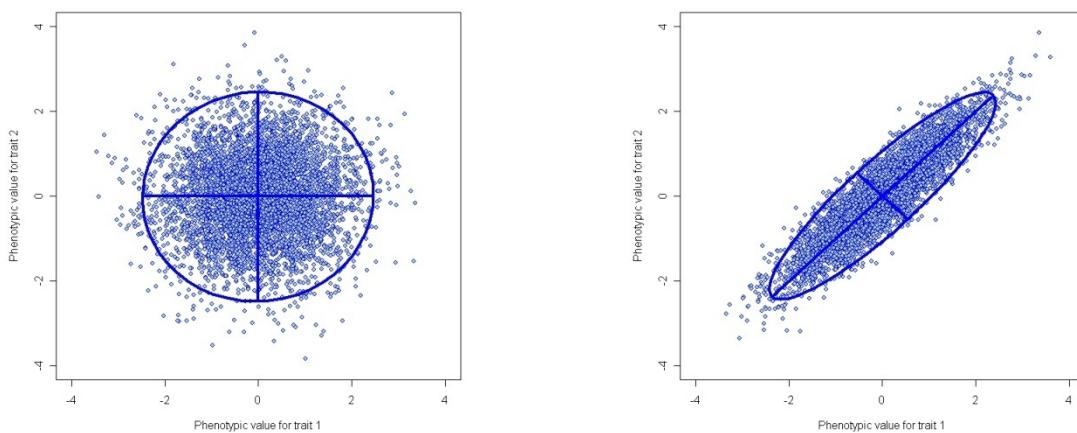


Figure 2.2. Changes in hypothetical bivariate trait distributions induced by selection. All of the distributions are normal before and after selection; 95% confidence ellipses are shown before selection (blue) and after selection (black). The position of bivariate means is shown with crosses. (a) Contractions in both variances with no shift in mean. (b) Contraction in both variances with an increase in bivariate mean: $\bar{z} = (0, 0)$. (c) Contractions in both variances and covariance with no shift in mean. (d) Contractions in both variances and covariance with an upward shift in mean. (e) Substantial contractions in both variances and covariance with no shift in mean. (f) Substantial contractions in both variances and covariance with an upward shift in mean.

To get a better feel for the bivariate normal distribution consider the views in Fig. 2.3, which show large samples from distributions with no correlation (Fig. 2.3 a and c) and a strong positive distribution (Fig. 2.3b and d). With no correlation and equal variances for the two traits, the distribution is a symmetrical hill. Positive correlation converts the distribution into a symmetric, ridge-shaped hill.



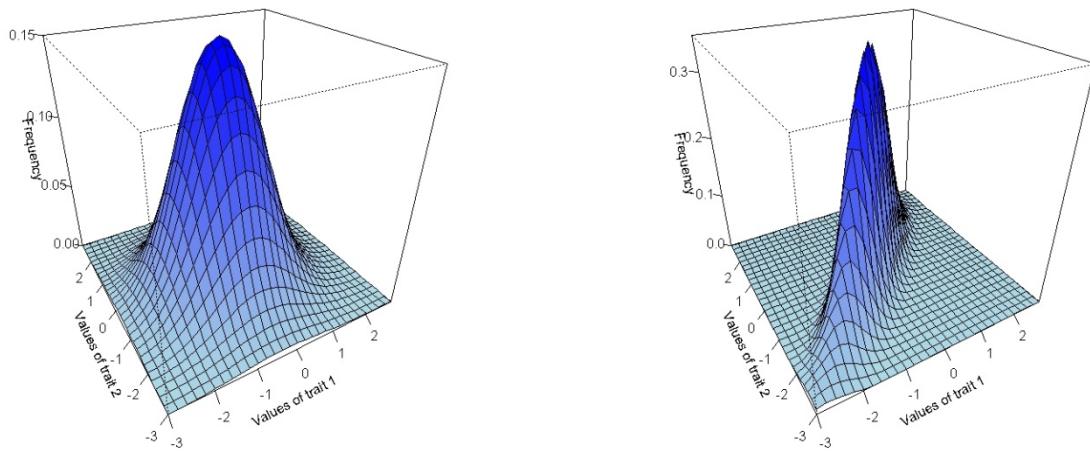


Figure 2.3. Large samples from two bivariate normal distributions of two traits, one with no trait correlation and the other with a strong positive correlation. (a) A bivariate scatter plot of the individuals sampled from a distribution with no correlation ($\bar{z}_1 = 0.0$, $\bar{z}_2 = 0.0$, $P_{11} = 1.0$, $P_{22} = 1.0$, $P_{12} = 0.0$, $r=0.0$). (b) A bivariate scatter plot of the individuals sampled from a distribution with a strong positive correlation ($\bar{z}_1 = 0.0$, $\bar{z}_2 = 0.0$, $P_{11} = 1.0$, $P_{22} = 1.0$, $P_{12} = 0.9$, $r=0.9$) (c) A 3-dimensional view of the probability distribution with no

correlation . (d) A 3-dimensional view of the probability distribution with a strong positive correlation

The points on the 3-dimensional surfaces shown in Fig. 2.3 are $p(z)$, the probabilities of observing a particular phenotypes as a function of particular traits values z_1 and z_2 . We now want to consider the formula for those probabilities for the case of a bivariate normal distribution. For convenience we can represent those two values as a column vector (Appendix 1 = basic conventions and rules of matrix algebra, see Arnold 1994), which we will call z . In the present case, the phenotype is represented by just two values, but in general it might be represented by n values, so that z might be a very tall vector. We can write an expression for a normal probability distribution in this general, n -trait case,

$$p(z) = \sqrt{(2\pi)^{-n} |P^{-1}|} \exp\left\{-\frac{1}{2}(z - \bar{z})^T P^{-1}(z - \bar{z})\right\}, \quad (2.0)$$

where P^{-1} is the inverse of the $n \times n$ variance-covariance matrix P , $| \cdot |$ denotes determinate, \bar{z} is the column vector of means with n elements and T denotes transpose, the conversion of column vector into a row vector (Appendix 1). As in the univariate case (1.02), the square root term is a normalization factor that insures that the probabilities sum to one.

2.2 The directional selection differential, s , a vector.

We need to consider the effects of selection on the multivariate distribution, $p(z)$. We recall from 1.1 that relative fitness is the variable that translated the distribution before selection, $p(z)$, into the distribution after selection, $p(z)^*$. In our multivariate world, absolute fitness, $W(z)$, and relative fitness, $w(z)$, are functions of a multi-trait phenotype, z . Indeed, we can substitute z for z , and a vector of trait means for \bar{z} in eq. (1.0-1.7) and those same expressions apply, without an assumptions about the multivariate distribution of z . We will pause to consider the selection differential, which is now an n -element column vector. In the 2-trait case it is

$$s = Cov(w, z) = \begin{bmatrix} Cov(w, z_1) \\ Cov(w, z_2) \end{bmatrix} = \bar{z} - \bar{z}^* = \begin{bmatrix} \bar{z}_1 - \bar{z}_1^* \\ \bar{z}_2 - \bar{z}_2^* \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \quad (2.01)$$

2.3 The directional selection gradient, β , a vector.

We can now consider a new measure of selection, one that will account for correlations among the traits. That new measure is the *directional selection gradient*, an n -element column vector. In general and in the 2-trait case it is

$$\beta \equiv P^{-1}s = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \quad (2.02)$$

where β_1 and β_2 are the selection gradients for traits 1 and 2, respectively. We can rearrange this last expression in a way that disentangles the direct and indirect effects of directional selection,

$$s = P\beta = \begin{bmatrix} P_{11} & P_{12} \\ P_{12} & P_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} P_{11}\beta_1 + P_{12}\beta_2 \\ P_{12}\beta_1 + P_{22}\beta_2 \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}. \quad (2.03)$$

This portrayal of s tells us that the selection differential for a trait is composed of one term that represents the direct effect of selection on that trait on its mean, e.g., $P_{11}\beta_1$, and another term that represents the indirect effect of selection on another trait, acting through the covariance between the two traits, e.g., $P_{12}\beta_2$ (Fig 2.z = path diagram). Of course, in the general case, with n traits, the indirect effects can be very numerous; the selection differential for trait 1 is

$$s_1 = P_{11}\beta_1 + P_{12}\beta_2 + P_{13}\beta_3 + \dots + P_{1n}\beta_n = P_{11}\beta_1 + \sum_{i=2}^n P_{1i}\beta_i. \quad (2.04)$$

Indeed, we see that it would be possible for the sum of indirect terms in (2.04) to overwhelm the direct effect, so that the selection differential, s_1 , could have opposite sign to the selection gradient, β_1 ! In other words, when we consider the actual value for a particular selection differential, e.g., s_1 , that value may reflect selection on other traits rather than selection on the trait in question.

2.4 The nonlinear selection differential, C , a matrix.

In general, selection can change the variances and covariances of the traits, as well as shift their means (Fig. 2.ww = series of panels showing changes in 95% confidence ellipses of various bivariate Gaussian distributions; this same set of panels could correspond, in a coming chapter, to a set of panels showing the corresponding ISS and AL surfaces!). Change in second moments (variance and covariance) is considerably more complicated than in the univariate case, so we will need to consider it in detail. Focusing on the 2-trait case for concreteness, the phenotypic variance-covariance matrix before selection is

$$P = \int p(z)(z - \bar{z})^T(z - \bar{z})dz = \begin{bmatrix} P_{11} & P_{12} \\ P_{12} & P_{22} \end{bmatrix}, \quad (2.05)$$

where P_{11} and P_{22} are the variances of z_1 and z_2 , and P_{12} is their covariance before selection. Selection may change any or all of the elements in P , so that after selection the matrix becomes

$$P^* = \int p(z)^*(z - \bar{z})^T(z - \bar{z})dz = \begin{bmatrix} P_{11}^* & P_{12}^* \\ P_{12}^* & P_{22}^* \end{bmatrix}. \quad (2.06)$$

Following the argument in 1.4, we will want to employ a correction factor for effects of directional selection on variances and covariances. Here and hence forth we assume multivariate normality of the trait distribution before selection. In the 2-trait case, that correction factor is

$$ss^T = \begin{bmatrix} s_1^2 & s_1 s_2 \\ s_1 s_2 & s_2^2 \end{bmatrix}, \quad (2.07)$$

where the diagonal terms are effects on the trait variance (which will always be positive) and the off-diagonal term is the effect on the trait covariance (which may be positive or negative). In other words, we want to correct for the fact that directional selection on trait i will reduce its variance by an amount s_i^2 , regardless of the sign of s_i , while directional selection on traits i and j will change their covariance by an amount $s_i s_j$. In the 2-trait case, the stabilizing selection differential is

$$C = P^* - P + ss^T = \begin{bmatrix} P_{11}^* - P_{11} + s_1^2 & P_{12}^* - P_{12} + s_1 s_2 \\ P_{12}^* - P_{12} + s_1 s_2 & P_{22}^* - P_{22} + s_2^2 \end{bmatrix} = \begin{bmatrix} Cov(w, \tilde{z}_1^2) & Cov(w, \tilde{z}_1 \tilde{z}_2) \\ Cov(w, \tilde{z}_1 \tilde{z}_2) & Cov(w, \tilde{z}_2^2) \end{bmatrix}. \quad (2.08)$$

Directional selection shifts the expected value of z , the mean (1.7); nonlinear selection shifts the expected values of the quadratic variables \tilde{z}_1^2 , \tilde{z}_2^2 , and $\tilde{z}_1 \tilde{z}_2$. Notice that by the nature of ss^T (2.07), the nonlinear selection differential for a particular trait will be $C_{ii} = P_{ii}^* - P_{ii} + s_i^2$, regardless of how many traits are under selection. In other words, the diagonal elements in C correct for the effects of directional selection on the trait in question and not for effects exerted through correlations with other traits.

2.5 The nonlinear selection gradient, γ , a matrix.

Just as we can solve for a vector, β , that corrects for trait correlations in measuring directional selection, we can by similar operations obtain a set of selection coefficients that correct for trait correlations in measuring nonlinear selection. Those coefficients constitute the *nonlinear selection gradient*, an $n \times n$ symmetric matrix, defined here in the general case and illustrated in the 2-trait case,

$$\gamma \equiv P^{-1}CP^{-1} = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{12} & \gamma_{22} \end{bmatrix}, \quad (2.09)$$

where γ_{11} and γ_{22} represent, respectively, the direct effects of stabilizing/disruptive selection on the variances of traits 1 and 2, and γ_{12} represents the direct effect of correlational selection on the covariance of traits 1 and 2, i.e., P_{12} . To see what constitutes direct and indirect effects, we can rearrange the 2×2 version of (2.09) to obtain

$$C = P\gamma P = \begin{bmatrix} P_{11} & P_{12} \\ P_{12} & P_{22} \end{bmatrix} \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{12} & \gamma_{22} \end{bmatrix} \begin{bmatrix} P_{11} & P_{12} \\ P_{12} & P_{22} \end{bmatrix} = \begin{bmatrix} P_{11}^2\gamma_{11} + 2P_{11}P_{12}\gamma_{12} + P_{12}^2\gamma_{22} & P_{11}P_{12}\gamma_{11} + P_{12}^2\gamma_{12} + P_{11}P_{12}\gamma_{12} + P_{12}P_{22}\gamma_{22} \\ P_{11}P_{12}\gamma_{11} + P_{12}^2\gamma_{12} + P_{11}P_{12}\gamma_{12} + P_{12}P_{22}\gamma_{22} & P_{12}^2\gamma_{11} + 2P_{12}P_{22}\gamma_{12} + P_{22}^2\gamma_{22} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{bmatrix}. \quad (2.10)$$

The $P_{11}^2\gamma_{11}$, $P_{22}^2\gamma_{22}$, and $P_{12}^2\gamma_{12}$ terms represent, respectively, direct effects on \tilde{z}_1^2 , \tilde{z}_2^2 , and $\tilde{z}_1\tilde{z}_2$. All of the other terms represent indirect effects mediated through covariances between the quadratic variables \tilde{z}_1^2 , \tilde{z}_2^2 , and $\tilde{z}_1\tilde{z}_2$.

2.55 The canonical form of the gamma-matrix

The picture of multivariate nonlinear selection presented by γ may be easier to visualize if we use a different coordinate system. A natural alternative to the original trait axes is one in which the γ -matrix takes a diagonal form in which all the off-diagonal terms, which describe correlational selection, are zero. This diagonal form of the γ -matrix is known as its canonical form (Phillips & Arnold 1989)

$$\Lambda = \mathbf{M}^T \gamma \mathbf{M}, \quad (2.11)$$

where Λ is matrix with the eigenvalues of γ , λ_i , on its diagonal and zeros as its off-diagonal elements, and \mathbf{M} is a matrix whose columns are the eigenvectors of γ (normalized to unit length). The new axes, the eigenvectors, are a rotation of the original axes, and like those axes, they are orthogonal to one another.

2.6 Estimates of multivariate selection gradients

Selection on beak dimensions (mm) in a Galapagos finch, illustrates a surprising case in which a selection gradient is strikingly different from the corresponding selection gradient.

Table 2.2 Directional selection on body size and beak dimension in a Galapagos finch (*Geospiza fortis*) arising from drought conditions on Daphne Major (Price et al. 1984). The sample size before selection is 640; after selection the sample is 96. An asterisk indicates significance at the 0.05 level. The test statistic for the selection differential is a T-test comparing the means of samples before and after selection.

trait	Selection differential, s	Selection gradient, $\beta \pm \text{s.e.}$
Weight (g)	0.62*	0.51±0.14*
Beak length	0.49*	0.17±0.18
Beak depth	0.60*	0.79±0.23*
Beak width	0.49*	-0.47±0.21*

The focus is on directional selection associated with a change in environmental conditions and, indeed, this form of selection is dramatically stronger than in the preceding example. For example, selection associated with drought conditions shifted average body weight by 62% of the standard deviation in this trait before selection. Although this and the other selection differentials suggest that heavier birds with larger beak dimensions were favored by selection, the directional selection gradients tell a different story. They suggest that while selection favored heavier birds, it also favored birds with deep, narrow beaks. The picture of selection revealed by the selection gradients is consistent with ecological observations during the drought. Hard seeds became disproportionately common during the drought, and only large birds could crack these seeds. Birds with narrow beaks are apparently favored by selection because birds with this morphology could twist and open the woody seed pods of *Tribulus cistoides* (Price et al. 1984). But how can the selection gradient for beak width have a different sign than its selection differential? We see from expression 2.04 that a shift in the mean, s_i , is composed of contributions from correlated traits. If those contributions ($\beta_j P_{ij}$) have positive signs, they can overwhelm the direct effect of selection on beak width ($\beta_i P_{ii}$), which has a negative sign. Calculating β can help untangle the direct and indirect effects and reveal the actual targets of selection. Finally, this example illustrates the value of standardized selection coefficients. In the case of weight, $\beta=0.51$ means that if direct selection increased body weight by a before-selection standard deviation, relative fitness would be increased by 51%. Decreasing beak width by a standard deviation would increase relative fitness by 47%.

2.7 Functional complexes

Correlation among traits is a fact of life that motivates the calculation of the selection gradients that we introduced in the preceding sections. Those calculations enable us to account for trait correlations in assessing the effects of selection. But, trait correlations also have a deeper meaning. We can view trait correlations as the product of selection that occurs when traits work together to perform a common function (Olson & Miller 1958). Functional interactions of this kind can help to build trait correlations (covariances) (cite Lande, Cheverud, Wagner). Consequently, it follows that we can detect functional interactions by assessing the effects of selection on trait covariances. The tools for that assessment are the off-diagonal elements of the C-matrix and especially of the γ -matrix. But, before turning to the assessment value of those two matrices, let's focus on the biology of functional interaction and the grouping of traits into functional complexes. As we shall see in the chapters ahead, these functional interactions have diverse and important effects on evolutionary processes and patterns.

Functional interaction between characters is a ubiquitous feature of life, especially in metazoans. The organization of interacting characters into functional complexes is particularly obvious if we focus on complexes that are dedicated to well-defined ecological or social functions. Spider web ... venom delivery in *Conus* snails ...

pheromone delivery in

plethodontid salamanders ...

pollinator attraction and

attachment of pollinia in orchids

... In all of these instances a

particular measure of whole

organism performance

summarizes the interactions of

characters in the complex ... prey

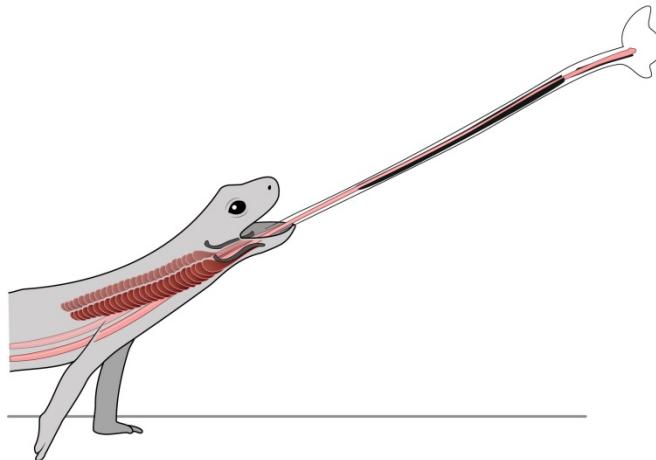
capture in spiders and *Conus* ...

tongue projection in chameleons

and bolitoglossine salamanders

... mating success in male

salamanders ... pollen export in orchids ...



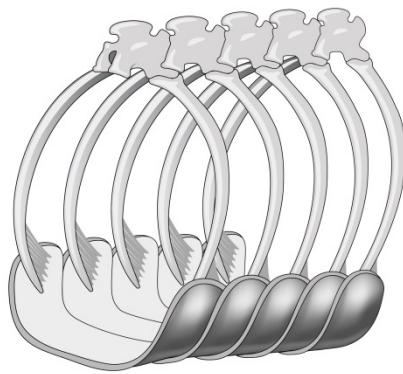
{In the next few paragraphs, portray traits showcased in multiple chapters as parts of functional complexes: snake vertebrae in locomotion ...}

Fig x.yy The tongue projection mechanism of a plethodontid salamander. The tongue skeleton (black) is normally folded inside the body cavity but is projected forward during prey capture by contracted protractor muscles (dark red). Retractor muscles (light red), attached to the tongue skeleton, run the full length of the body and attach to the pelvis. The contraction of these muscles pulls the tongue and attached prey back into the body (after Deban et al. 1997)

Functional complexes are equally recognizable when sets of characters participate in well understood physiological functions ... in vertebrates for example ... cardio-vascular system ... heart, vessels, lungs ... delivery of oxygen to peripheral tissues

during exercise ... water transport system in plants ... phloem, xylem, root system ... water delivery to

Figure 4.qq. A five-vertebrae segment of the vertebral column of a natricine snake (*Natrix natrix*), showing functional connection between three elements. Ribs articulate with the vertebral column, but muscles also connect the tips of the ribs of the ventral scales on the snake's ventral surface. Furthermore, a complicated system of muscles (not shown) connects between the ribs. (after a 19th century instructional poster).



peripheral tissues ... {one or two additional examples} ... As in the case of functional complexes directed at mates and prey, the performance of physiological character complexes can often be summarized by one or a few metrics that capture the salient features of operation in the entire complex.

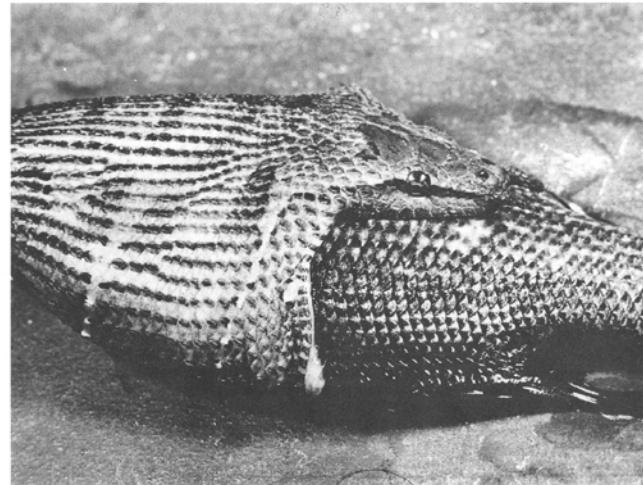
The unifying feature in all of these examples is interaction and coordination between the elements in the complex that can be summarized and expressed in measures of performance. The ease with which such functional complexes can be listed suggests that trait organization of this kind is so common as to be virtually ubiquitous. In any case, the ease of inventory suggests that functional complexes are extremely common, so that they subsume many if not all traits in organisms. The significance of this ubiquity is that the mode of selection required to account for functional complexes is likewise ubiquitous and important. In the chapters that follow we will identify multivariate stabilizing selection, and especially correlational selection, as the mode of selection chiefly responsible for the assembly and maintenance of functional complexes. To see this connection more clearly we must first tackle the problem of how the performance of functional complexes is related to lifetime fitness.

2.8 Morphology, performance, and fitness

Introduce performance here, at the start of this paragraph Important measures of performance have been called ‘surrogates for fitness’. This characterization is unfortunate because it suggests that we are pretending that performance is fitness. Instead of that pretense, we wish to understand how performance is related to fitness, for with that understanding we will also be able characterize selection on functional complexes.

The solution to our problem lies in recognizing that selection on a trait in a functional complex can be subdivided into two parts: the relationship between the trait and performance, and the relationship between performance and fitness. Consider the example of the traits that come into play when a snake swallows a large prey item. Snakes ingest their prey whole

Figure 4.xx A tentacled snake (*Erpeton tentaculum*) in the process of swallowing a large fish (from Arnold 1983)



(Fig. 4.xx) and consequently the maximum size that can be ingested is a function of the sizes of a series of five structural elements (Fig. 4.xy). In an experimental study of swallowing performance our model of performance, f , might be

$$f = \beta_{fz1}z_1 + 2\beta_{fz2}z_2 + 2\beta_{fz3}z_3 + 2\beta_{fz4}z_4 + \beta_{fz5}z_5 + \varepsilon_f, \quad (zz)$$

where the subscripted β s represent directional coefficients describing the effect of each of the 5 traits on performance, and the factors of 2 account for traits that participate twice in effects on performance. These β coefficients are not directional selection gradients, which would represent effects of traits on fitness, so we will call them *directional performance gradients*.

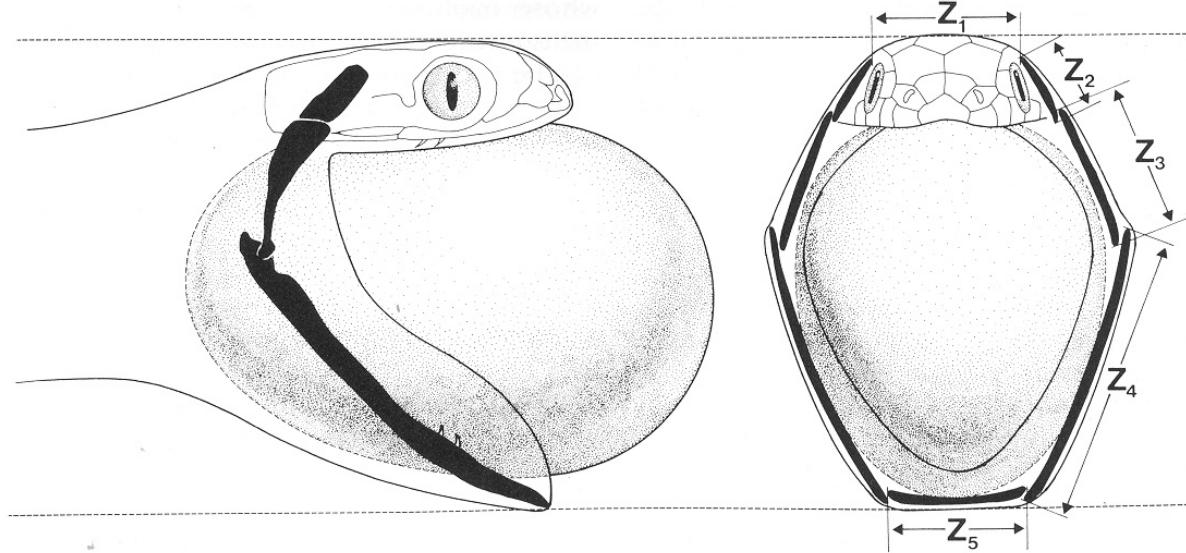


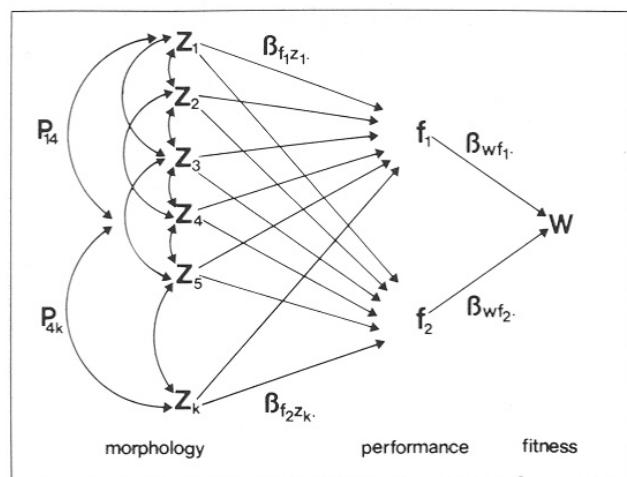
Figure 4.xy An African egg-eating snake (*Dasypeltis* sp.) in the process of swallowing its prey. The lengths of 5 structural elements determine the maximal cross-sectional area of prey than can be ingested: width of the brain case (z_1), and the lengths of the supratemporal (z_2), quadrate (z_3), mandible (z_4), and mandibular symphysis (z_5) (from Arnold 1983).

The relationship between a performance gradient and an ordinary selection gradient becomes clear if we construct a path diagram. In Fig. 4.yy, we have made our example slightly more general by including two performance measure, f_1 and f_2 , and by increasing the number of traits from 5 to k . Figure 4.yy is a diagram that portrays a nested regression model. We have taken our regular expression for multivariate directional selection,

$$w = \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_k z_k + \varepsilon_w$$

and converted it into a model in which all of the k traits exert their effects on w via two performance measures. In other words,

$$w = \beta_{wf_1} f_1 + \beta_{wf_2} f_2 + \varepsilon_w,$$



where the two subscripted β s represent directional effects of performance on fitness, which we shall call *fitness gradients*. By substituting two expressions like (zz) into this last expression, we see that each directional selection gradient is the product of a performance gradient and a fitness gradient (Fig. 4.yz).

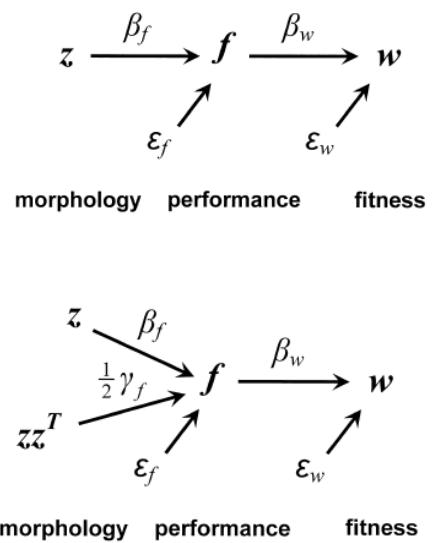
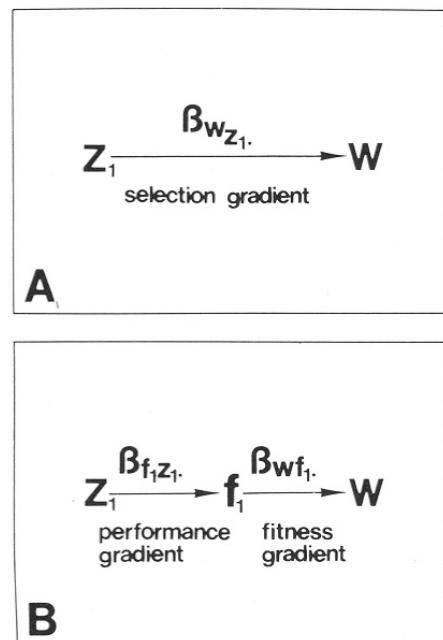
Figure 4.yy. Path diagram showing causal relationships (paths with single arrowhead) between k morphological traits (z_1, z_2, \dots, z_k), two measures of performance (f_1 and f_2), and relative fitness (w). Double-headed arrows represent phenotypic covariances between traits, before selection. Two of these covariances are labeled, P_{14} and P_{4k} . (from Arnold 1983)

Figure 4.yz. Total directional selection gradient on a trait (A) can sometimes be partitioned into two parts (B): a performance gradient and a fitness gradient.

The significance of this product rule (Fig. yz) is that even when it is not feasible to estimate selection gradients or fitness gradients, it may often be possible to estimate performance gradients. Performance gradients are more routinely tractable because they can be estimated in the laboratory via performance testing, whereas the other two kinds of gradients require assessment of fitness or its components under field conditions. Under ideal circumstances, all three kinds of gradients can be estimated in a combined laboratory and field study (e.g., Tsuji et al's *Sceloporus* study).

Clearly morphological and other kinds of traits can have curvilinear as well as directional effects on performance, so we need to consider a more general model that includes those effects.

The algebra of such a model is detailed by Arnold (2003). Without recounting those details here, we can note that in general the linear model that we have just considered can be represented by a multivariate path diagram in which all of the elements, except fitness, are vectors (Fig. 4.wz). If we add curvilinear



effects of traits on performance, that extended model can be represented by the path diagram shown in Fig. 4.wz. Although this model is tractable and estimable, it is obviously a considerable simplification. We have assumed that fitness is a planar rather than a curvilinear function of performance. Although that assumption is justified and appropriate if our goal is to estimate directional fitness gradients, we need empirical inquiries to establish whether the surface relating performance to fitness can indeed be approximated by a plane.

Figure 4.wz. A path diagram showing multivariate directional performance and fitness gradients in which all of the elements, except fitness, are vectors. (from Arnold 2003).

Figure 4.zw. A path diagram showing multivariate directional (β_f , a vector) and curvilinear (γ_f , a matrix) performance gradients. Other conventions as in Fig. 4.wz. (from Arnold 2003)

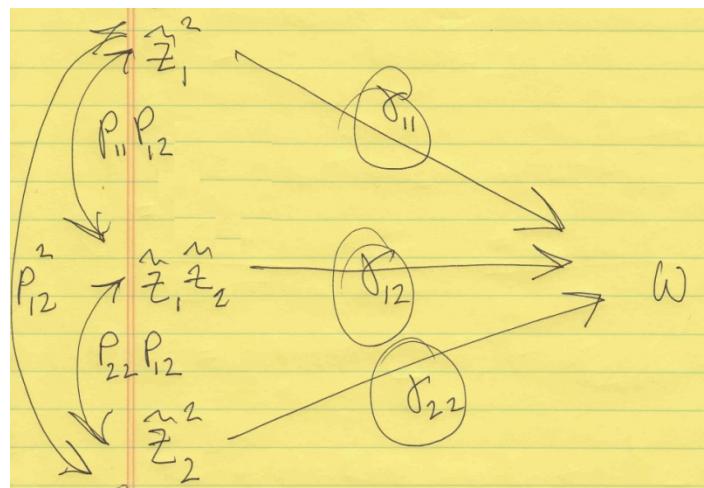


Figure 4.zz Path diagram view of (2.10) showing the effects of the quadratic variables \tilde{z}_1^2 , \tilde{z}_2^2 , and $\tilde{z}_1\tilde{z}_2$ on relative fitness, w . Coefficients are shown on each path. Double-headed arrows denote covariances between quadratic variables ##Include something like this figure? At a minimum it needs some work, e.g. coefficients of $\frac{1}{2}$ on gamma11 and 22##

{need a concluding paragraph here for this section, relating the above results to the issue of the mode of selection acting on functional complexes}

2.9 Morphology and performance in garter snakes

Returning to the example of crawling speed in newborn garter snakes introduced in Chapter 1, we now consider how accounting for trait covariance affects our concept of selection. This example is also an exercise in identifying the most important trait dimension for functional integration. In the first place, we have a new nonlinear selection differential to consider, C_{12} , which describes the total effect of nonlinear selection on the trait covariance, P_{12} . In the sample after selection, trait covariance has increased by 19.2%. Bootstrapping reveals that in only 1 boot sample out of 1000 did C_{12} take a value less than zero, so the change in covariance is highly significant ($P=0.001$). Virtually all of the 19.2% expansion is due to nonlinear selection; directional selection contributes only a 0.8% decrease in covariance. The selection gradients, which account for trait covariance, give much the same picture of selection as the selection differentials. Although most gradients are nonsignificant, the nonlinear selection gradient describing the effect of the trait product, z_1z_2 , on crawling speed is positive and

highly significant ($P < 0.001$). The positive value for this coefficient, γ_{12} , indicates that selection directly increases trait covariance. In other words, comparing samples before and after selection we see a change in trait covariance (from 0.073 to 0.265), but that change largely reflects direct effects on covariance rather than indirect effect via trait means or variances. The direct effect of selection is to increase the integration of the two traits.

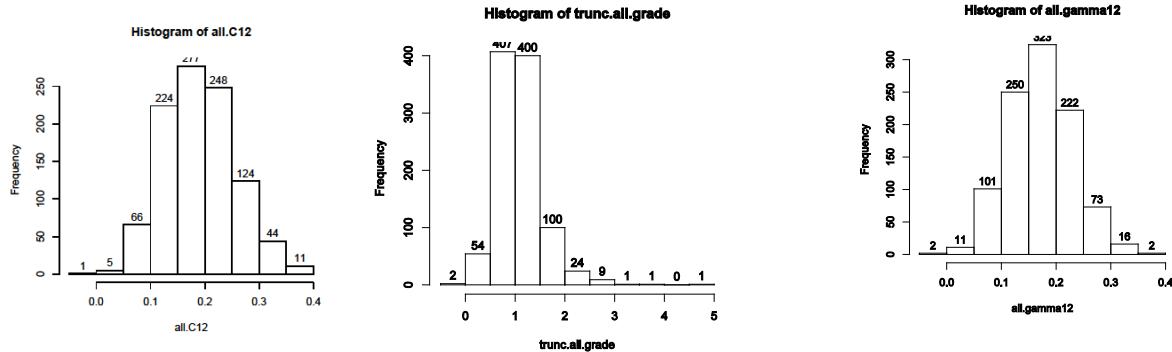
A useful way to visualize the picture presented by the γ -matrix is to rotate trait axes so that γ takes a so-called canonical form (Phillips & Arnold 1989). In this canonical form, the off-diagonal elements are zero, so no direct effects act on trait covariance. The γ coefficients (eigenvalues) corresponding to the new major and minor axes are, respectively 0.253 (0.95 CI = 0.136, 0.489) and -0.617 (0.95 CI = -0.221, 0.088). In other words, selection on the major axis is disruptive and highly significant, while selection on the minor axis might be either stabilizing (as it is in the point estimate) or disruptive, but non-significant. The major axis (leading eigenvector) is inclined at about a 45 degree angle in trait space with a slope of 1.055 (0.95 CI = 0.380, 2.156); the minor axis is perpendicular to the major axis. The leading eigenvector of γ is important to us because we have shown that it is the trait dimension of greatest functional integration (see next section), i.e., the trait direction in which trait covariance is most affected by selection exerted through crawling speed. We will return to the issue of visualization in Chapter 4 when we consider the topic of multivariate selection surfaces.

Using this example to illustrate (2.11), taking the values for M and γ from Table 2.1, we find that the canonical form of the the γ -matrix is

$$\Lambda = M^T \gamma M = \begin{bmatrix} 0.688 & -0.726 \\ 0.726 & 0.688 \end{bmatrix}^T \begin{bmatrix} 0.097 & 0.176 \\ 0.176 & 0.116 \end{bmatrix} \begin{bmatrix} 0.688 & -0.726 \\ 0.726 & 0.688 \end{bmatrix} = \begin{bmatrix} 0.282 & 0 \\ 0 & -0.069 \end{bmatrix}.$$

Table 2.1 Selection differentials and gradients describing effects of numbers of body and tail vertebrae on crawling speed in newborn garter snakes (*Thamnophis radix*) (n=143). Selection gradients were estimated using expressions 2.02 and 2.09. The standard errors of β and γ were estimated from the standard deviations of corresponding bootstrap distributions (n=1000 samples with replacement). Bootstrap estimates of 95% confidence limits are shown in parentheses. Significance levels: *, P<0.05; **, P<0.01; ***, P≤0.001. Eigenvalues of the γ -matrix corresponding to the leading and minor eigenvector are denoted λ_1 and λ_2 , respectively. 'Eigenvector' is the slope of the leading eigenvector.

	Body vertebrae		Tail vertebrae			
	mean	variance	mean	variance	covariance	correlation
Before selection	153.414	11.437	74.257	9.094	0.743	0.073
Before selection'	0.000	1.000	0.000	1.000	0.073	0.073
After selection'	0.031	1.122	0.004	1.142	0.265	0.234**
	s ₁	C ₁₁	s ₂	C ₂₂	C ₁₂	
selection differential	0.031 (-0.032, 0.096)	0.123 (-0.062, 0.348)	0.004 (-0.058, 0.063)	0.142 (-0.015, 0.327)	0.192*** (0.074, 0.328)	
correction term, ss^T		0.077		0.001	0.008	
	β_1	γ_{11}	β_2	γ_{22}	γ_{12}	eigenvector
selection gradient+	0.031 (-0.033, 0.095)	0.097 (-0.087, 0.301)	0.002 (-0.065, 0.059)	0.116 (-0.028, 0.292)	0.176** (0.066, 0.288)	1.055** (0.380, 2.156)
standard error++	0.033	0.098	0.031	0.082	0.059	$\lambda_1=0.282$ (0.136, 0.489) $\lambda_2=-0.069$ (0.136, 0.489)



*Use these histograms; probably not**

2.10 Technical issues in estimating and interpreting selection gradients

Several cautions, common to all multivariate statistical analyses, should be kept in mind in estimating β and γ and interpreting those estimates. As we will see in Chapter 4, the key formulas for β and γ (2.2 and 2.9) are equivalent to formulas for sets of multiple regression coefficients. Consequently, the issues we need to consider are usually discussed in the context of estimation by multiple regression (Lande & Arnold 1983, Mitchell-Olds & Shaw 1987); we will view them in that light in Chapter 4.

Not including correlated traits in the study can lead to biased estimates of β and γ . In particular, we need to consider the possibility that traits under directional and/or stabilizing selection are correlated with the measured traits but are not included in the study. This circumstance can cause us to over- or under-estimate our selection gradients, depending on the sign and magnitude of selection on the unmeasured trait and the pattern of correlation with the measured traits. In other words, what we interpret as direct effects of selection on our traits can be affected by traits that are excluded from the analysis for one reason or another. In practice, most investigators live with this limitation on interpretation because traits are included in the analysis precisely because they are likely to be under selection. In other words, prior information is brought to bear in choosing traits that partially mitigates the problem of influence from unmeasured traits. Nevertheless, the possibility of this kind of complication can never be completely eliminated and should be borne in mind in interpreting results.

Unmeasured environmental variables can produce an illusion of selection. This problem is related to the one just discussed, but here we are concerned with environmental effects that produce correlations between fitness and traits. Mitchell-Olds & Shaw (1987) discuss and tell a hypothetical example in which growing

conditions vary spatially and cause some plants to be both large and fecund and others to be small and barren. If we fail to include growing conditions in our analysis (e.g., as covariates), we might erroneously conclude that plant size is under strong selection. As in the case of correlated, unmeasured traits, biased estimates have lead us to a false conclusion. One antidote is the realization that correlational studies may not reveal causal relationships. In multivariate statistical analysis, we attempt to correct for correlations, but we may not succeed. For this reason, it is always wise to do a companion experimental study that manipulates traits of interest and gets us closer to an inference of causality.

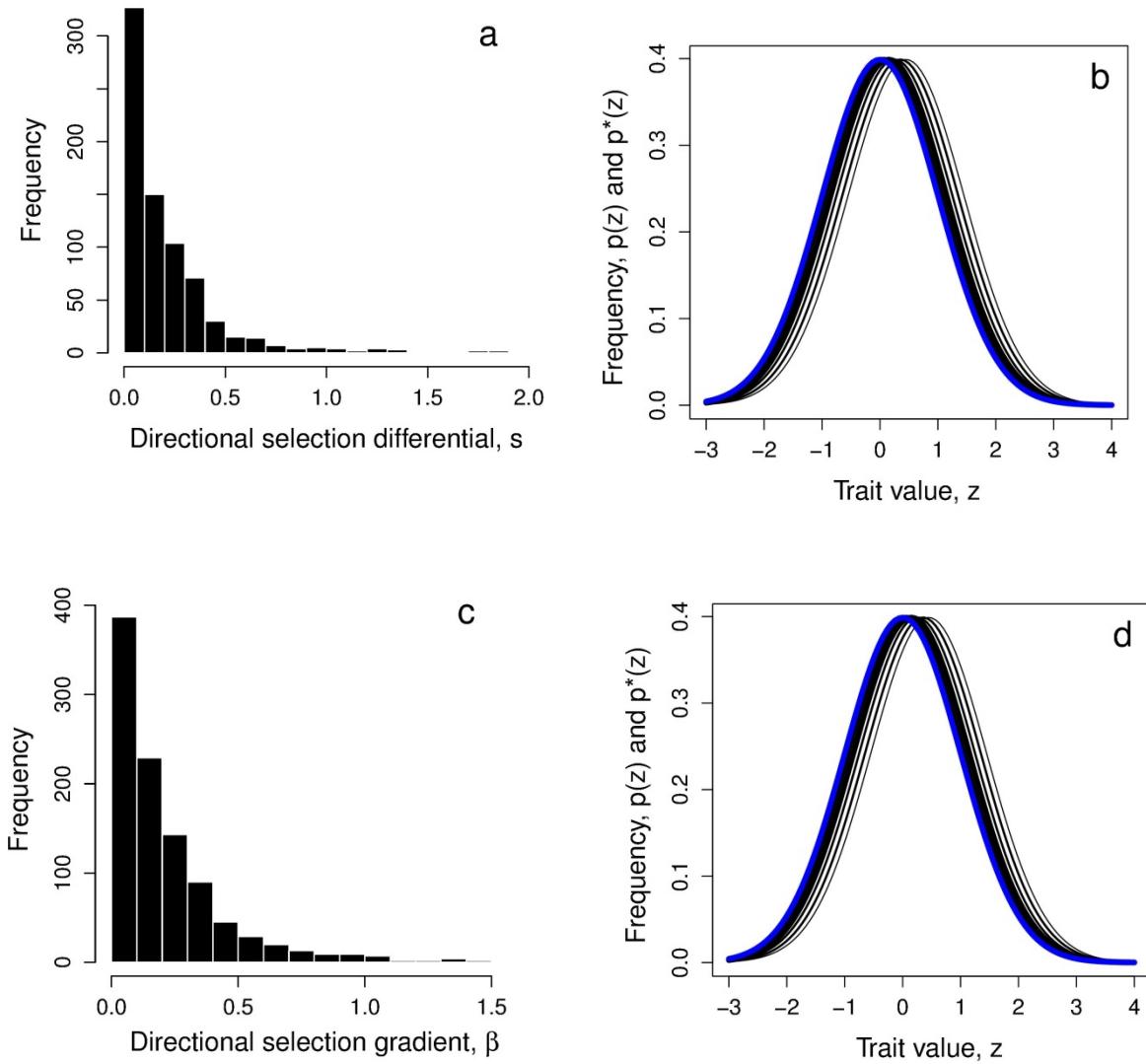
It is possible to make a strong inference of causality within the framework of correlational study. The argument is vividly portrayed in court cases that challenged the claim of tobacco companies that the link between smoking and lung cancer was not causal. Attorneys for cancer victims argued that three conditions help make a strong case for causality: plausibility, strength of correlation, and prevalence of correlation across replicate studies. All of these conditions can be considered in deciding whether a particular selection gradient represents a causal effect of phenotype on fitness.

The expressions presented in this chapter for estimating β and γ are useful when a population is sampled before and after selection (i.e., for the case of cross-sectional data). Those estimations require making assumptions about the two samples since they may no individual in common (Lande & Arnold 1983). A key assumption then is that the sample before selection closely resembles the actual set of individuals after selection before those individuals were exposed to selection. Alternatively, individuals may be followed through time so that their individual fitness values are assessed, bypassing these assumptions. In the case of such longitudinal data, a multiple regression approach can be used to estimate selection coefficients. That regression approach is described in Chapter 4. The notorious factor of 1/2 problem surrounding the estimates of γ in the literature before 2009 will also be discussed in Chapter 4 (Stinchcombe et al. 2009).

2.9 Surveys of selection gradients

The point of the surveys in this section is that the β and γ estimates correct for the effects of selection on multiple traits, unlike the surveys in Chapter 1 which showed shifts in mean and variance that reflect both direct and indirect effects. The advantage is that comparisons between s and *beta* and between C and *gamma* allow us to see how much the aggregated indirect effects contribute to the overall distribution of selection coefficients.

The surveys summarized here were compiled by Hoekstra et al. (2001), Kingsolver et al. (2001), and Stinchcombe (2009). Kingsolver et al. (2001) and Hoekstra et al. (2001) updated Endler's (1986) survey, using similar criteria and obtaining a sample about an order of magnitude larger. This more recent survey of 62 studies (63 species) focused on selection acting on natural populations in natural circumstances. Like Endler's sample, the survey included a wide range of organisms (plants, invertebrates, vertebrates) and an even wider range of traits (most were morphological measurements and counts, but some behavior and life history traits were included). The sample consists of studies published between 1984 and 1997 and so slightly overlaps with Endler's sample. Stinchcombe (2008) uncovered errors in the estimation of γ that



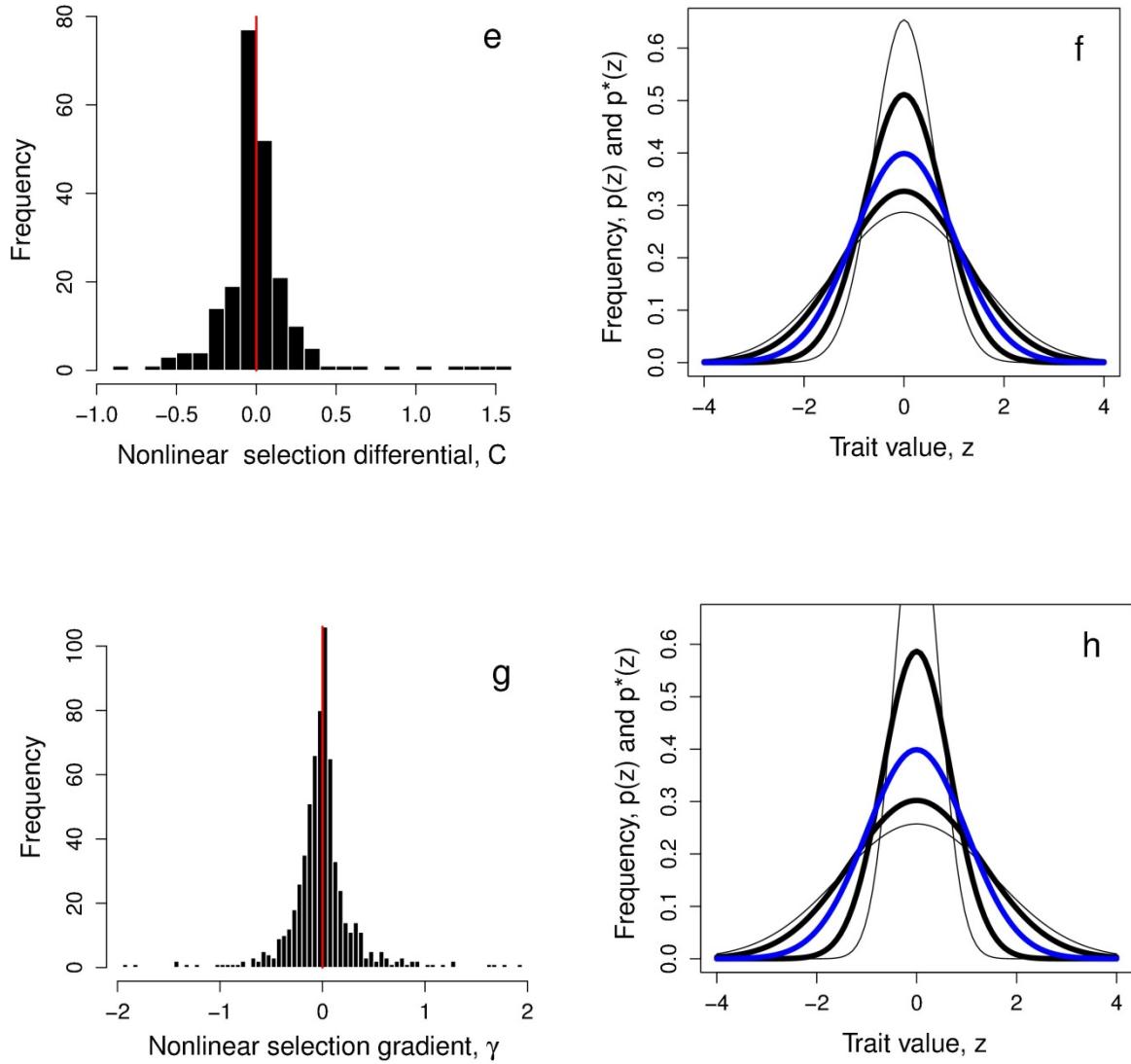


Figure 2.5 Histograms of selection gradients and differentials paired with frequency distributions that portray the magnitude of effects on means and variances. (a) Histogram of the absolute values of directional selection differential estimates, s ; 746 values from Kingsolver et al 2001 database. Three values greater than $\text{abs}(2.0)$ are not

included. (b) Shifts in mean corresponding to the directional selection differentials in Fig. 2.5a. Shifts corresponding to the five bins on the right-most side of the distribution (0.05-0.45) are shown in black and account for 92% of the observations. (c) Histogram of the absolute values of directional selection gradient estimates, β ; 992 values from Kingsolver et al 2001. Three values for s greater than $\text{abs}(2.0)$ were not included. (d) Shifts in mean corresponding to the directional selection gradients in Fig. 2.5c. Shifts corresponding to the five bins on the right-most side of the distribution (0.05-0.45) are shown in black and account for 90% of the observations. (e) Histogram of nonlinear selection differential estimates, $C = (P^*-P+s^2)/P$; 220 values from Kingsolver et al 2001 database. Five values greater than $\text{abs}(2.0)$ are not included. The vertical red line separates the distribution into negative and positive estimates. (f) Shifts in variance corresponding to the nonlinear selection differentials shown in Fig. 2.5e. Shifts corresponding to the four most populated bins in the center the distribution (-0.15 to 0.15) are shown in black and account for 78% of the observations. (g) Histogram of nonlinear selection gradient estimates, γ ; 653 values from the Stinchcombe et al 2008 database, after deleting 7 values of $\gamma > 2$ and 4 values < -2 . (h) Shifts in variance corresponding to the nonlinear selection gradients shown in Fig. 2.5g. Shifts corresponding to the 16 most populated bins in the center the γ distribution (-0.4 to 0.4) are shown in black and account for 86% of the observations.

permeate the literature prior to 2008. They resampled the studies compiled by Kingsolver et al (2001) and corrected those errors in γ and added many additional studies. That new database was used to make the histogram shown in Fig. 2.5g.

From a statistical point of view, distributions of selection differentials and gradients are extremely similar. We get the same overall picture of directional selection, for example, whether we look at the differential, s , or the gradient, β . In both cases we see a distribution that is negative exponential in appearance with a modal value close to zero and with the vast majority of estimates in the range 0 to 0.5 (Fig. 2.5 a, Fig. 2.5c). In qualitative terms, directional selection tends to be weak, rarely shifting mean by more than half a within-population phenotypic standard deviation. The similarity between the distributions of s and β suggest that the indirect effects of selection, arising from phenotypic correlations between traits, make at most a minor contribution to the overall picture.

Turning to the distributions of nonlinear selection differentials and gradients, we again see much the same distributional picture. Both selection coefficients show leptotic (peaked) distributions that are almost symmetrically centered about zero (Fig 2.5e, Fig 2.5g). The selection differential C is slightly biased towards negative values (stabilizing selection), while the gradient γ has a slight bias in the other direction, towards

disruptive selection. The overall effect of indirect effects appears to inflate the absolute values of C , so that its distribution is less leptotic than that of γ . Judging from the distribution of γ , it is unusual for direct nonlinear selection to increase or decrease trait variance by more than 40%. A comparison of the distribution of γ compiled by Kingsolver et al. (2001) with the one compiled by Stinchcombe (2008) yields a surprising result. The two distributions are virtually identical! Although individual estimates of γ may be inaccurate by a considerable amount prior to 2008, in the aggregate the errors appear to cancel out.

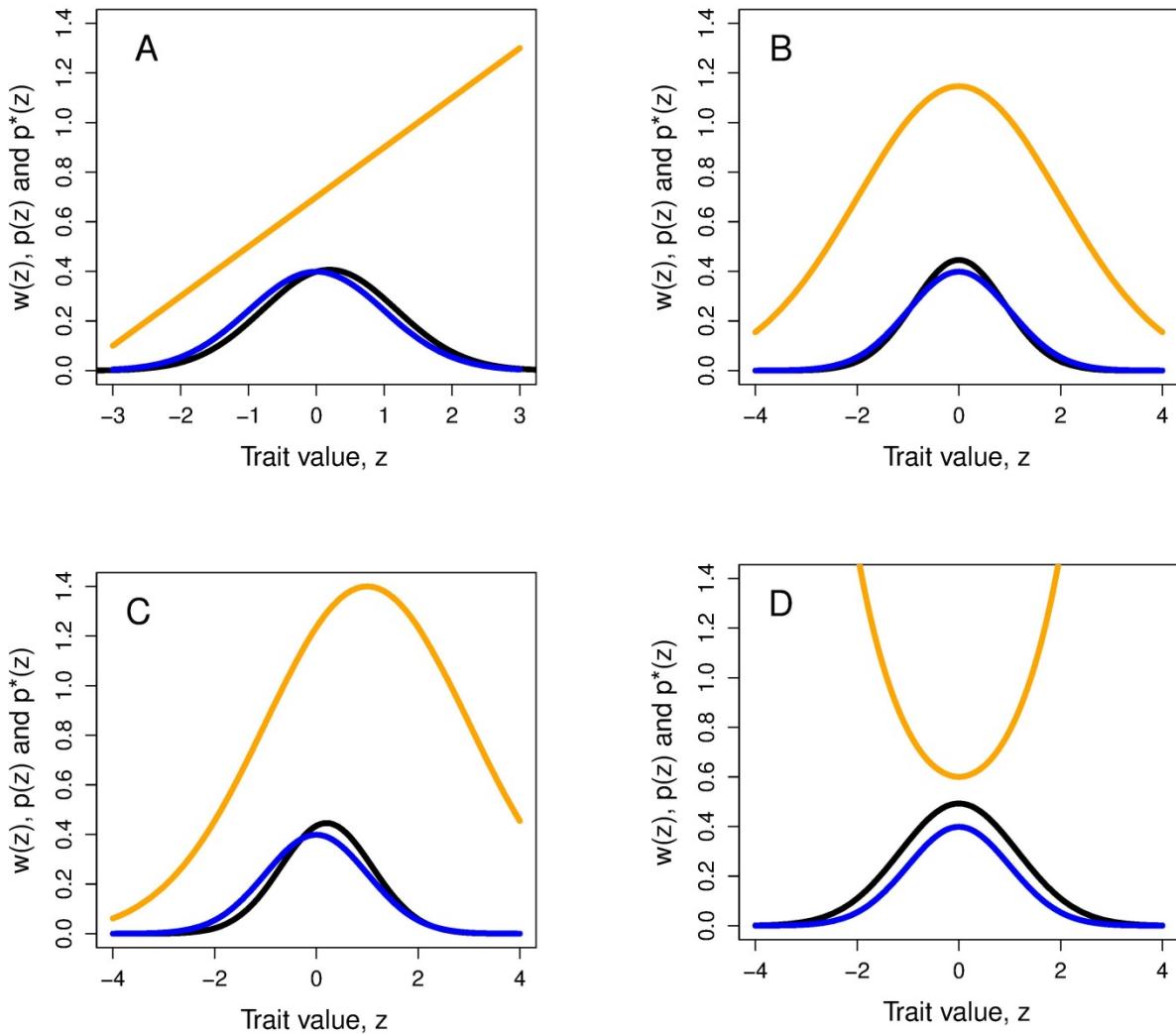
Chapter 3: The Selection Surface and Adaptive Landscape for a Single Trait

© Stevan J. Arnold 2016

Overview.- Selection on a single trait can be visualized as a curve that relates trait values to fitness, the ISS. The fundamental properties of this curve (slope and curvature) are intimately related to the changes in trait mean and variance that are induced by selection within a generation. For example, a steep selection curve dramatically shifts the mean. A \cap -shaped curve reduces trait variance within a generation. Although we generally will not know the shape of the ISS, we can deduce its properties and approximate its shape with various functions. To make predictions about how the trait will change from one generation to the next, we need a related curve, one that is averaged over the trait distribution. This curve is called the adaptive landscape (AL).

3.1 The individual selection surface, ISS.

The idea that selection is some kind function is implicit in the names used to describe selection; e.g., directional, stabilizing, truncation, etc. In this chapter we will make this idea explicit in a way that helps us visualize selection in its many guises. Imagine selection described by some continuous function. In particular, imagine relative fitness of individuals of phenotype z as a function of trait values, z . An overall positive slope implies directional selection favoring higher values (Fig. 3.0a), while an overall negative slope implies directional selection for smaller values. Downward curvature that straddles the phenotypic mean implies stabilizing selection in the sense that selection encourages the mean to reside in a specified range (Fig. 3.0b, c). Upward curvature has the opposite effect, the sign of disruptive selection

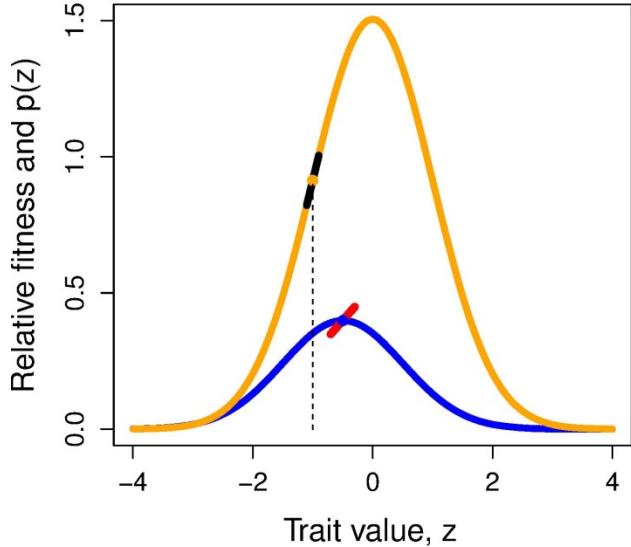


(Fig. 3.0d) . We will refer to the function in question as $w(z)$ and call it the *individual selection surface*, the ISS.

Figure 3.0 Examples of individual selection surfaces, ISSs, showing their effects on trait distributions. The statistics of the trait distributions before selection (blue), $p(z)$, and after (black) selection, $p(z)^*$, are given in Fig. 1.1. The ISS, $w(z)$, is shown as an orange curve. (a) Directional selection. The ISS is a linear function, $w(z) = \alpha + \beta z$, $\alpha = 1$, $\beta = 0.2$. The orange line has been shifted down 0.3 units for graphic effect. (b) Stabilizing selection. The ISS is a Gaussian function with $\theta = 0$, $\omega = 4$. (c) Directional and stabilizing selection. The ISS is a Gaussian function with $\theta = 1$, $\omega = 4$. (d) Disruptive selection. The ISS is a Gaussian function with $\theta = 0$, $\omega = -4$. The orange curve has been shifted down 0.8 units for graphic effect.

Our concern with the ISS is local in the sense that we will focus on its shape in the region of trait values in which individuals are likely to be observed within our focal population. For example, consider the issue of the average slope of the ISS. To calculate that average slope, we would need to take the slope or first derivative of the function at a particular value of z , $\partial w(z)/\partial z$ (Fig. 3.1), weight that slope by

Figure 3.1. The directional selection gradient, β , as the weighted average of first derivatives of the ISS. The ISS is the orange curve; the trait distribution before selection, $p(z)$, is shown in blue. The first derivatives of the ISS for $z=-1$ is shown as the slope of the black, straight line segment superimposed on the ISS. The average of all such slopes, weighted by $p(z)$, is the directional selection gradient, β , shown as the slope of the red line segment. In this illustration, the ISS is a Gaussian function ($\theta = 0$, $\omega = 1$) and $p(z)$ is a normal distribution ($\bar{z} = -0.5$, $P = 1$).



the frequency of individuals at that point, $p(z)$, repeat these operations over the entire range of occupied values of z , and then add up all those weighted slopes, $p(z)\partial w(z)/\partial z$. This average slope turns out to be a familiar commodity, the directional selection gradient,

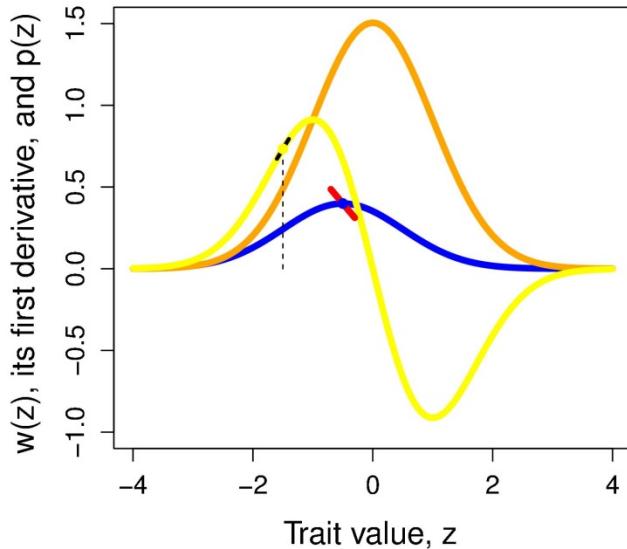
$$\beta = \int p(z) \frac{\partial w(z)}{\partial z} dz \quad (3.00)$$

(Lande & Arnold 1983). By a similar set of operations, one can evaluate the average curvature of the ISS, $\partial^2 w(z) / \partial z^2$ at each point, and calculate the average of those curvatures (Fig. 3.2). This average is equivalent to the nonlinear selection gradient,

$$\gamma = \int p(z) \frac{\partial^2 w(z)}{\partial z^2} dz \quad (3.01)$$

Surprisingly, these equivalencies hold whatever the form of the ISS so long as it is continuous and differentiable, and the trait is normally distributed before selection (Lande & Arnold 1983).

Figure 3.2. The nonlinear selection gradient, γ , as the weighted average of second derivatives of the ISS. The ISS is the orange curve, $w(z)$; the trait distribution before selection, $p(z)$, is shown in blue. The first derivative of the ISS is the yellow curve. A first derivative of this yellow curve (which is the second derivative of the ISS) at $z=-1.5$ is shown as a black line segment superimposed on the yellow curve. The average of all such slopes, weighted by the trait distribution, is the nonlinear selection gradient, $\gamma = -0.4375$, shown as the slope of the red line segment. In this illustration, the ISS is a Gaussian function ($\theta = 0, \omega = 1$) and $p(z)$ is a normal distribution ($\bar{z} = -0.5, P = 1$).



3.2 Linear and quadratic approximations to the ISS.

Unless we are prophets, as we pretended to be in Figd. 3.1 and 3.2, the ISS is not revealed to us directly, so we must deduce the properties of the ISS from data on relative fitness, $w(z)$, as a function of trait values, z . Suppose, for example, instead of an ISS revelation, we have instead the data plot shown in Fig. 3.3. We can bypass the problem of the actual shape of the ISS and try to estimate the selection gradients, β and γ . It turns out that these gradients can be estimated by the familiar statistical procedures of linear and quadratic regression (Lande & Arnold 1983). Before fitting the regressions, it is useful to standardize the trait before selection so that it has a zero mean and a standard deviation of one. Fitness should be standardized in a different way so that it has a mean of one. These standardizations yield selection gradients in readily interpretable units. Having accomplished these standardizations , we first estimate β by fitting a linear approximation to the ISS,

$$w(z) = \alpha + \beta z + \varepsilon, \quad (3.02)$$

ε is the deviation of a particular observation from the regression line (Fig. 3.3a). The usual assumption in statistical inference is that ε is normally distributed with a mean of 0. To estimate γ , we need to use a second-order polynomial function, a quadratic function, to approximate the ISS,

$$w(z) = \alpha + \beta z + \frac{1}{2} \gamma z^2 + \varepsilon \quad (3.03)$$

(Fig. 3.3b). The factor of $\frac{1}{2}$ is included so that γ will be the second derivative of the function, a measure of curvature.

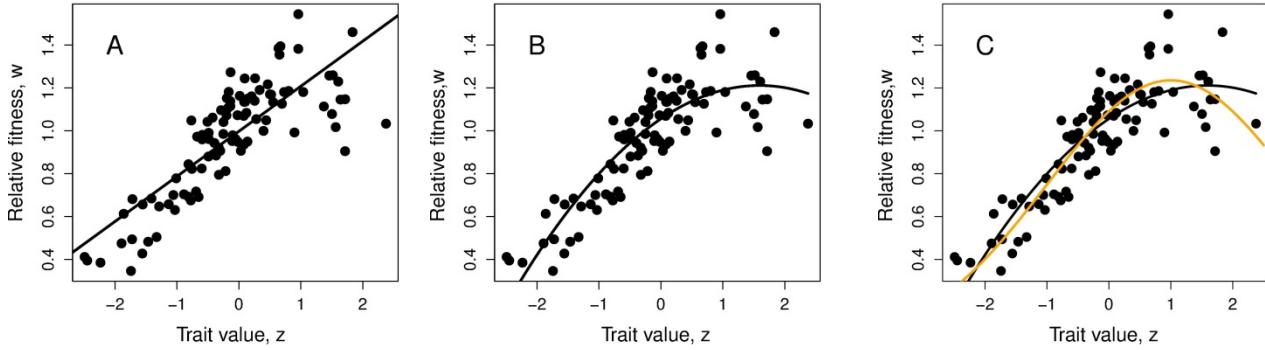


Figure 3.3. Linear and quadratic approximations to the ISS in a hypothetical example. (a) A linear fit to the data (Fig. 1.x) using (3.02), $\alpha = 1$, $\beta = 0.21 \pm 0.02$ s.e. (b) A quadratic fit using (3.03), $\alpha = 1$, $\beta = 0.20 \pm 0.01$ s.e., $\gamma = -0.12 \pm 0.02$ s.e. (c) The actual ISS is a Gaussian function (3.07), shown in orange, $\theta = 1$, $\omega = 4$. The data points were generated by taking a random sample of trait values, z , from a normal distribution ($\bar{z} = 0$, $P=1$). Those trait values were used in conjunction with the specified Gaussian function to produce corresponding, expected values of w . Values of ε were drawn from a normal distribution (mean = 0, variance = 0.01) and added to expected value of w to produce the points in the figure ($n=100$).

A few words about the actual steps in performing these two regressions may be useful. For the standardization of trait values before selection, one simply subtracts the raw trait mean from each value and divides each of those new values by the raw trait standard deviation, producing a desired new trait with zero mean and standard deviation of one. The raw values of fitness are divided by the raw fitness mean to produce a new fitness variable (relative fitness) with a mean of one. If one wished to draw or express the two regressions, (3.02) or (3.03), one ignores the estimated value of α and sets its value to one. Why is this? The value of ε on the regression curve is zero, so it is dropped from the equation. Finally, to draw or express (3.03), one takes the value of β estimated from (3.02) and the value of γ estimated from (3.03). This last point may seem mysterious, but it is a practical solution to a bias problem. In general, if the trait distribution is not perfectly normal before selection, the standardized trait values, z , will be correlated with their squared values, z^2 . This correlation will bias the estimate of β obtained from (3.03). A simple solution to this problem is to use the value of β estimated from the linear regression (3.02), which will produce an unbiased estimate (Lande & Arnold 1983).

Neither of these approximations, (3.02) and (3.03), is meant to actually imitate asymmetry or the bumps and grinds that might be present in the ISS. Linear and quadratic functions are used because they enable us to estimate parameters of selection. For example, in the hypothetical example shown in Fig. 3.1, the ISS was actually a Gaussian function (Fig. 3.3c), which is only roughly approximated by a quadratic function. Despite the relatively poor fits of the linear and quadratic functions to this ISS, those fits do provide correct estimates of β and γ . Nevertheless, the actual shape of the ISS is also of interest. A variety of procedures and functions might be used to more accurately describe the shape of the ISS. We will consider one of these in the next section.

In passing, we note that in the univariate case, selection gradients and selection differentials are equivalent when fitness and trait values are standardized. Under that standardization scheme, relative fitness has a mean of one, and the trait values have a mean of one, with $P=1$ (section 1.4). With this standardization, we see from (2.2 and 2.9) that $\beta=s$ and $\gamma=C$, but these relationships will not hold in the multivariate case.

3.3 Cubic spline approximation to the ISS.

The power of the quadratic approximations just described is that we can use them to estimate the selection coefficients β and γ . It is a remarkable fact that this estimation is legitimate and accurate even if the ISS is not quadratic. In some situations, however, quadratic approximation may lead us astray even though it does its job of estimating β and γ . One such case is shown in Fig. 3.4. In this hypothetical example, truncation selection acts on the population. The actual shape of the ISS is step function with a step at a phenotypic value of +1. Quadratic regression will do a poor job of approximating the ISS for it will yield a disruptive curve with an inflection point near -1. In general, when the shape of the ISS is of interest, regression techniques more sophisticated than quadratic regression can be used to approximate its shape. Schlüter (1988) introduced the use of regression with cubic splines to describe fitness functions, and many subsequent studies have successfully used this approach to approximate the ISS.

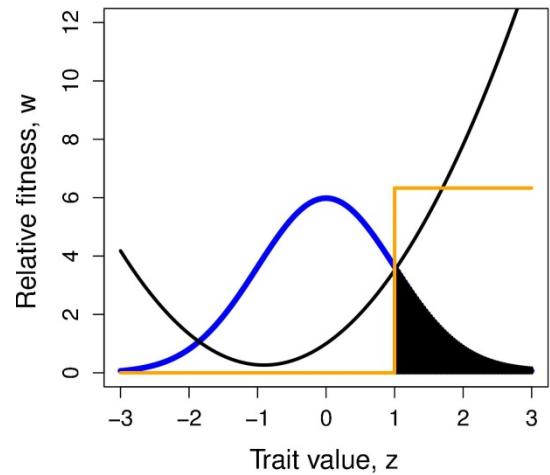
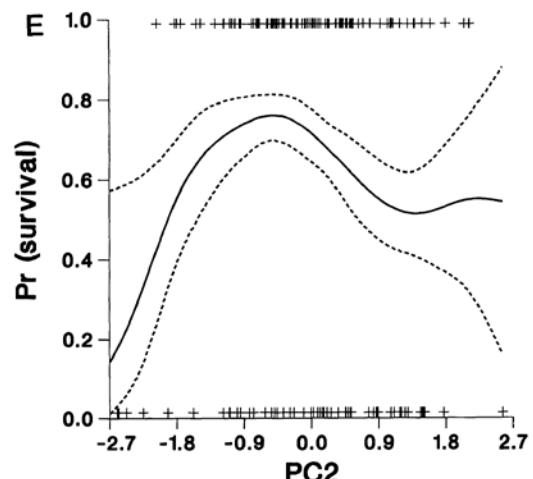


Figure 3.4. A hypothetical example in which a quadratic approximation to the ISS is misleading in shape. The sample before selection is shown as a blue curve, a normal distribution with a mean of zero and a standard deviation of 1. The actual ISS, which exerts truncation selection, is shown in orange (estimated using 3.03). The sample after selection is shown in black. The quadratic approximation to the ISS is shown as a black curve. This curve was estimated by fitting a quadratic regression to a sample of 1000 individuals drawn from the normal distribution before selection and assigned relative fitnesses according to the truncation-function ISS. Although the black curve gives a poor idea of the ISS, it does give a good estimate of the selection gradients ($\beta = 1.62 \pm 0.04$ s.e., $\gamma = 1.79 \pm 0.07$ s.e.) (after Schlüter 1988).

An example of a cubic spline approximation to an unknown ISS is shown in Figure 3.5. Here survivorship of song sparrows (*Melospiza melodia*) is shown as a function of a linear combination of morphological measurements (PC2). As in the preceding example, fitness is binary. Measurements in the sample after selection are clustered near the middle of the range in trait values. The cubic spline approximation appropriately shows a convex function with an inflection near the trait mean (stabilizing selection). The minor curves in the fitted function were not anticipated by the authors and are probably of no interest to readers. In this sense, a simple quadratic approximation, which would be a smooth convex function, would do just as well, while estimating β and γ . In some situations, however, minor curves or non-quadratic shapes will have biological significance.

Figure 3.5. An example of a cubic spline approximation to an unknown ISS. The solid black curve is a cubic spline of overwinter survival of male song sparrows ($n=152$) as a function of morphological measurements (PC2). The dashed curves show ± 1 standard error of the



function, estimated by bootstrapping. The crosses at the top and bottom show the absolute fitness and trait values of males that did and did not survive winters. From Schluter 1988.

3.3 The adaptive landscape, AL.

Another fitness function, distinct from the ISS, will be important to us. This important function is known as the *adaptive landscape*, the AL. In the AL, average fitness of the population, \bar{W} , is a function of its average trait value, \bar{z} . For our focal population, we can evaluate the AL at a single point, the trait mean, \bar{z} . That narrow perspective on the AL, however, is sufficient to tell us the slope and curvature of the AL at that point. In particular,

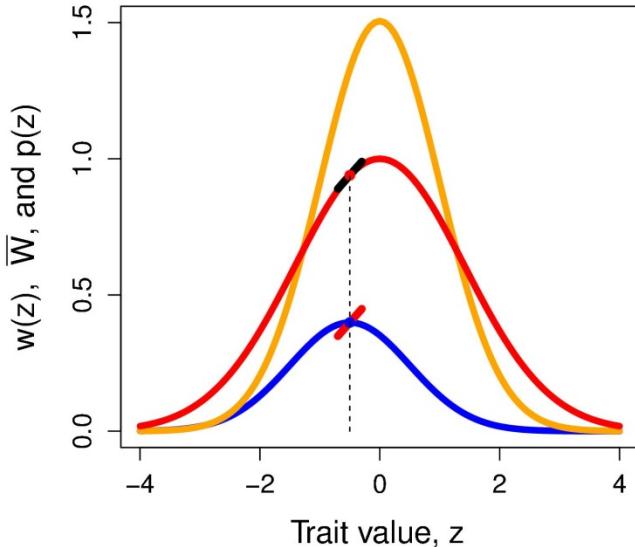
$$\beta = \frac{\partial \bar{W}}{\bar{W} \partial \bar{z}} = \frac{\partial \ln \bar{W}}{\partial \bar{z}} \quad (3.04)$$

and

$$\gamma - \beta^2 = \frac{\partial^2 \bar{W}}{\bar{W} \partial \bar{z}^2} = \frac{\partial^2 \ln \bar{W}}{\partial \bar{z}^2}, \quad (3.05)$$

where $\ln \bar{W}$ means the natural logarithm of average absolute fitness in the population (Lande 1979, Arnold & Lande 1983). The directional selection gradient, β , as the slope of the AL evaluated at the trait mean is portrayed in Fig. 3.6.

Figure 3.6. The directional selection gradient, β , as the first derivative of the AL, evaluated at the trait mean before selection, $\bar{z} = -0.5$. The AL is shown as a red Gaussian curve with an optimum $\theta = 0$ and width parameter $(\omega + P) = 2$. The directional selection gradient is shown at two sites connected by a dashed vertical line: at the mean of $p(z)$, where it is shown as a red line segment, and at the point of evaluation on the AL, where it is shown as a black line segment. Other conventions as in Fig. 3.1.



Just as we can imagine that the ISS is a function that extends beyond the occupied region of phenotypes for our population, the same might be so for the the AL. We also need to imagine that hypothetical probability distributions for z , so that we can average the ISS over those hypothetical distributions to visualize the correponding AL. An easy way to generate these hypothetical distributions of z is to keep the variance, P , constant and simply shift the mean before selection, \bar{z} . The extension is especially simple if we assume that $p(z)$ is a normal distribution, and the ISS is a Gaussian function (Haldane 1954, Lande 1976) so that

$$p(z) = (\sqrt{2\pi P})^{-1} \exp\left\{-\frac{(z - \bar{z})^2}{2P}\right\} \quad (3.06)$$

and

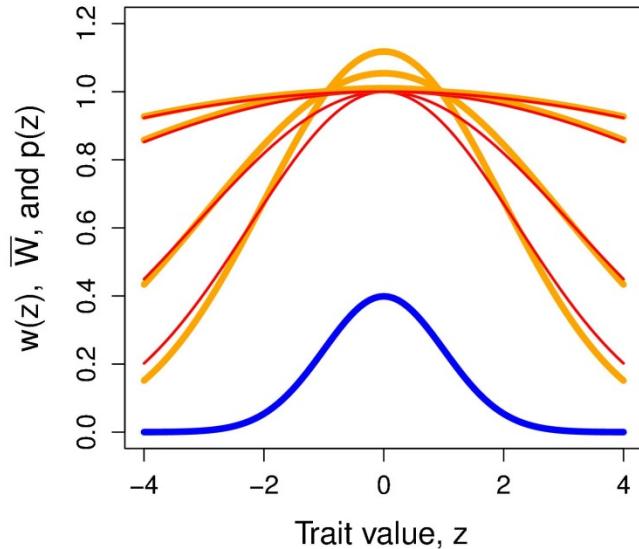
$$w(z) = \exp\left\{-\frac{(z-\theta)^2}{2\omega}\right\}, \quad (3.07)$$

where θ is the optimum and $\sqrt{\omega}$ is the width of the function, analogous to a standard deviation (ω is analogous to a variance) (Fig. 3.0e). If $p(z)$ keeps its shape (constant variance) while translating it mean, averaging the ISS over this translated function yields a Gaussian-shaped AL,

$$\bar{W} \propto \exp\left\{-\frac{(\bar{z}-\theta)^2}{2(\omega+P)}\right\}, \quad (3.08)$$

with the same optimum as the ISS, θ , but a larger ‘variance’, $\omega + P$ (Fig. 3.65). The first derivative or

Figure 3.65 Gaussian ISSs and their corresponding Gaussian ALs for a range of values of ω . The blue curve shows a normal trait distribution before selection, $p(z)$, with a mean of 0 and a variance, $P=1$. The wide orange curves show the individual selection surfaces for $\omega = 99$ (at the top), 49, 9, and 4 (at the bottom). The narrow red curves show the corresponding adaptive landscapes with width parameters $\omega+P = 100, 50, 10$, and 5.



slope of the AL, evaluated at the population mean, is

$$\frac{\partial \ln \bar{W}}{\partial \bar{z}} = (\omega + P)^{-1}(\theta - \bar{z}) \quad (3.09a)$$

and its curvature is

$$\frac{\partial^2 \ln \bar{W}}{\partial \bar{z}^2} = -(\omega + P)^{-1} = \gamma - \beta^2 \quad (3.09b)$$

(Lande 1979, Phillips & Arnold 1989, Jones et al. 2004). When the trait mean is at the optimum and $\omega \gg P$,

Lande (pers. comm.) showed that

$$\gamma = -\frac{1}{\omega} \left(\frac{\omega}{\omega + P} \right)^{\frac{3}{2}}.$$

A useful approximation for converting between γ and ω under Gaussian assumptions is

$$\gamma \approx -\frac{1}{\omega}, \quad (3.10)$$

An expression that works best when γ is close to zero or with a trait value close to the optimum when $\gamma > -1$ (Fig 3.7a).

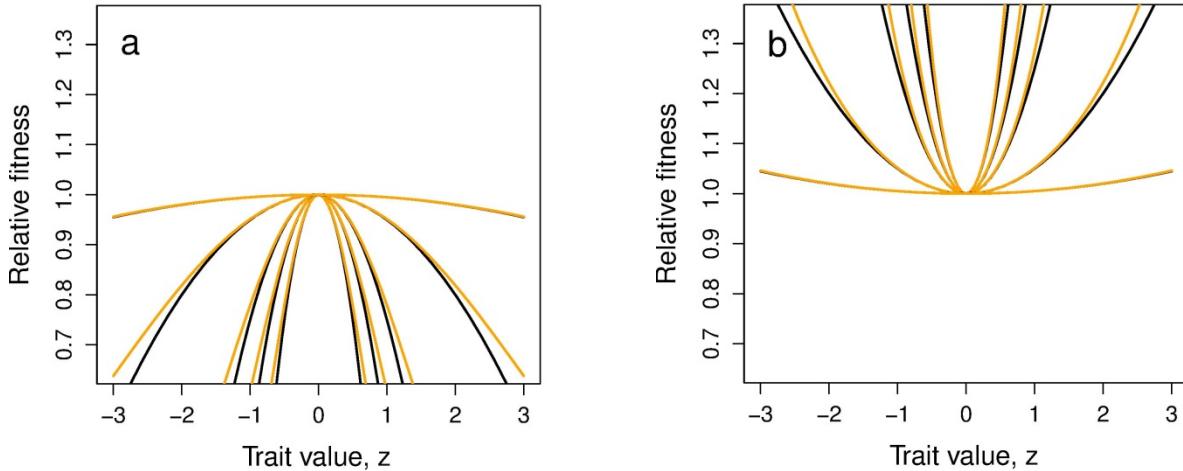


Figure 3.7. Gaussian approximations to quadratic individual selection surfaces (ISSs). Quadratic ISSs are shown as black curves and their Gaussian approximations are shown as orange curves. Expression 3.10 was used to approximate ω from γ . Optimum or pessimum of quadratic surface set at zero. When ω is 0.5 or -0.5 the approximation is so close that the quadratic curves are barely visible. (a) Gaussian approximations of stabilizing quadratic selection curves. From the top curve to the lowest, $\gamma(\omega)$ values are -0.01(100), -0.1(10), -0.5(2), -1(1), and -2 (0.5). (b) Gaussian approximations of disruptive quadratic selection curves. From the lowest curve to top, $\gamma(\omega)$ values are 0.01(-100), 0.1(-10), 0.5(-2), 1(-1), and 2 (-0.5).

In general, by taking the second derivative of (3.03) and setting it equal to zero, we find that the distance to the optimum (or pessimum) from the trait mean is

$$(\bar{z} - \theta) = \beta / -\gamma \quad (3.11)$$

(Mitchell-Olds & Shaw 1987, Phillips & Arnold 1989).

The Gaussian form of the ISS (3.07) is also useful in theoretical work because it allows us to immediately solve for the mean and variance after selection. If a trait is normally-distributed before

selection and subjected to Gaussian selection (3.07), it will be normally-distributed after selection with mean and variance,

$$\bar{z}^* = (\bar{z}\omega + \theta P) / (\omega + P) \quad (3.12a)$$

and

$$P^* = \omega P / (\omega + P) \quad (3.12b)$$

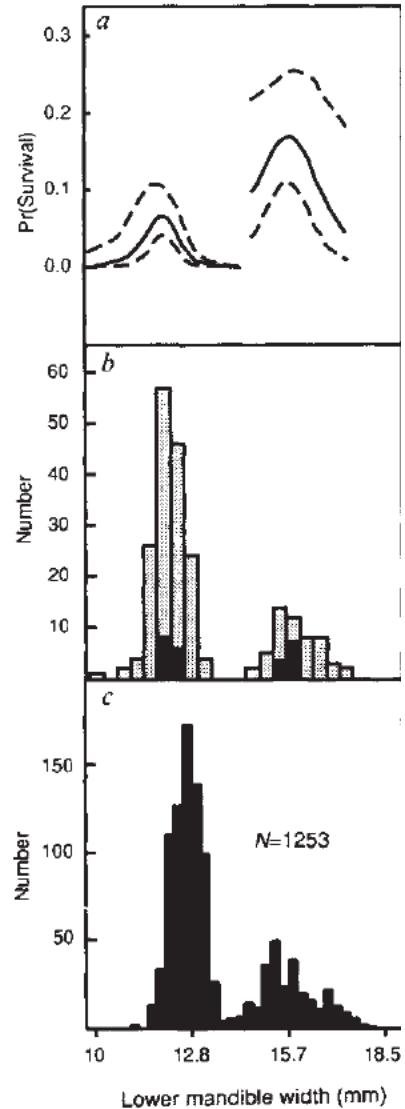
Lande (1981).

3.4 Empirical approximations to the ISS

Fig. 3.5 = Univariate examples of linear, quadratic and cubic spline approximations to ISS, do this with the Galapagos finch data?

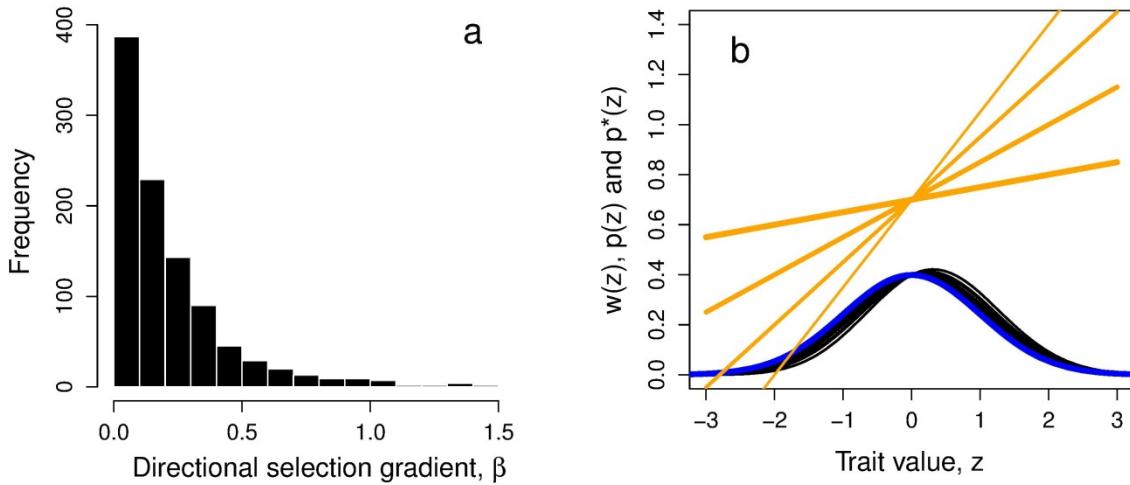
.... empirical focus on β , neglect of γ ... related to the sample size problem, exacerbated by including many traits ...

Figure 3.75. Disruptive selection on bill size in the African finch *Pyrenestes ostrinus*. (a) Cubic spline approximation to the ISS fitted separately for small and large billed morphs. Confidence limits (dashed lines) estimated by bootstrapping. (b) Trait distribution of juveniles that did not survive (shaded) and that did survive (black). (c) Trait distribution in adults. From Smith (1993).



3.5 Surveys of univariate selection surfaces and adaptive landscapes

Here we focus on the shapes of ISSs that emerge from the two surveys summarized in Chapter 2, Kingsolver et al. 2001 and Stinchcombe et al. 2008. The most basic result is that the most frequent result from empirical studies is an almost flat curve with almost no slope and no curvature. We reach this conclusion by noting that the most frequent estimate of β is very close to zero (Fig. 3.8a). This fact is poorly reflected in Fig 3.8b, in which the widest ISS is for a bin category centered at $\beta=0.05$. Also note that the histogram in Fig. 3.8a is for the absolute value of β . If we had plotted the estimated values, in which the signs are arbitrary functions of measurement scale, the plot would be approximately symmetrical about zero. Turning to γ , we see such an approximately symmetrical histogram in Figure 3.8c, which implies that the most frequent γ estimate is very close to zero, which implies a nearly flat curve that is bowed slightly upward or downward. Curves corresponding to four bin categories ($\gamma = 0.3, 0.1, -0.1$, and -0.3) are shown in Fig. 3.8d. In other words, the most frequent result in empirical studies is to find that the individual selection surface is nearly flat. It is unusual to estimate a slope steeper than the thinnest orange line in Figure 3.8b or more curved than the thinnest curves in Figure 3.8d.



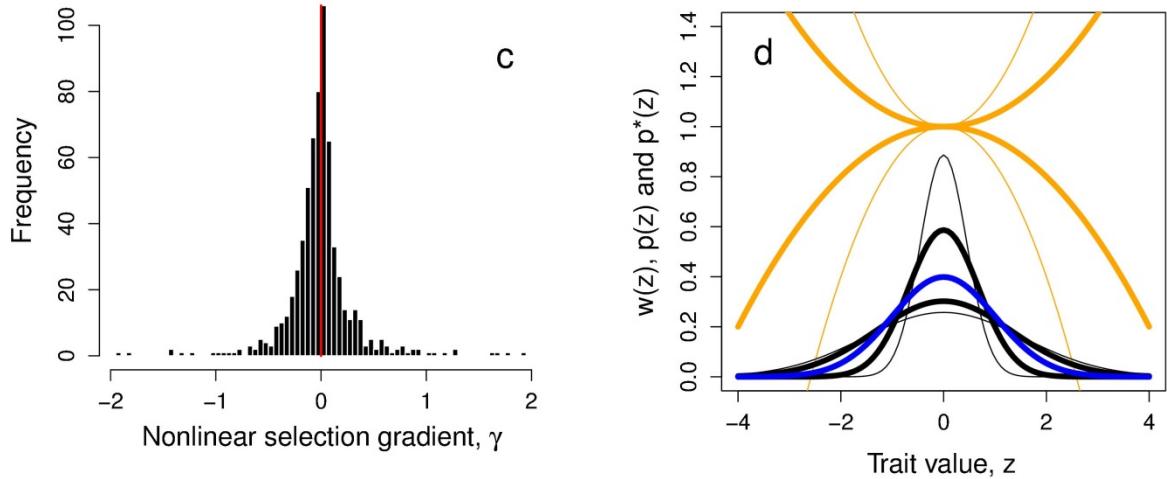


Figure. 3.8 Histograms of selection gradients paired with frequency distributions that portray corresponding selection surfaces and the magnitude of effects on means and variances. Trait distributions before and after selection are shown in blue and black, respectively. (a) Histogram of the absolute values of directional selection gradient estimates, β ; 992 values from Kingsolver et al 2001. (b) Shifts in mean corresponding to the directional selection gradients in Fig. 3.8a. Shifts corresponding to the four bins on the right-most side of the distribution (0.05-0.45) are shown as black curves whose widths correspond to bin frequency. Orange curves show the fitness functions (ISSs) corresponding to those 4 bin categories, with widths corresponding to bin frequency. (c) Histogram of nonlinear selection gradient estimates, γ ; 653 values from the Stinchcombe et al 2008 database, after deleting 7 values of $\gamma > 2$ and 4 values < -2 . (d) Shifts in variance corresponding to the nonlinear selection gradients shown in Fig. 3.8c . Shifts corresponding to the 16 most populated bins in the center the γ distribution (-0.4 to 0.4) are shown as four black curves (accounting for 86% of the observations) with width representing four bin category frequency. Orange curves show the fitness functions (ISSs) corresponding to those 4 bin categories, with widths corresponding to bin category frequency.

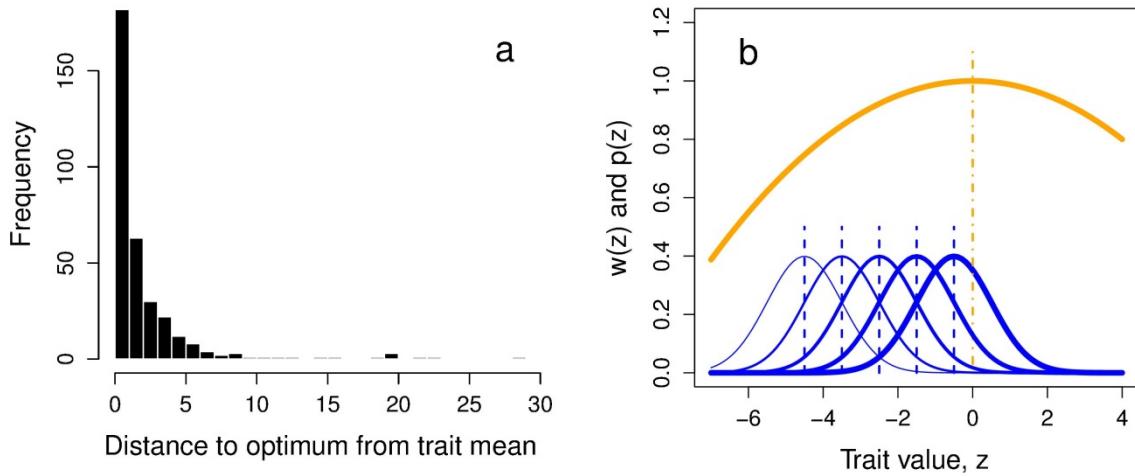


Fig. 3.9 Histogram of distances of the phenotypic mean from the intermediate optimum, θ , in natural populations, with blue curves portraying the most frequent values of those distances. (a) Values were calculated using (3.11) with paired values of negative γ and absolute values of β from the Stinchcombe et al. (2008) database ($n = 339$, dropping two values greater than 50). Distance to the optimum is measured in units of within-population phenotypic standard deviation before selection. The median value is 0.906. (b) The orange curve shows the most frequency value of γ with a negative value in the histogram shown in Fig. 3.8c (bin centered at $\gamma = -0.025$ with a count of 80). The optimum, θ , is shown with a vertical dash-dot line. The blue curves show means of $p(z)$ at distances of -0.5, -1.5, -2.5, -3.5 and -4.5 from θ , corresponding to the first 5 bins in the histogram with relative frequencies portrayed with line widths. The positions of trait means are shown with vertical dashed lines.

We can also calculate the distances to the optimum of the ISS (which are equivalent to distances from the optimum of the AL), assuming a convex ISS, using (3.11). Taking the median values of those distances, we find that most trait means are within one trait phenotypic standard deviation of the optimum and that it is unusual for the trait mean to be more than 5 standard deviations from the optimum (Fig. 3.9a).

3.6 The ecology of selection

The simplicity of the function describing stabilizing selection (probably the commonest mode of selection) gives conceptual power but it should not hide the fact that many ecological factors and features of biological design are responsible for the particular location of the optimum and

downward curvature at the two ends of the function (Travis 1989). If we consider the length of the tail in a hypothetical bird, for example, it seems likely that a tail that is too short fails to provide aerodynamic lift at take off, while a tail that is too long interferes with maneuverability and landing. In other words, different kinds of performance and selective agents are likely to cause downward curvature at the two ends of the ISS. The ISS is complex in the sense that it summarizes the interaction of a set of traits with an environment of selective agents and contexts.

3.7 Technical issues in estimating and interpreting selection surfaces

No one mode of selection analysis is likely to satisfy all the expectations of the investigator or reader. Selection gradients, for example, provide useful measures of the intensity of selection in the same form that appears in response to selection equations, but they may fail to describe the actual shape of the ISS. Consequently, a quadratic approximation (section 3.2) to the ISS might be used to estimate β and γ , but a cubic spline (section 3.3) or some other averaging function may be needed to accurately describe the shape of the ISS.

The ISS is a multivariate beast whose overall appearance cannot generally be appreciated by viewing it one trait at a time. Even in the simple case in which only two traits are under selection, each of the two univariate views of selection may be misleading. Plotting fitness as a function of a shape trait may show a flat trend line because strong selection for increasing values of shape prevails in small individuals, while strong selection for decreasing values of shape prevails in large individuals. In this case, averaging selection across one trait (shape) masks the true mode of selection acting on another trait (size). In general, we must be wary that a univariate view or coefficient accurately describes selection. Usually it will not. In the next chapter we will consider truly multivariate visions of the selection beast.

Chapter 4: The Selection Surface and Adaptive Landscape for Multiple Traits

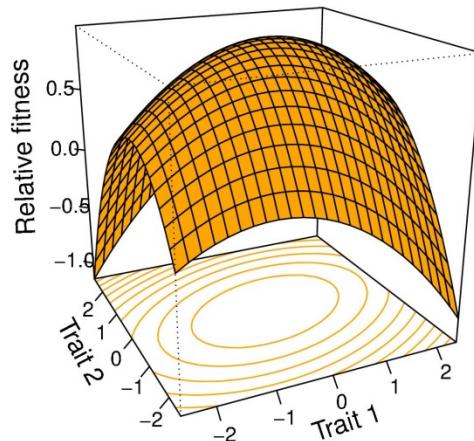
© Stevan J. Arnold 2016

Overview.- Selection on multiple traits can be visualized as a surface that relates fitness to the values of two or more traits in individuals. Such a surface can help us visualize the direct effects of selection on trait distributions. Indeed, under some conditions, this surface can take account of the indirect effects of selection induced by trait correlations. In the case of two traits, analysis of selection in natural populations often reveals a hill- or ridge-shaped surface that depicts bivariate stabilizing and correlational selection. The corresponding adaptive landscape has a similar shape but with less curvature.

Natural selection acts simultaneously on many traits. An important consequence of this fact of nature is that we need a multivariate conceptualization of selection to deal with selection and evolution in the natural world. Textbooks in evolutionary biology have been slow to embrace these multivariate inevitabilities. Instead, depictions of selection remain locked on a vision of selection in which selection acts on single traits and is purely directional, a vision developed in the 1940-1960 world of plant and animal breeding. That world included the origin of selection indicies, a multivariate tool for handling directional selection on multiple traits (Hazel 1944), but the conceptualization of multivariate stabilizing selection is a relatively recent development (Lande & Arnold 1983, Phillips & Arnold 1989).

Many important features of the effects of multivariate selection on trait distributions can be appreciated by considering bivariate selection that is stabilizing. The selection surface corresponding to this form of selection is convex (Fig. 4.0) with the optimum situated near the bivariate mean of the trait distribution before selection. Fitness falls off in all directions away from the optimum.

Figure 4.0 Bivariate stabilizing selection represented as a convex selection surface. Relative fitness, $w(z)$, corresponds to points on the surface as a function of the values of two traits, z_1 and z_2 . A contour representation of the surface is projected onto the $z_1 \times z_2$ plane with contours at increments of 0.2 in relative fitness. This surface is quadratic with $\alpha = 1$, $\beta_1 = 0.22$, $\beta_2 = -0.08$, $\gamma_{11} = -0.31$, $\gamma_{22} = -0.18$, and $\gamma_{12} = 0.07$.

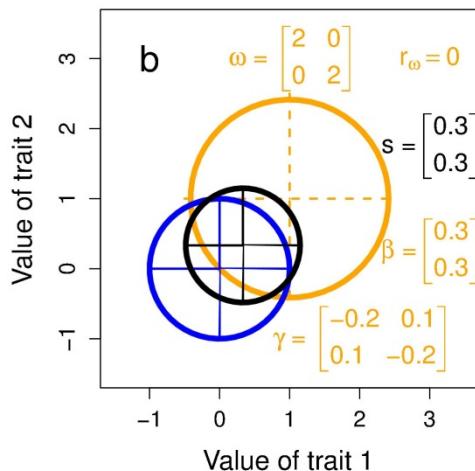
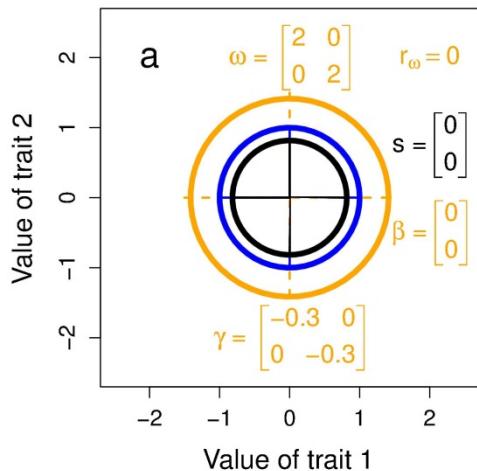
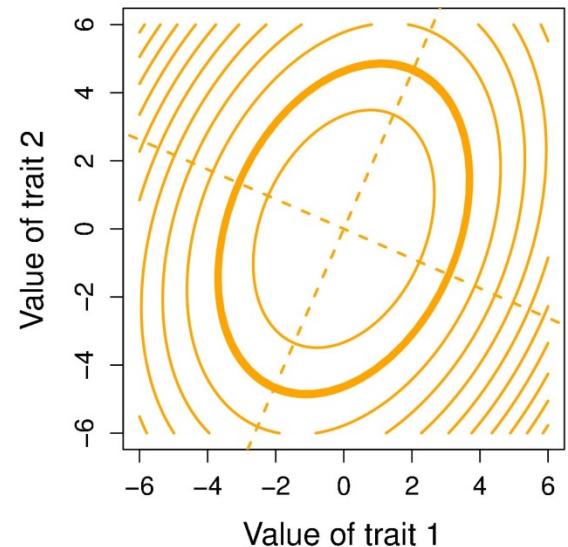


Another convention that will prove useful later is the representation of bivariate stabilizing selection with a single contour line and its eigenvectors. In Fig. 4.1 we show this representation for an inclined surface that exerts strong correlational selection. Consider first a familiar bivariate normal distribution of phenotypes, $p(z)$. The 95% confidence ellipse for the bivariate mean is 1.96 phenotypic standard deviations from the mean along each of its principal axes or eigenvectors. The eigenvalues corresponding to each eigenvector are analogous to variances. Likewise, a selection surface has principal axes (section 2.55), shown in Fig. 4.1 as dashed orange lines. The eigenvalues (λ) for this surface are also analogous to variances. Consequently an ellipse analogous to the 95% confidence ellipse for $p(z)$, is situated $1.96\sqrt{\lambda_i}$ along each

of the two eigenvectors, $i = 1$ and 2 , from the stationary point (where the eigenvectors intersect). This confidence ellipse is shown in Fig. 4.1 as a bold orange ellipse. The smaller this ellipse, the stronger the stabilizing selection. The further the optimum of the surface from the bivariate mean, the stronger the directional aspect of selection.

Figure 4.1 Contour and confidence ellipse portrayal of a selection surface. The optimum of the surface is at $z_1=0$ and $z_2=0$. The thin orange ellipses represent equal values of relative fitness at increments of 0.2. The bold orange ellipse represents the analog to a 95% confidence ellipse. See text for more details.

Other examples of stabilizing selection surfaces are shown in Fig. 4.2 using the confidence ellipse portrayal. Such convex surfaces can be circular (Fig. 4.1a,b), elliptical with a positive (Fig. 4.1 c, d) or a negative inclination (Fig. 4.1e, f). Many other configurations are possible as well. When the optima of these concave surfaces are located away from the bivariate mean before selection, directional selection is imposed and the bivariate mean is pulled in that direction (Fig. 4.0b, c, f). In the rest of this chapter we will elaborate on these concepts of multivariate selection and make them more precise. In the process we will see that the selection depicted in Fig. 4.0 is substantially stronger than we expect to see in nature.



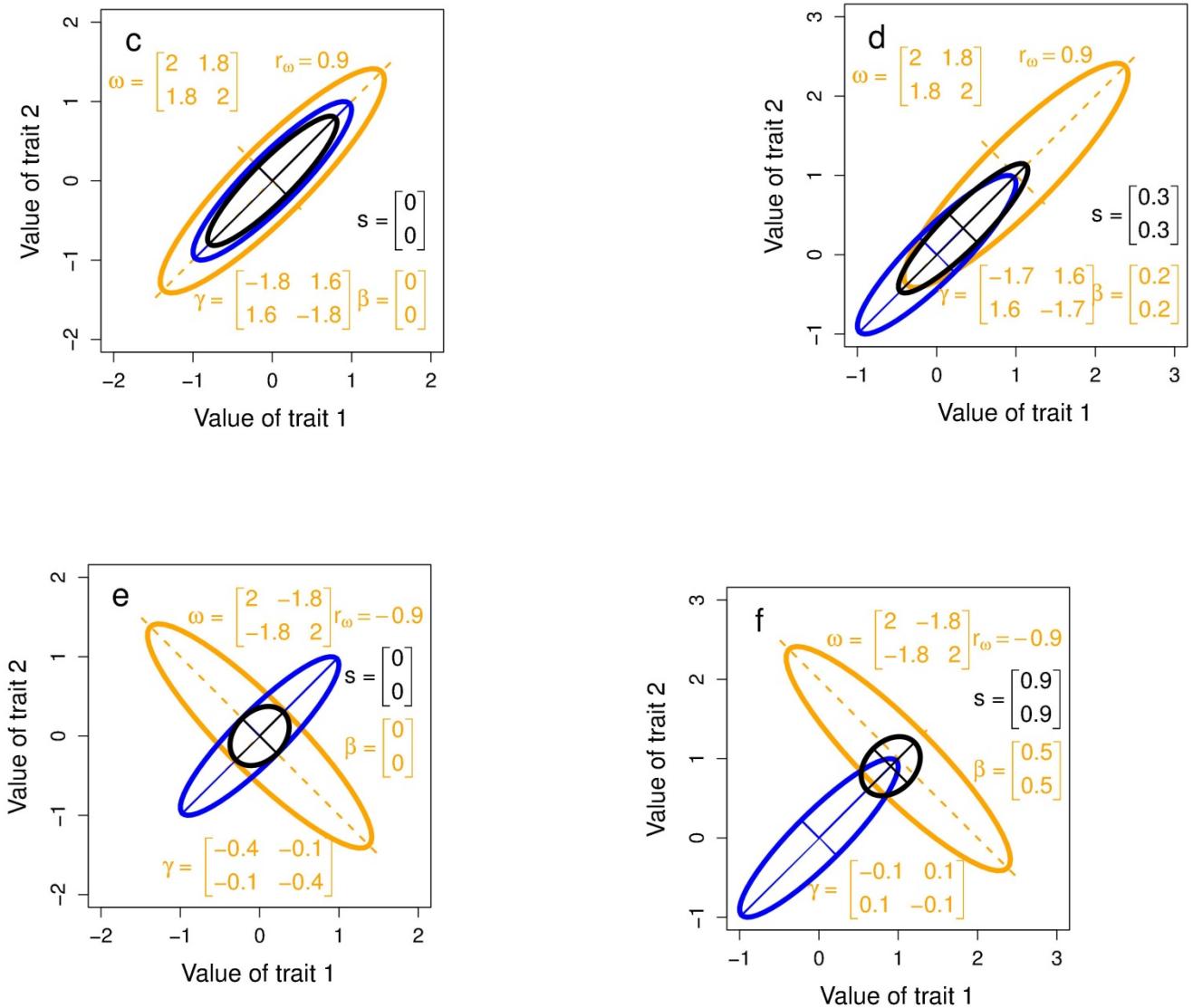


Figure 4.2. Various forms of Gaussian selection and their effects on bivariate trait distribution. Ellipses comparable to 95% confidence ellipses are shown for Gaussian ISSs (orange). Phenotypic trait distributions are bivariate normal before (blue 95% confidence ellipses) and after selection (black 95% confidence ellipses). Bivariate means, variances and covariances, before and after selection, are given in the caption for Fig. 2.2. The relationship of the γ -matrix to the ω -matrix will be discussed in section 4.4. In the left-hand panel, the trait distribution experiences nonlinear selection, but no directional selection. In the right-hand panel, directional selection has been added by shifting the position of the optimum, θ . (a) Symmetrical stabilizing selection contracts the trait variance without shifting the mean. (b) When the optimum is displaced from trait mean in both directions, the trait distribution after selection shifts towards the optimum. (c) Positive correlational selection is reflected in the positive inclination of the γ -ellipse. When this ellipse is aligned with the P -

ellipse (blue), the inclination of the P^* -ellipse (black) is not changed by selection. (d) Displacement of the optimum from the trait mean, combined with aligned γ - and P -matrices, shifts the trait distribution towards the optimum. (e) Nonalignment of the γ - and P -matrices causes a dramatic change in trait covariance (compare P - and P^* -ellipses). (f) Nonalignment of the γ - and P -matrices, combined with a displaced optimum, shifts the trait distribution towards the optimum.

In Figures 4.0 and 4.1, we specified selection, imposed it on trait distributions and observed the consequences within a generation. We now want to consider the problem of deducing the shape of the selection surface from actual data. Those data might take one of two forms: (1) a sample of individuals together with their trait measurements before selection and a estimate of each individual's fitness, or (2) one sample of individuals and their trait measurements before selection and a second sample, with measurements, taken after selection. In either case, we want to deduce the properties of the ISS from the data. The properties we have in mind are the average slopes and curvatures of the ISS, properties that we wish to mathematically connect to actual observations. We will need to begin by defining the properties of multivariate slope and curvature

4.1 Key properties of the individual selection surface, ISS, for multiple traits.

In general, we want to think of the relative fitness of individuals, $w(z)$, as a continuous function or surface of the values of two or more traits, z . This surface might have a complicated shape, but we will be concerned only with simple quadratic shapes ... (possible new figure = use as an example a complex surface with just one peak but with a few bumps, perhaps a sum of two or three Gaussian surfaces. Such an example will make clear the distinction between the actual ISS and its quadratic approximation). If we consider a particular point on a 2-trait version of this surface (Fig 4.0), the slope at that point is a vector,

$$\begin{bmatrix} \partial w(z)/\partial z_1 \\ \partial w(z)/\partial z_2 \end{bmatrix}, \quad (4.0a)$$

and the curvature at that point is described by a matrix,

$$\begin{bmatrix} \partial^2 w(z)/\partial z_1^2 & \partial^2 w(z)/\partial z_1 \partial z_2 \\ \partial^2 w(z)/\partial z_1 \partial z_2 & \partial^2 w(z)/\partial z_2^2 \end{bmatrix}. \quad (4.0b)$$

The two elements in the vector (4.0a) give the slope in each of the two trait directions, z_1 and z_2 , with a positive sign indicating that fitness increases with trait values, a negative sign indicating the opposite. The diagonal elements in the matrix (4.0b) give the curvature in those same two directions, with negative sign denoting downward (stabilizing) curvature and positive sign denoting upward (disruptive) curvature. In our illustrated example, Fig 4.0, the signs of two diagonal curvature measures is negative (downward curvature). The off-diagonal element in the matrix describes a curvature phenomenon, correlational selection, that has no analog in the univariate case. A positive sign for this element, $\partial^2 w(z)/\partial z_1 z_2$, means that the surface is titled so that it promotes a positive correlation between the two traits, while a negative sign means that the surface is tilted so that it promotes a negative correlation. In the illustrated case $\partial^2 w(z)/\partial z_1 z_2$ is positive.

As in the univariate case, the average slope and curvature of the multivariate ISS are equivalent to our familiar selection gradients, β and γ , which are now, respectively, a column vector and a matrix.

$$\beta = \int p(z) \frac{\partial w(z)}{\partial z} dz = \begin{bmatrix} \int p(z) \frac{\partial w(z)}{\partial z_1} dz \\ \int p(z) \frac{\partial w(z)}{\partial z_2} dz \end{bmatrix} \quad (4.1a)$$

and

$$\gamma = \int p(z) \frac{\partial^2 w(z)}{\partial z^2} dz = \begin{bmatrix} \int p(z) \frac{\partial^2 w(z)}{\partial z_1^2} dz & \int p(z) \frac{\partial^2 w(z)}{\partial z_1 \partial z_2} dz \\ \int p(z) \frac{\partial^2 w(z)}{\partial z_1 \partial z_2} dz & \int p(z) \frac{\partial^2 w(z)}{\partial z_2^2} dz \end{bmatrix} \quad (4.1b)$$

(Lande & Arnold 1983). These are the same averaging functions as in (3.00, 3.01), except that now the averaging is for each of two elements in β and for each of three distinct elements in γ . {address qualifiers discussed in Taylor expansion sect of Phillips & Arnold 1989, p. 1214, or in the next section}

4.2 Linear and quadratic approximations to the ISS.

The elements in the multivariate selection gradients β and γ can be estimated by approximating the ISS with linear and quadratic surfaces. (Note that in the discussions that follow, we assume that each of the traits has been standardized so that each mean is zero, and $w(z)$ has been standardized so that its mean is 1). For example, in the 2-trait case, we can estimate β by fitting a linear regression model to the data on relative fitness and trait values,

$$w(z) = \alpha + \beta^T z + \varepsilon = \alpha + \beta_1 z_1 + \beta_2 z_2 + \varepsilon. \quad (4.2)$$

The fitted surface is a plane, with $\alpha=1$ describing its elevation, β_1 and β_2 its inclination, and ε representing the departure of individual data points from the regression surface in the vertical dimension (Fig. 4.1). We can estimate the elements in γ fitting a curvilinear regression model that corresponds to a quadratic surface,

$$w(z) = \alpha + \beta^T + \frac{1}{2} z^T \gamma z + \varepsilon = \alpha + \beta_1 z_1 + \beta_2 z_2 + \frac{1}{2} \gamma_{11} z_1^2 + \frac{1}{2} \gamma_{22} z_2^2 + \gamma_{12} z_1 z_2 + \varepsilon, \quad (4.3)$$

where z_1^2 , z_2^2 , and $z_1 z_2$ are the squares and products of trait values, so-called quadratic variables. (Note that these are the same as \tilde{z}_1^2 , \tilde{z}_2^2 , and $\tilde{z}_1 \tilde{z}_2$ in sec. 2.5, but here we have deleted the tildas for simplicity.) As in the univariate case, γ_{11} and γ_{22} are *stabilizing selection gradients*, describing downward curvature (stabilizing selection) when their signs are negative and upward curvature (disruptive selection) when their signs are positive. And, as in the univariate case, the factors of $\frac{1}{2}$ are present so that the stabilizing selection gradients are second derivatives (Stinchcombe et al. 2008). A new kind of coefficient is represented by γ_{12} . This *correlational selection gradient* describes orientation of the ISS in the z_1 by z_2 dimensions, with a positive sign corresponding to an upward tilt and a negative sign corresponding to a downward tilt.

Note that (4.2) provides a statistical model that can be used to estimate the elements of β and γ . The requisite data are measures of trait values and relative fitness for each individual in a sample. Such data are *longitudinal* in the sense that ordinarily individuals must be literally or figuratively followed through time to estimate relative fitness. Fitting the quadratic regression model (4.2) with least-squares is a standard problem in multivariate statistics than can be accomplished in many statistics packages. Extracting actual estimates of β and γ , however, requires some attention to details, which are discussed below.

The rationale for treating the estimation of β as a regression problem (4.2) is very strong.

Because s is a vector of covariances between relative fitness and traits (2.2), comparison of $P^{-1}s$ with the definition of partial regression coefficients (Kendall & Stuart 1979) reveals that β is a vector of partial regression coefficients (Lande & Arnold 1983). For example, the directional selection gradient for the first of n traits, β_1 , is the partial regression of relative fitness on z_1 , holding all the other traits constant. By analogy and assuming multivariate normality of the trait distribution, γ is a matrix of partial regression coefficients, where each element represents a partial regression of relative fitness on quadratic variables, holding all the other quadratic variables constant (see eq. 2.8 and 2.9).

Since all the elements of β and γ are contained in one quadratic regression model, one is tempted to use just (4.3) to estimate all the elements in these gradients. One should resist this temptation! If the distributions of the traits are not perfectly symmetrical, as they are in the multivariate normal case, skewness in the distributions will cause the regular trait values, z_1 and z_2 , to be correlated with the quadratic variables, z_1^2 , z_2^2 , and $z_1 z_2$. These correlations will distort the estimates of β . A simple way to avoid this problem (the heartbreak of multivariate skewness) is to use (4.2) to estimate β and (4.3) to estimate γ . Another, more complicated solution involving orthogonal polynomials is discussed by Lande & Arnold (1983).

Despite the simplicity of the quadratic approximation to the ISS (4.3), it can be used to represent large variety of selection possibilities. In the 2-trait case, for example, quadratic surfaces may be hill (Fig. 4.2a), a saddle (Fig. 4.2b), a nearly level ridge (Fig. 4.2c), a rising ridge (Fig. 4.2d), as well as other possibilities. Fig. 4.2 also illustrates the point that it is generally difficult to visualize the quadratic surface that is specified by even a 2×2 matrix of γ values. All four surfaces illustrated in Fig. 4.2 have the same value and sign (negative) for their stabilizing selection coefficients, γ_{11} and γ_{22} . All of the correlational selection coefficients, γ_{12} , have the same sign (positive), but differ in magnitude from surface to surface. The position of the bivariate mean (denoted with an x) varies from surface to surface. The directional selection gradient, β , is the direction of steepest uphill slope from that position, and consequently it varies from surface to surface in Fig. 4.w. If the bivariate mean were located at a stationary point on the surface (the intersection of the two eigenvectors), both of the elements in β would be zero, i.e., there would be no directional selection.

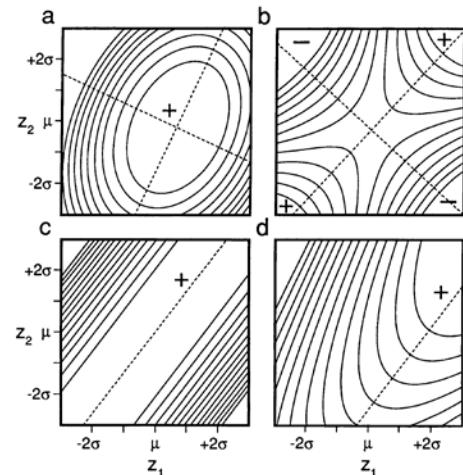


Figure 4.2. Hypothetical quadratic ISSs for two traits. Peaks are denoted with a + sign and depressions with a - sign. Dotted lines represent the canonical (principal) axes of the surface. Traits means are denoted with μ and phenotypic standard deviation with σ . Despite the differences in the appearance of these surfaces, the values of the elements in their γ -matrices are only slightly different. See Phillips & Arnold 1989 for details. From Phillips & Arnold 1989.

4.3 Canonical analysis of the quadratic approximation to the ISS

Figure 4.2 illustrates the important point that we can not deduce the shape of the fitted quadratic surface from a simple inspection of the γ -matrix. The four surfaces illustrated in Fig. 4.2 differ radically in shape yet their γ -matrices are extremely similar. One can determine the shape of the surface by plotting expression (4.2) or, more elegantly, by conducting a canonical analysis of the γ -matrix (section 2.65) that yields its eigenvalues, λ , and eigenvectors. Recalling our earlier discussion (2.65), the eigenvectors are a rotation of the original trait axes which have the property that the first eigenvector is in the direction that has the greatest curvature, the second eigenvector is in an orthogonal direction with the next greatest

curvature, and so on. The eigenvectors of the surfaces illustrated in Fig. 4.2 are shown as dotted lines. These eigenvectors and their associated eigenvalues are calculated directly from γ -matrix (Phillips & Arnold 1989). Using the γ -matrix one can also determine the *stationary point*, z_0 , on the fitted quadratic surface, which may be a fitness maximum, minimum or saddle point. The distance from the mean to this stationary point is

$$z_0 = -\gamma^{-1}\beta, \quad (4.03)$$

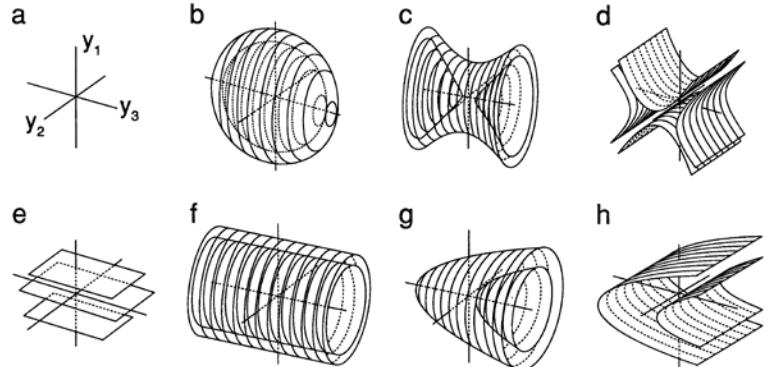
and the value of relative fitness at the stationary point is

$$w_0 = \alpha + \frac{1}{2}\beta^T z_0 \quad (4.04)$$

(Phillips & Arnold 1989). By calculating the eigenvalues of a surface, one can deduce its properties. If all the eigenvalues of γ are negative the surface is a dome (convex); if they are all positive the surface is a bowl (concave). Mixed signs indicate a saddle-shaped surface, with a stationary point at the saddle. See Phillips & Arnold (1989) for further discussion.

When the eigenvalues of γ are of mixed sign, it will often be useful to order them using their absolute values, $|\lambda_i|$, that is to say, by greatest curvature, whether concave or convex (Brooks & Blows 2003). The contrasting convention used by most computational algorithms is to rank eigenvalues by their raw values, so that all eigenvalues with positive sign receive higher rank than all eigenvalues with negative sign, irrespective of the magnitude of curvature. The distinction becomes important when we wish to determine the direction on the surface that has the least effect on relative fitness. We can call this direction γ_{min} . It is the direction given by the eigenvector with the minimum value of $|\lambda_i|$, which can be thought of as a *selective line of least resistance*. In all the plots shown in Fig. 4.2, the selective lines of least resistance are inclined at about a 45 degree angle in trait space.

In the 3-trait case, the variety of ISSs that can be approximated with quadratic surfaces is also very large (Phillips & Arnold 1989). Seven varieties are shown in Fig. 4.4, where the canonical axes are denoted y_1 , y_2 , and y_3 , but are not ordered by the size of their eigenvalues. Each of the nested surfaces in each figure represents equal values for relative fitness. Perhaps the easiest 3-trait case to visualize is stabilizing selection on all three axes, shown in Fig. 4.4a. Here the the fitness optimum, θ , a stationary point, is situated at the intersection of the axes. Fitness falls off as concentric spheres about this point. In Fig. 4.4f, the optimum is a line corresponding to y_3 , with stabilizing selection on y_1 and y_2 , but no selection along the y_3 -axis.



4.4 The adaptive landscape.

The adaptive landscape relates population mean fitness to average trait values in the multivariate case, just as it does in the univariate case. The slope and curvature of the AL, evaluated at the trait mean, \bar{z} , are related to the selection gradients. In general and in the 2-trait case,

$$\boldsymbol{\beta} = \frac{\partial \bar{W}}{\bar{W} \partial \bar{z}} = \frac{\partial \ln \bar{W}}{\partial \bar{z}} = \begin{bmatrix} \partial \ln \bar{W} / \partial \bar{z}_1 \\ \partial \ln \bar{W} / \partial \bar{z}_2 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \quad (4.05)$$

and

$$\gamma - \boldsymbol{\beta}\boldsymbol{\beta}^T = \frac{\partial^2 \bar{W}}{\bar{W} \partial \bar{z}^2} = \frac{\partial^2 \ln \bar{W}}{\partial \bar{z}^2} = \begin{bmatrix} \partial^2 \ln \bar{W} / \partial \bar{z}_1^2 & \partial^2 \ln \bar{W} / \partial \bar{z}_1 \partial \bar{z}_2 \\ \partial^2 \ln \bar{W} / \partial \bar{z}_1 \partial \bar{z}_2 & \partial^2 \ln \bar{W} / \partial \bar{z}_2^2 \end{bmatrix} = \begin{bmatrix} \gamma_{11} - \beta_1^2 & \gamma_{12} - \beta_1 \beta_2 \\ \gamma_{12} - \beta_1 \beta_2 & \gamma_{22} - \beta_2^2 \end{bmatrix}. \quad (4.06)$$

The directional selection gradient, $\boldsymbol{\beta}$, gives the direction of steepest uphill slope (first derivatives) from the multivariate mean, \bar{z} (Lande 1979). The matrix $\gamma - \boldsymbol{\beta}\boldsymbol{\beta}^T$ describes the curvature (second derivatives) of the AL at the multivariate mean (Lande 1979, Lande & Arnold 1983). Notice that the sign elements in the second term in (4.06) are always negative, so we can conclude that the curvature of the AL ($\gamma - \boldsymbol{\beta}\boldsymbol{\beta}^T$) is always less than the curvature of the ISS (γ).

To visualize the AL as a surface, we need to average the ISS over translations (lateral shiftings) of the trait distribution, $p(z)$. Building on our discussion of the univariate case (sec. 3.3), this averaging is easy if $p(z)$ is multivariate normal (2.0) and if the ISS is approximated by a multivariate Gaussian surface,

$$W(z) = \exp\{-\frac{1}{2}(z - \theta)^T \boldsymbol{\omega}^{-1}(z - \theta)\} \quad (4.07)$$

(Lande 1979, 1980, or 1981?). The $\boldsymbol{\omega}^{-1}$ term is the inverse of the $\boldsymbol{\omega}$ -matrix, which in the 2-trait case takes the form

$$\boldsymbol{\omega} = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{bmatrix}, \quad (4.08)$$

with ω_{11} and ω_{22} analogous to variances. When these two coefficients are positive, the ISS may be a bell-shaped hill (or any one of the other surfaces shown in Fig. 4.w!). The off-diagonal term, ω_{12} , is analogous to a covariance. When it is positive, the hill is tilted upward in the $z_1 \times z_2$ dimension. When ω_{12} is negative the hill is tilted downward. We can produce a multivariate AL by averaging this Gaussian ISS over a multivariate normal trait distribution. The resulting adaptive landscape is also multivariate Gaussian,

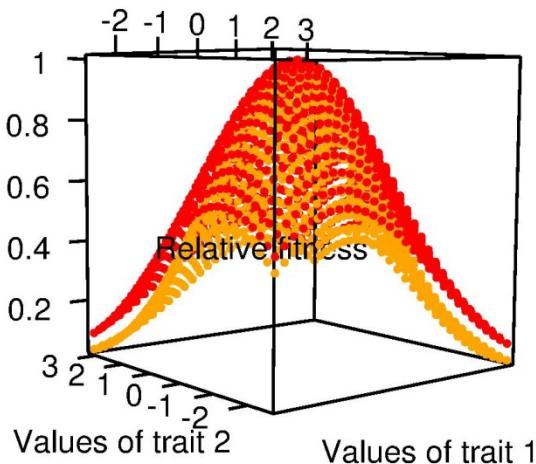
$$\bar{W} \propto \exp\{-(\bar{z} - \theta)^T (\boldsymbol{\omega} + P)^{-1}(\bar{z} - \theta)\}, \quad (4.09)$$

where $\boldsymbol{\omega} + P$ is a shape and orientation matrix

(Lande 1979). In the 2-trait case,

$$\boldsymbol{\omega} + P = \begin{bmatrix} \omega_{11} + P_{11} & \omega_{12} + P_{12} \\ \omega_{12} + P_{12} & \omega_{22} + P_{22} \end{bmatrix}, \quad (4.10)$$

where, as in $\boldsymbol{\omega}$, the diagonal terms are analogous to variances and the off-diagonal terms are analogous to a covariance, and both kinds of terms carry the



same kinds of geometric interpretations. Because P_{11} and P_{22} are always positive, the AL will be somewhat flatter than the ISS, but its tilt may differ if ω_{12} and P_{12} differ in sign. If stabilizing selection is weak, however, so that $\omega \gg P$, the ISS and AL will be very similar in configuration (Fig. 4.5).

Figure 4.5. A Gaussian ISS and corresponding Gaussian AL are similar in configuration if stabilizing selection is weak, as in this example: Gaussian ISS (orange) with $\theta = (0,0)$ and $\omega = (5,2.5,2.5,5)$ and the corresponding Gaussian AL with $\omega + P = (6,2.5,2.5,6)$, where $P = (1,0,0,1)$.

A Gaussian form for the ISS is also useful in theoretical work because it enables us to solve for the multivariate trait distribution after selection. If the trait distribution, $p(z)$, is multivariate normal before selection and the ISS is Gaussian (4.07), the phenotypic trait distribution is normal after selection with means and variance-covariance matrix given by

$$\bar{z}^* = (\omega + P)^{-1}(\omega\bar{z} + P\theta)$$

and

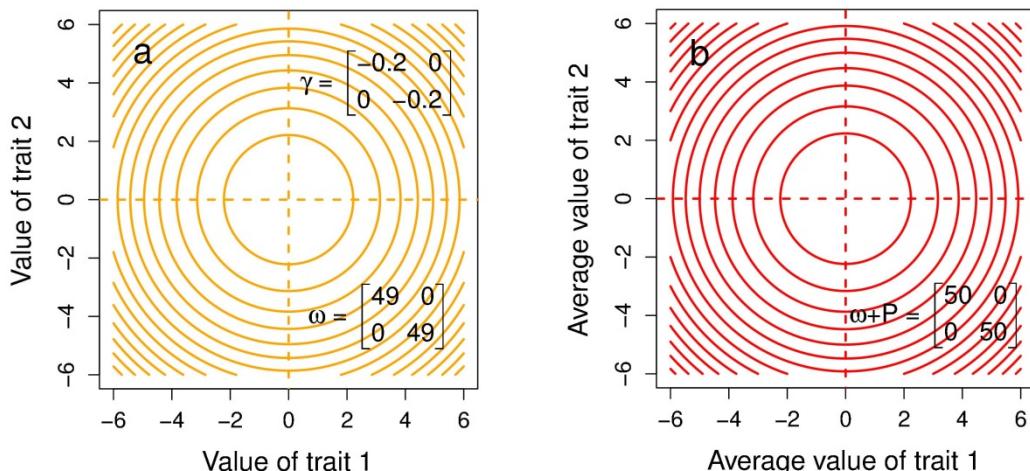
$$P^* = (\omega + P)^{-1}\omega P,$$

which are generalizations of Lande's (1981) results for the univariate case (3.12). In these forms, we see that the AL exerts its effects on $p(z)^*$ via its shape matrix, $\omega + P$. **What is the formula for beta when the ISS is Gaussian?****

When multivariate stabilizing selection is weak, we can estimate ω from the negative inverse of the γ -matrix,

$$\omega = -\gamma^{-1}, \quad (4.11)$$

which is the multivariate generalization of (3.10). From this approximation of ω , we can easily approximate $\omega + P$. Ordering eigenvalues by their absolute values (section 4.3), the ω - and γ -matrices have the same eigenvectors but in reverse order, so ω_{max} corresponds to γ_{min} , the direction in trait-space in which the ISS has the weakest curvature. In other words, ω_{max} is the eigenvector of ω with the largest eigenvalue, giving the direction in trait space that is most forgiving with respect to selection, a selective line of least resistance. Similarly, $\omega + P_{max}$ is the eigenvector of $\omega + P$ with the largest eigenvalue, which is likely to be similar to ω_{max} if multivariate stabilizing selection is weak (Fig. 4.6).



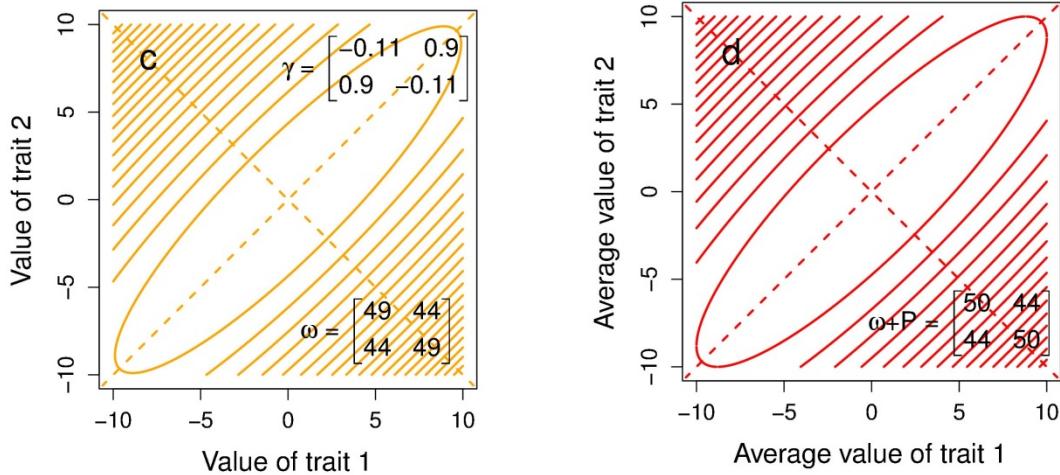


Figure 4.6 Contour plot portrayals of weak bivariate Gaussian ISSs and their corresponding adaptive landscapes. In each figure the optimum is at $z_1=0$ and $z_2=0$, $P=(1,0,0,1)$. Eigenvectors are shown as dashed lines. Matrix representations are superimposed on each surface. (a) An ISS with equally strong stabilizing selection on each trait with no correlational selection. Contours show equal values of relative fitness, $w(z)$. (b) The AL corresponding to Fig. 4.6a. Contours show equal values of average absolute fitness, \bar{W} . (c) An ISS with equally strong stabilizing selection on each trait and strong correlational selection, $r_\omega=0.9$. (d) The AL corresponding to Fig. 4.6c.

In conclusion, we note while the ISS is useful as a local description of selection, the AL is useful because it is the surface on which \bar{z} evolves. Because of this distinction, we will need a vision of the AL in our later discussion of evolution. We adopt a Gaussian framework because it gives us that needed vision.

4.5 Examples of quadratic approximations of the ISS

We can appreciate the relationship of selection gradients to the ISS by returning to the example of how crawling speed in newborn garter snakes is affected by body and tail vertebral counts (Arnold 1988, Arnold & Bennett 1988). In Chapter 2 we showed how the gradients in this example could be estimated by comparing samples before and after selection (2.02, 2.09), but here we show estimation by linear and quadratic regression. The directional selection gradients were estimated by linear regression (4.2) and the nonlinear selection gradients were estimated by quadratic regression (4.3) (Table 4.1). Bootstrap estimates of 95% confidence limits suggest that the point estimates of β_1 , β_2 , γ_{11} and γ_{22} are not different from zero. Point estimates of standard errors provided by regression analysis without bootstrapping provide a similar picture (Table 4.1). In contrast, the correlational selection gradient, γ_{12} , is positive and substantially different from zero. To visualize the ISS that corresponds to these selection gradients, we plot the corresponding surface,

$$w(z) = 1.0 + 0.031z_1 - 0.011z_2 + \frac{1}{2}(-0.011)z_1^2 + \frac{1}{2}(-0.006)z_2^2 + 0.079z_1z_2,$$

and compare that surface with the one described by the selection gradients estimated in Chapter 2,

$$w(z) = 1.0 + 0.031z_1 + 0.002z_2 + \frac{1}{2}0.097z_1^2 + \frac{1}{2}0.116z_2^2 + 0.176z_1z_2.$$

The gradient estimates are subtly different as a consequence of different estimation procedures, but how different are the two surfaces? Contour plots reveal that the surfaces are remarkably similar and saddle-shaped in both cases (Fig. 4.2). Both surfaces curve slightly upward from the bottom lefthand corner to

the upper, right-hand corner, and slightly downward from the upper left-hand corner to the bottom right-hand corner.

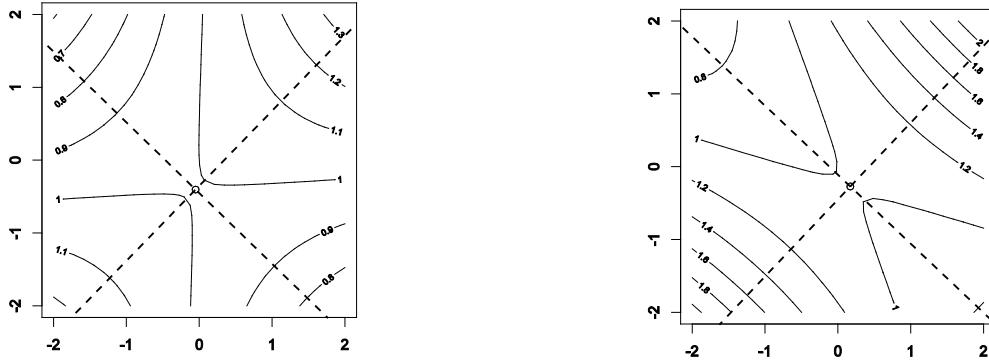


Figure 4.z Crawling speed performance surfaces as a function of body and tail vertebral numbers, estimated by two different procedures. Eigenvectors are shown in dotted lines. Stationary points, calculated using (4.03), are shown as open circles, near the center of each plot. A. (left) Contour surface drawn using selection gradients estimated by linear and quadratic regression. The slope of the leading eigenvector is 0.971 ($\lambda_1=0.071$, $\lambda_2=-0.087$). B. (right) Contour surface drawn using selection gradients estimated by comparing samples before and after selection (2.02, 2.09). The slope of the leading eigenvector is 1.055 ($\lambda_1=0.282$, $\lambda_2=-0.069$).

The new axes described by the eigenvectors are biologically informative. Eigenvectors are linear combinations of the original axes. In the case of selection surfaces, each eigenvector consists of a series of weights (loadings) each of which describes how a particular trait contributes to the new axis. When all the weights are positive the weighting produces a combination that is the weighted average (sum) of the traits. When the weights have a mixture of signs, the weighting produces a combination that is difference or contrast in traits. Thus, the leading eigenvectors of the surfaces in Fig. 4.z, with slopes close to one, represent the sum of the two vertebral counts, whereas the second eigenvectors, with slopes close to minus one, represent the difference in the two vertebral counts. In Fig. 4.za, curvature is about the same in both the sum and difference directions ($|0.071| \approx |-0.087|$), but in Fig. 4.zb, curvature is substantially stronger in the sum direction ($|0.282| > |-0.069|$) telling us that variation in the sum of counts is tolerated less than variation in direction of the count difference.

This example also illustrates the utility of eigenvectors and eigenvalues in comparing surfaces. Although the similarity of the two surfaces is not apparent from their γ -matrices, it is readily deduced from their eigenvectors and eigenvalues (Fig. 4.z). In both cases the slope of the leading eigenvector is nearly one, selection in that direction is disruptive, as indicated by the positive sign of the corresponding eigenvalue, whereas selection in the opposite direction is stabilizing, as indicated by the negative sign and magnitude of the eigenvalue corresponding to that eigenvector. In contrast, the values of β_1 and β_2 tell us little about the surface. Instead, they describe slope of surface at the location of the bivariate mean, which in the present case is very close to the stationary point.

The illuminating properties of the eigenvectors and eigenvalues suggest that they are best platform to explore the sampling properties of the ISS, however it is estimated. In particular, a bivariate plot of the two eigenvalues estimated for each bootstrapped sample can tell us how often the ISS takes the

form of a saddle (versus other shapes). When the ISS is estimated by multiple regression, the ISS most often takes a saddle shape (λ_1 positive, λ_2 negative), but a bowl-shaped surface is relatively common (λ_1 positive, λ_2 positive) (Fig. 4.zzb). The distribution of slopes for the leading eigenvector is bimodal, and straddles zero with a slight preponderance of negative values (Fig. 4.zza). When the ISS is estimated from samples before and after selection, the shapes of the ISS show the same split between saddle and bowl shapes. In this case, however, the distribution of slopes for the leading eigenvector is unimodal with a mode close to one, and nearly all values are positive.

Bootstrap sampling of the two estimation procedures is revealing because it tells us that our visualization of the ISS is fragile. Our vision of the ISS changes shape from boot to boot. The slope of the leading eigenvector is, however, a more stable feature of the ISS when it is estimated by samples before and after selection. In that case, a confidence interval could easily be specified about the strong mode (near one). In the case of estimation by regression, however, a confidence interval would be misleading in and of itself, because the distribution of slopes shows two modes (one near 2 and one near -2). The results in the two case do not recommend one estimation procedure over the other. Instead, this case study emphasizes the point that when curvature of the ISS is slight, our confidence in shape of the ISS and in its major axis is compromised.

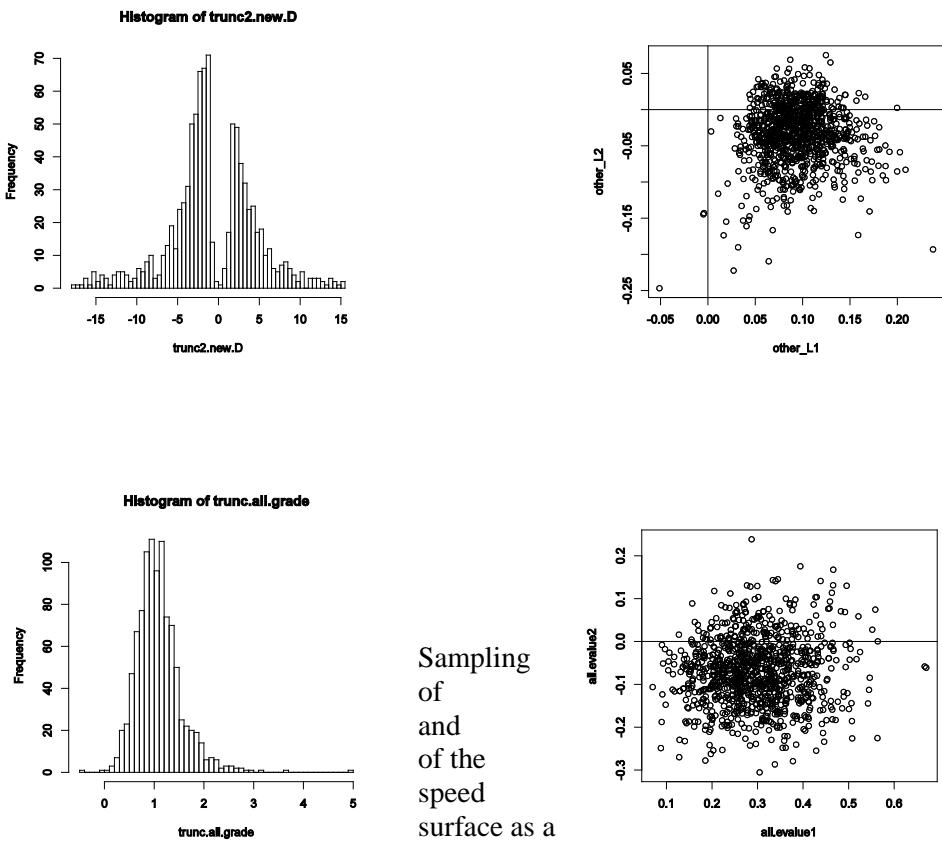


Fig. 4.zz
distributions
eigenvalues
eigenvectors
crawling
performance
function of

vertebral numbers, estimated by bootstrapping two different estimation procedures ($n=1000$ samples with replacement). A. (upper left) The bootstrap distribution of slopes of the leading eigenvector of the γ -matrix, estimated by multiple regression. B. (upper right) Bootstrap values of γ_2 plotted against γ_1 , when the γ -matrix is estimated by multiple regression. C. (lower left) The bootstrap distribution of slopes of the leading eigenvector of the γ -matrix, estimated from samples before and after selection. D. (lower right)

Sampling
of
and
of the
speed
surface as a
body and
tail

Bootstrap values of γ_2 plotted against γ_1 , when the γ -matrix is estimated from samples before and after selection.

Finally, we can use our regression estimate of the γ -matrix to illustrate approximations of ω and $\omega+P$. Using 4.xx, we find that **Note that the values of gamma here and in Table 4.1 need to be reconciled with the values used in plotting the selection surface**

$$\omega = -\gamma^{-1} = -\begin{bmatrix} -0.011 & 0.079 \\ 0.079 & -0.006 \end{bmatrix}^{-1} = \begin{bmatrix} -0.957 & -12.784 \\ -12.784 & -1.711 \end{bmatrix}.$$

The negative inverse operation converts a matrix of second derivatives for a quadratic surface, γ , into a matrix of variance- and covariance-like elements that describes a Gaussian surface, ω . When the variance-like diagonal elements of ω are positive the Gaussian surface is convex, like a familiar bell curve, but when the diagonal elements are negative, the Gaussian surface is concave, like an upside down bell curve. The canonical form of γ , with eigenvalues on the main diagonal, is

$\begin{bmatrix} -0.088 & 0 \\ 0 & 0.071 \end{bmatrix}$, with the eigenvectors of γ given by the columns of $\begin{bmatrix} -0.718 & -0.696 \\ 0.696 & -0.718 \end{bmatrix}$. The canonical form of ω is $\begin{bmatrix} -14.123 & 0 \\ 0 & 11.455 \end{bmatrix}$, with eigenvectors given by the columns $\begin{bmatrix} 0.697 & -0.717 \\ 0.717 & 0.697 \end{bmatrix}$.

As expected, the eigenvectors of γ and ω are the same but in reverse order. The selective line of least resistance is given by $\gamma_{min} = [-0.696 \quad -0.718]^T$, or equivalently $\omega_{max} = [0.697 \quad 0.717]^T$

Because our vertebral count distributions are approximately normal before selection, and because the trait variances have been standardized to one before selection, we can readily visualize the curvature of the Gaussian ω -surface in relation to our trait variances. A first eigenvalue of -14.123 tells us that the Gaussian surface is concave in the direction of the first eigenvector with a ‘variance’ about 14 times larger than our trait variance. Similarly, a second eigenvalue of 11.455 tells us that the surface is convex in an orthogonal direction with a ‘variance’ about 11 times larger than our trait variances.

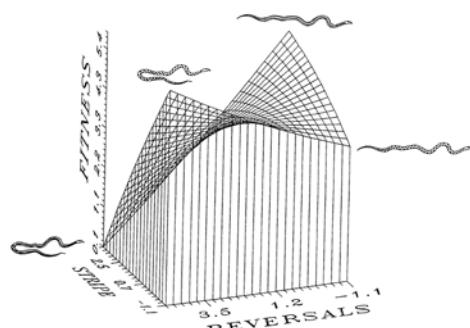
Our corresponding estimate of $\omega+P$, using values of P from Table 4.1, is

$$\omega + P = \begin{bmatrix} -0.957 & -12.784 \\ -12.784 & -1.711 \end{bmatrix} + \begin{bmatrix} 1.000 & 0.073 \\ 0.073 & 1.000 \end{bmatrix} = \begin{bmatrix} 0.043 & -12.711 \\ -12.711 & -0.711 \end{bmatrix}$$

with canonical form $\begin{bmatrix} -13.051 & 0 \\ 0 & 12.383 \end{bmatrix}$ and eigenvectors given by the columns $\begin{bmatrix} 0.697 & -0.718 \\ 0.718 & 0.697 \end{bmatrix}$.

As expected, the canonical forms of ω and $\omega+P$ are very similar. The selective line of least resistance determined from $\omega + P$ is $[0.697 \quad 0.718]^T$.

Brodie’s (1992) study of how viability in garter snakes is affected by coloration and behavior provides another illuminating example of selection surface analysis. This study was motivated by the observation that coloration pattern and antipredator behavior coevolve in snakes, so that higher taxa fall out along a bivariate continuum (Jackson et al. 1976). At one end of



the continuum are slow-moving snakes with blotched color patterns that rely on crypsis to evade predators. At the other end are fast-moving snakes with stripes or no pattern that rely on speed and optical illusion to escape predators. Remarkably, this same coloration-behavior continuum occurs within individual populations of the garter snake *Thamnophis ordinoides*. On the continuum, the population includes blotched snakes that tend to reverse directions and striped snakes that crawl in a straight line, but off-continuum snakes are also well represented (striped, reversing snakes and blotched, straight-crawling snakes). Despite this bivariate smear before selection, the connection to the interspecific continuum becomes clear when we visualize the viability selection surface. That visualization reveals that snakes that fall out along the continuum are favored by selection, but selection acts against snakes with the maladaptive combinations of striped-reversal and blotched-straight crawl (Fig. 4. zw). Turning to the selection coefficient that exerts the most influence on the portrayal of the surface, this example bears many similarities to the *T. radix* crawling speed-vertebral count example that we just discussed. As in the *T. radix* example, it is a correlational selection differential and corresponding gradient that are sufficiently large to attain statistical significance. Bivariate selection on coloration pattern and behavior affects the correlation between those traits. In a later section we will discuss the significance of the fact that this bivariate selection has also resulted in a genetic correlation between these two traits (Brodie 1993).

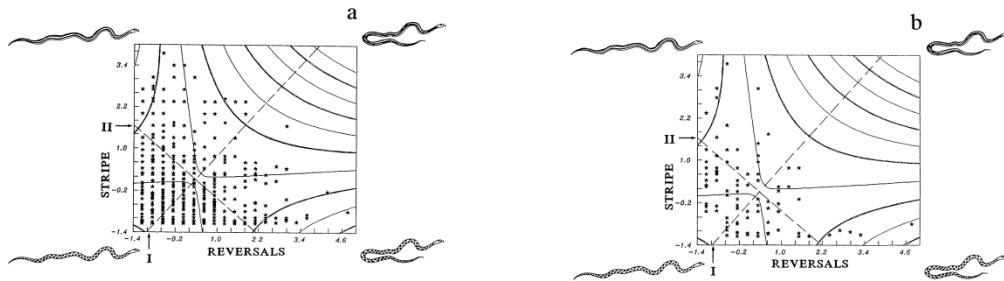


Figure 4.zx Viability selection as a function of coloration pattern and antipredator behavior in *T. ordinoides*. Survival in the field is shown as a function of reversals (the tendency to reverse directions during simulated predation exposure in the laboratory) and stripe (overall stripedness of the color pattern). Contours show relative fitness (survival) as a quadratic function of reversals and stripe. Eigenvectors are shown as dashed lines. Selection is stabilizing (concave) along the leading eigenvector (I) and disruptive (convex) along the second eigenvector (II). (a) The bivariate sample before selection ($n=646$). (b) The bivariate sample after selection ($n=101$). (c) Surface plot of the quadratic selection function.

4.9 Surveys of quadratic approximations to the ISS

We would like to know what multivariate form of selection is most prevalent in nature. From a theoretical standpoint we can expect quadratic approximations to take a variety of shapes (bowl, dome, saddle, rising ridge, etc.), but which of these shapes is most commonly encountered? Blows & Brooks (2003) have taken a major step towards answering this question by surveying studies that tackled the issue of shape in three or more dimensions (Table 4.az). The surprising answer was that in 17 of 19 cases the surface was saddle-shaped (one bowl and one dome accounted for the other two cases). This answer is surprising because a saddle is inherently unstable. For this very reason theoreticians often employ a multivariate dome to describe selection on continuous traits. Dome-shaped selection tends to move the trait mean towards a stable local optimum. Why is this multi-stabilizing form of selection so rare? One possible reason – among many – for the disconnect between theoretical expectation and empirical realization is that we nearly always measure components of fitness, not lifetime fitness in studies of selection surfaces. Arnold & Bennet (1988) and Brodie (1992), for example, assessed selection over the first few weeks or years of life, not over the entire life span.

... another possibility is that when curvilinear selection is weak, a point estimate may often have the opposite curvature to the parametric surface ... *illustrate with boot-strapping of radix performance surface*

Table A1: Comparative data set

n	Largest γ_{ii}	λ	Type of surface	Type of selection	Reference
5	.044	.062	Saddle	O	Mitchell-Olds and Bergelsson 1990
4	-.457	-1.262	Saddle	F	Moore 1990
4	-.550	-.714	Saddle	M	Moore 1990
4	-.707	-1.093	Saddle	F	Moore 1990
4	-.498	-.729	Saddle	M	Moore 1990
4	.102	.155	Saddle	M	Moore 1990
4	-.538	-.650	Saddle	F	Moore 1990
4	-.122	-.273	Saddle	S	Brodie 1992
3	-.874	-.875	Saddle	F	Nunez-Farfan and Dirzo 1994
3	.370	.552	Saddle	F	O'Connell and Johnston 1998
3	1.180	1.709	Bowl	F	O'Connell and Johnston 1998
3	.770	1.124	Saddle	F	O'Connell and Johnston 1998
3	.260	.283	Saddle	F	O'Connell and Johnston 1998
3	.200	.305	Saddle	F	O'Connell and Johnston 1998
3	.23	.26	Saddle	F	O'Connell and Johnston 1998
5	.994	.999	Saddle	F	Simms 1990
3	-.019	-.021	Peak	S	Kelly 1992
4	.016	.027	Saddle	S	Kelly 1992
5	.112	.214	Saddle	F	Kelly 1992

Note: Analysis of nonlinear selection in 19 data sets identified by Kingsolver et al. (2001) involving three or more traits. We have presented the largest γ_{ii} and λ_i from each γ . The type of surface is based on the signs of all eigenvalues, following the descriptions used by Phillips and Arnold (1989); if eigenvalues were all negative, the surface has a peak; if all were positive, the surface is a bowl; and if some were negative and positive, the surface is a saddle. The type of selection follows the categories used by Kingsolver et al. (2001): S = survival, M = mating success, F = fecundity/fertility, O = other. n = number of traits measured in each study.

4.10 Cubic spline approximation to the ISS

The cubic spline solution to the problem of estimating the shape of the ISS (section 3.3) can be generalized to multiple traits (Schluter & Nychka 1994). A 2-trait example is shown in Fig. 4.5

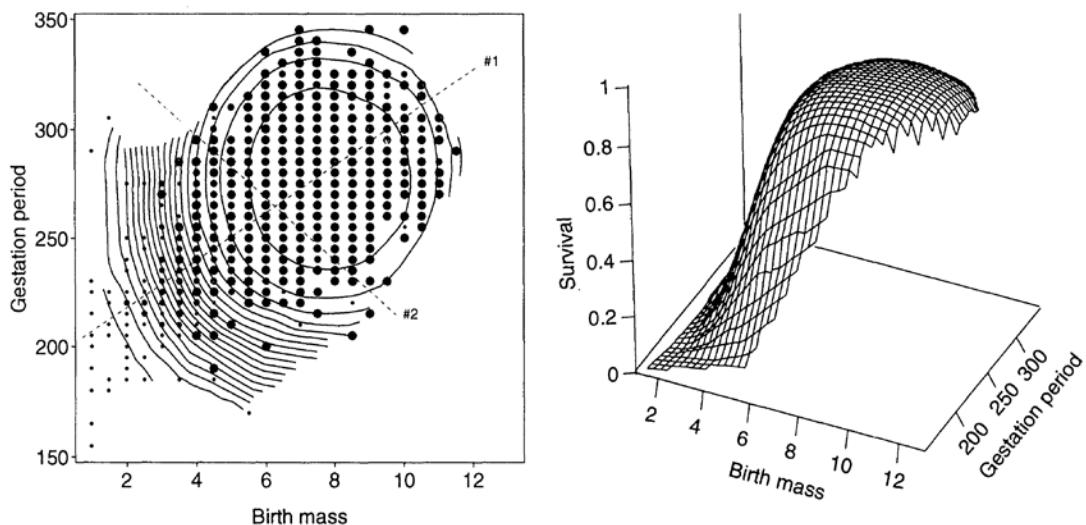
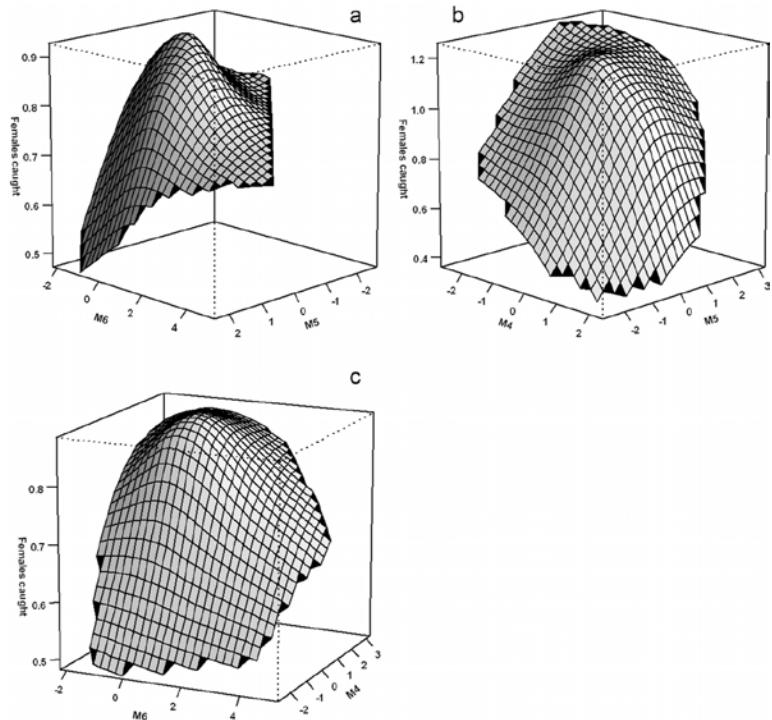


FIG. 5.—Survival probability of male human infants as a function of birth mass (pounds) and gestation period (days). The surface was approximated using two directions (table 3) indicated by the *dashed lines*. Fitness contours in the *left panel* are in increments of 0.5. They describe a dome rising steeply from lower left to a broad plateau above and right of center, declining slightly along its outer edges. *Right panel* gives three-dimensional perspective. Symbol size increases with increasing average survival of individuals in each mass (rounded to nearest 0.5 lb.) and gestation class (rounded to nearest 5 d); differences in symbol size are exaggerated in the range 0.90–1.00 to highlight slight survival differences. The most extreme observation at large mass was excluded from the analysis. $\ln(\lambda) = -10$; $n = 7,036$.

Bentson et al. (2006) ... cricket vocalizations ...

Responses of crickets to artificial calls



Bentson et al. 2006 Figure 2: Thin-plate spline visualizations of the fitness surfaces, which demonstrate significant convex stabilizing selection between m_5 and m_6 (a), m_4 and m_5 (b), and m_4 and m_6 (c).

4.11 Technical issues in estimating and interpreting the ISS {Consider moving this section to before the survey section}

Over the last few decades, the trend in selection studies has been to include portrayals of the ISS, as well as tables of β and γ estimates and their standard errors. In part this trend reflects increasing appreciation of the fact the ISS cannot be visualized from a table of β and γ estimates (Phillips & Arnold 1989). The other growing realization is that the ISS is an object of interest in its own right (Blows 2007). Accurate portrayals of the ISS can summarize complex modes of selection, provide a vision of the AL, and may point to important avenues of evolution during adaptive radiations (Chapter 17). Nevertheless, a disturbing fraction of selection studies fail to report coefficients of nonlinear selection (often estimates of γ_{ij} are missing, but sometimes estimates of γ_{ii} as well).

Using multivariate regression to capture a vision of the ISS immediately places us in a quandry of how many traits to include. The quandry arises because the vagaries of sampling force us to consider the over-all shape of the ISS rather than just the values and significance of individual elements in β and γ and because the actual targets of selection (the real subset of traits under the strongest directional or nonlinear selection) are usually unknown to us. To capture the real targets, we could include more traits in the selection analysis, but – because γ is a matrix – the number of coefficients to be estimated goes up as the square of the number of traits. Increasing the number of traits may increase our chances of including real targets, but as we attempt to estimate more coefficients our power to detect selection goes down (i.e., more of our estimates are nonsignificant). Unfortunately, this multivariate quandry has no universal solution, although it can often be mitigated by two considerations. The first is that ecological or biomechanical arguments may enable us to narrow the field of possible traits. For example, in the case of Galapagos finches, a host of ecological observations and biomechanical analogies, bolster a priori the choice of beak depth and length as probable targets of selection (Grant xxxx, Grant & Grant yyyy). On this basis the investigators restricted their selection analysis to just two traits. The second consideration is that selection can sometimes be profitably viewed as acting on combinations of traits. Taking this approach, one can analyze selection on principal components or other linear combinations of traits and thereby reduce the trait number to one, two or three. This tact works best if the linear combinations are readily interpretable. For example, if the traits are linear measurements, the first principal component can often be interpreted as a measure of overall size, while components with small eigenvalues represent measures of shape (Jolicoeur & Mosimann 1960).

A variety of other issues are in general endemic to multivariate statistical analyses, not just to selection analysis. Although these issues are not peculiar to selection analysis by multiple regression, they should be considered in using this approach or in interpreting results (Lande & Arnold 1983, Mitchell-Olds & Shaw 1987, Brodie et al. 1995). (1) The problem of *multicollinearity* arises if a subset of traits is highly correlated. In this circumstance, the analysis can fail to fully account for correlations among traits, with resulting distortions of estimated coefficients. As in the quandry discussed above, two useful approaches are to reduce the number of traits (using outside criteria) or to use principal components. (2) Multiple regression results are always conditional on the proviso that all correlated traits under selection have been included in the analysis. To the extent that correlated, unmeasured traits are exposed to selection, their exclusion may have distorted estimates of selection coefficients. While this proviso may at first sound fatal, most investigators feel that it is better to account for some correlations than to fall back on univariate analyses. (3) Stepwise regression and related strategies are useful statistical solutions to the problem of sorting through a large field of possible explanatory variables. These approaches are problematic in selection analyses, however, because of the biases they cause in the estimates of β and γ . (4) Variance in relative fitness (the *opportunity for selection*) can limit the power to detect selection and should be taken into account in designing selection studies (Hersch & Phillips 2004). (5) A dichotomous fitness measure (e.g., survivors and nonsurvivors) complicates the estimation of standard errors for β and γ . Janzen & Stern 1998 describe a solution that employs logistic regression.

A technical problem plagues published estimates of nonlinear selection coefficients, γ . Recall that in the quadratic regression equation (4.3) the nonlinear coefficient for z_i^2 is $\frac{1}{2}\gamma_{ii}$ so that γ_{ii} is a second derivative of the ISS. Stinchcombe et al (2008) queried the authors of 32 papers published in the period 2002-2007 and found that in a sizeable fraction of those papers (78%), the authors failed to take the factor of $\frac{1}{2}$ into account. The consequence is that the published estimates labeled γ_{ii} are actually $\frac{1}{2}\gamma_{ii}$. Because the diagonal elements of γ -matrix are under-estimated by a factor of 2, while the off-diagonal elements are not, portrayals of the ISS can be affected. Unless authors specifically state that the factor of $\frac{1}{2}$ was taken into account or reproduce (4.3), a cautious reader should assume that published values of γ_{ii} are likely to be in error.

Under certain circumstances one must assume that nonlinear selection is weak ($\omega \gg P$ or $P \gg |\gamma|$).

One such circumstance arises when one wishes to claim that a quadratic or Gaussian approximation provides a good representation of the ISS. If the actual ISS is asymmetric or stabilizing or disruptive selection is strong, neither of these approximations is likely to be satisfactory. Likewise, when ... {need to write a simple R program to devise some rules} the simple formula for converting between γ and ω (3.10) is inaccurate.

... computational strategies {see papers by Blows, Brooks and colleagues with focus on ISS rather than on the elements of γ } ... parts of fitness, not all ... only parts of the ISS and the AL

4.12 Why the multivariate view of selection is important.

... hazards in taking a univariate view of selection ... ridges not apparent in univariate views refer back to the discussion of the Phillips & Arnold 1989 quad plots ... slight change in γ -coefficients, big impact of appearance of surface ... and amplify on those conclusions

Chapter 7: Drift of a Single, Neutral Trait

© Stevan J. Arnold 2016

Overview.- Nonrandom sampling of parents can cause the phenotypic mean to fluctuate from generation to generation. Such fluctuations are most severe in small populations and in the absence of selection. Theory for drift in the mean is based on a model in which many replicate populations are derived from a single ancestral population and thereafter evolve independently. Such models allow us to predict the variance among replicate trait means at any generation in the future, as a function of genetic variance and effective population size. In nature, populations diverge more rapidly in the short term than predicted by drift. In the long term, populations and species diverge less than predicted by drift.

7.0 The concept of effective population size, N_e

Population size has a fundamental effect on random variation in the population mean from generation to generation. Such random or stochastic effects can be conveniently modeled using a single number called effective population size, N_e (Wright 1931, Barrowclough & Lande 1987, Charlesworth 2009). In the simplest kind of model, with no selection and equal family sizes, this number simply represents the number of parents. In more complicated models, effective population size can incorporate the sex ratio of breeding parents and variation in family size (Appendix 3). In all cases, however, the basic idea is that a single number, N_e , can account for stochastic variation in genetic properties that arises each generation from finite sampling of parents.

An evolving lineage is more than a single panmictic population with a particular N_e . In particular, we want to consider the case of a population subdivided into a set of demes of various effective sizes and exchanging migrants. As we follow this metapopulation through evolutionary time, some demes go extinct and their territory is colonized by other, more successful demes. Our intuition tells us that the total N_e for such a lineage composed of 10 demes, each with an N_e of 100, must be more than 100, perhaps more than 1000. For a particular model of migration between demes this intuition is known to be correct. Wright (1939; see Whitlock & Barton for ref; see also paper by Wright in 1942 with same title in Bull Math Stat) showed for the island model (... explicate ...) that the effective size of the total population is

$$N_e = \frac{n\bar{N}}{(1 - F_{ST})}, \quad (7.0)$$

where n is the number of demes, \bar{N} is their average effective size, and F_{ST} and $(1 - F_{ST})$ are, respectively, the among- and within-deme components of variance in gene frequency. In our thought experiment, N_e could be as small as 1000 if all of the variation was within demes ($F_{ST} = 0$), or large as 100,000 if nearly all of the genetic variation was among demes ($F_{ST} = 0.99$). When F_{ST} is this large, migration is slight, and the demes behave as nearly independent entities. Under more general assumptions, Whitlock & Barton (1997) have shown that

$$N_e = \frac{n\bar{N}}{\{1 + Var(\mathcal{G})\}[\sum_i \mathcal{G}_i^2(1 - F_{STi})]}, \quad (7.1)$$

where \mathcal{G}_i is the eventual contribution of the i th deme to the whole population, $Var(\mathcal{G})$ is the variance in those eventual contributions (which are scaled so that $\sum_i \mathcal{G}_i = n$), and F_{STi} is the F_{ST} of the i th deme (Whitlock & Barton 1997). In this expression \bar{N} is an average weighted by \mathcal{G}_i^2 . We see that (7.1) reduces to (7.0), if the eventual contributions of all demes are equal ($Var(\mathcal{G}) = 0$), but in general $Var(\mathcal{G})$ reduces total N_e below the simple expectation of $n\bar{N}$.

{end this section by applying the theory just developed to the case of N_e for evolution in an entire species = *T. elegans* ... lower bound from microsat data = Manier et al ... knock the estimate up one, two or three orders of magnitude to account for N_e across the geographic range ... knock the estimate down one or two orders of magnitude to account for temporal variation in N_e = geometric mean ... arrive at a final rough estimate, which will be used later in a MIPoD analysis; then bridge to the next section}

7.1 Estimates of N_e

{see table below from Charlesworth 2009 = long term, species wide estimates; find a review and incorporate results for short-term estimates from single local populations or arrays of populations; refs from E&A 2007 = Barrowclough 1980; Begon et al. 1980; Husband and Barrett 1992; Jorde and Ryman 1996; Storz et al. 2001; Turner et al. 2002; develop the distinction between these two kinds of estimates early on in this section!}

Table 1 | Effective population size (N_e) estimates from DNA sequence diversities

Species	N_e	Genes used	Refs
<i>Species with direct mutation rate estimates</i>			
Humans	10,400	50 nuclear sequences	145
<i>Drosophila melanogaster</i> (African populations)	1,150,000	252 nuclear genes	108
<i>Caenorhabditis elegans</i> (self-fertilizing hermaphrodite)	80,000	6 nuclear genes	41
<i>Escherichia coli</i>	25,000,000	410 genes	146
<i>Species with indirect mutation rate estimates</i>			
Bonobo	12,300	50 nuclear sequences	145
Chimpanzee	21,300	50 nuclear sequences	145
Gorilla	25,200	50 nuclear sequences	145
Gray whale	34,410	9 nuclear gene introns	147
<i>Caenorhabditis remanei</i> (separate sexes)	1,600,000	6 nuclear genes	43
<i>Plasmodium falciparum</i>	210,000–300,000	204 nuclear genes	148

For data from genes, synonymous site diversity for nuclear genes was used as the basis for the calculation, unless otherwise stated.

{End this section by applying the lessons of (7.1) to the case of Eagle Lk *T. elegans* metapopulation, bringing in the consequences of source-sink pop dynamics etc; Manier et al 2005 and ms on stochastic demography in the same system}

7.2 Sampling from a normal distribution of breeding values

If we take repeated samples of size N from a normal distribution with variance V , the variance among the means of those samples will be V/N . This result reinforces our intuition that differences due to sampling will be appreciable with small samples, but inconsequential if the samples are large. We can apply this result to model the consequences of drift for replicate populations that obey the same rules of inheritance and sampling. We imagine a base population in which breeding values are normally distributed with a mean $\bar{x} = \bar{z} = 0$ and a variance G . Suppose we establish replicate populations, each of size N_e , from this base population. The variance among those replicates in mean breeding value will be

$$Var(\bar{z}_l) = \frac{1}{N_e} G \quad (7.00)$$

and the overall mean will be zero (Lande 1976).

7.3 Projecting the distribution of breeding values into the future

Now allow each of the replicate populations to constitute a lineage. In each lineage G remains the same and the same process of sampling occurs each generation. If we follow a single lineage through time, its phenotypic mean each generation, \bar{z}_t , deviates from the mean in the preceding generation, \bar{z}_{t-1} , by an

amount that can be specified by taking a random draw from normal distribution of breeding values with zero mean and variance G/N_e . Following the mean through time, we see a random walk (Fig. 7.0a). Each

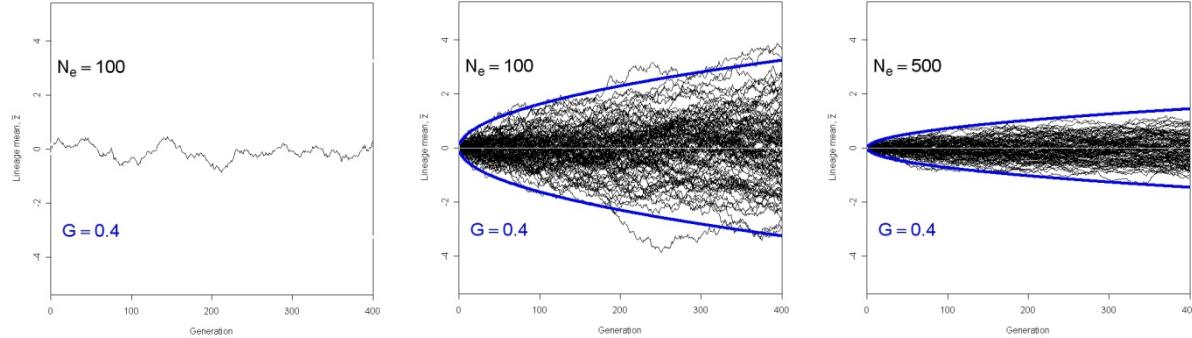


Figure 7.0 Genetic drift causes the phenotypic mean of a lineage to undergo a random walk. Simulations of genetic drift using 7.01 and 7.02. The lineage mean, \bar{z} , is shown in units of within-population phenotypic standard deviation, \sqrt{P} . (a) The random walk of a single lineage of small effective size. The random walks of 100 replicates lineages are shown for small (b) and moderate sized (c) populations. The theoretical 99% confidence limits for the mean of lineage means are shown in blue.

replicate lineage undergoes an independent random walk that obeys the same rules. We now wish to consider the statistical properties of the entire ensemble of replicate lineages at some generation, t , in the future. We began in generation 1 with an among-replicate distribution of means that was normally distributed with a mean of 0 and a variance of G/N_e (7.00). The distribution of means in generation 2 is the product of that distribution and normal distribution with the same properties, a so-called convolution of normal distributions, which is itself a normal distribution. By extending this logic, we can conclude that the among-replicate distribution of means at any future generation t , is a normal distribution

$$\Phi(\bar{z}_t) = \frac{\exp[-\frac{1}{2}(\bar{z}_t - \bar{z}_0)^2 / D(t)]}{\sqrt{2\pi D(t)}}, \quad (7.01)$$

with a mean of zero, $\bar{z}_0 = 0$, and a variance given by

$$D(t) = \frac{1}{N_e} G t, \quad (7.02)$$

where $D(t)$ stands for the dispersion (variance) of replicate means at generation t (Lande 1976). In other words, the variance among replicates increases linearly with time. This linear property means that the 99% confidence limits for the mean of means are ever-expanding, so that their graph forms a conic section (Fig. 7.0b, c) ... normal distributions of replicate means with increasing variance (Fig. 7.1, based on figs. in Lande 1976 and Estes & Arnold 2007).

The results shown in Figs. 7.0 and 7.1 are obviously simplifications since we have ignored the effects of sampling on G , a problem to which we will return in Chapter y. We note, however, that (7.02)

might be a reasonable approximation if N_e is large enough so that the sampling effects on G are relatively small.

Figure 7.1. Distributions of lineage means under neutrality. Each curve represents the distribution of lineage means after some number of elapsed generations, t . Genetic variance = 0.4 and $N_e = 1000$.

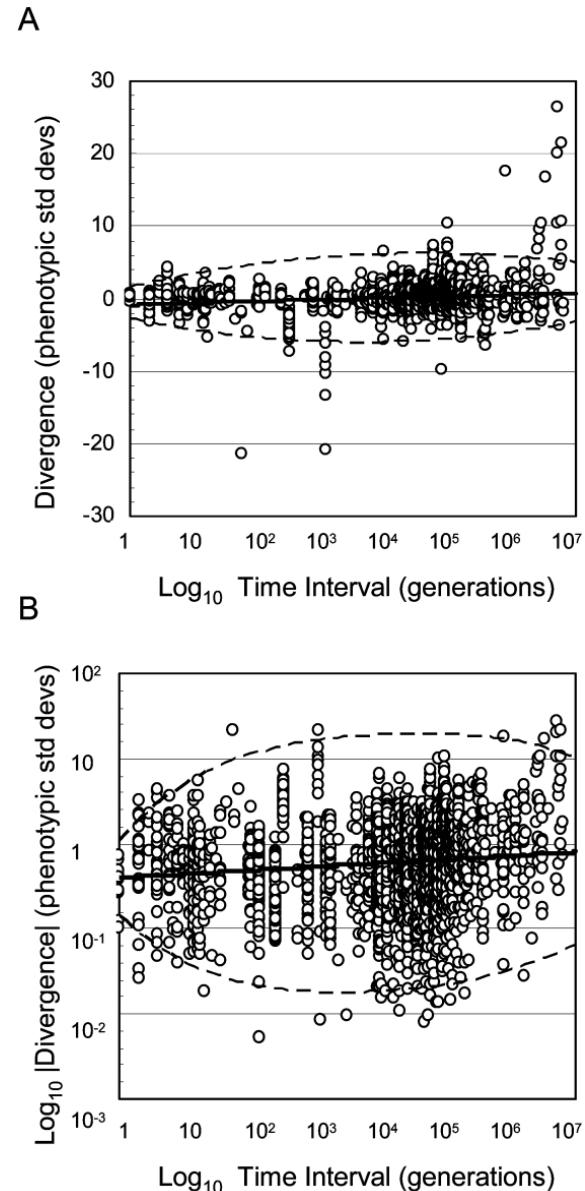
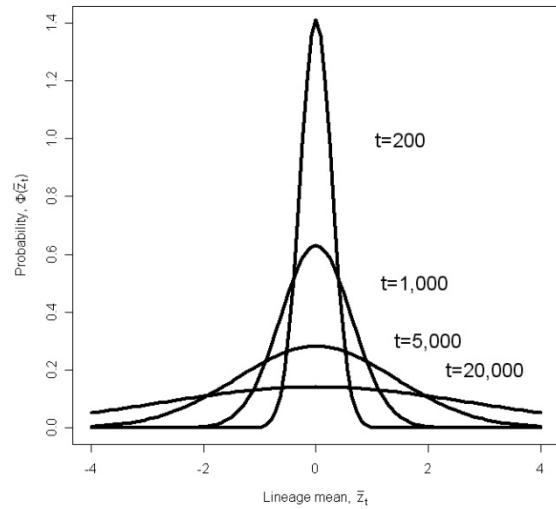


Figure 1: A, Plot of divergence as a function of time interval. Divergence is measured as the difference between the average trait values of an ancestral and a descendant population, expressed in units of pooled phenotypic standard deviation. The slope of the fitted regression line is 0.2407 (SE = 0.0268), $P < .0001$. B, Plot of \log_{10} absolute divergence on \log_{10} time interval. The slope of the fitted regression line is 0.0437 (SE = 0.0070), $P < .0001$. In both plots, time interval is measured in number of generations on a \log_{10} scale. The broken lines show the boundaries of the 99% confidence ellipse for the data. The data ($n = 2,639$) are from Gingerich (2001). The fitted slopes are equivalent to 0.84 phenotypic standard deviations per million generations on a raw scale.



We have also used a single number to represent population size throughout time in all replicate lineages, another obvious simplification. If we model the effects of temporal fluctuation in population size, it turns out that geometric mean population size provides a good overall approximation of sampling effects (Crow & Kimura 1970 or appropriate S. Wright ref). Bottlenecks have a disproportionately large effect that is not captured by the arithmetic mean but is captured by the geometric mean. Consequently, a single number, call it N_e , can be used to represent population size even when size fluctuates.

{End this section with a historical note on Brownian motion, ref general idea of stochastic processes, and general parameterization of BM in particular}

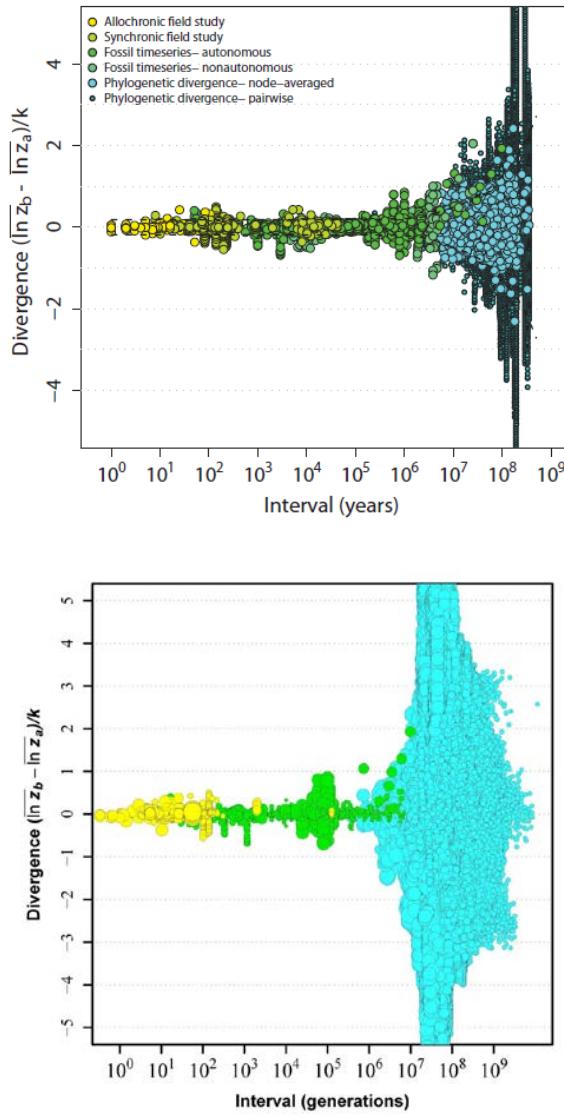
7.4 Tests for neutral evolution with large data sets

... tests on long time scales ... Lynch 1990 ... test for neutrality with Gingerich data (Estes & Arnold 2007)

... Fig. 7.2 (based on neutrality figure in Estes & Arnold 2007 spreadsheet).

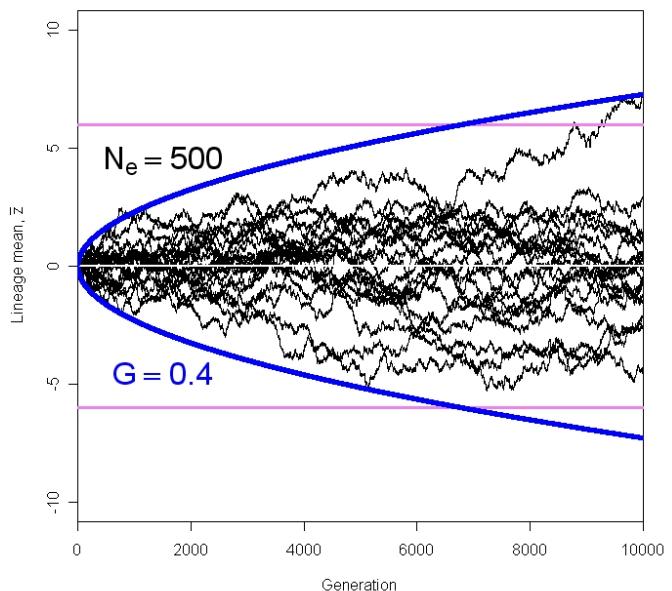
Fig. 7.x The relationship between evolutionary divergence and elapsed time. Divergence is measured as the difference between log-transformed size measurements in two populations, standardized by the dimensionality of the measurements. The interval on the x-axis is the time separating the two samples. (a) Divergence as a function of elapsed time measured in years.

(b) Divergence as a function of elapsed time measured in generations. From Uyeda et al 2011



The problem that genetic drift faces as a model of evolution can be seen as soon as we graph the process on long time scales. For example in Figure 7.5 we examine the consequences of drift on a moderately long time scale, using the same parameters that we employed in Fig. 7.0c on a short time scale. Although the process appeared restrained on a time scale of hundreds of generations, when we expand the time scale to thousands of generations, a substantial proportion of lineages diverge beyond the limits observed in large data sets (Fig. 7.5). On a time scale of millions of generations, this problem is exacerbated. Even populations with relatively large population sizes ($N_e = \text{xxxxxx}$) and minuscule genetic variances ($G = \text{xxxx}$) can drift beyond observed limits after 1-100 million generations (Estes & Arnold 2007).

Figure 7.5. A simulation of drift in 20 replicate populations for 10,000 generations based on 7.01 and 7.02. The blue lines show the theoretical 99% confidence limits for the mean of lineage means. The violet lines show the limits of phenotypic divergence that are observed on time scales less than 1 million generations (Estes & Arnold 2007, Uyeda et al. 2011), namely, less than ± 6 within-population phenotypic standard deviations, \sqrt{P} .



7.5 Drift on a phylogeny

Another simplification in (7.02) is that we imagined that the replicate lineages diverge from a single common ancestor at generation 0. In other words, the replicates evolved on a socalled star phylogeny in which all the branch lengths t generations. In applying neutral theory to evolving populations we will often want to account for a more complex phylogeny.

Independent contrasts represent one way to take account of phylogeny (Felsenstein 1985). To apply this method, one converts data at the tips into contrasts (differences between means) that can be assumed to be independent if the stochastic process of trait evolution on the tree is Brownian motion (e.g., drift, as characterized above). For the tree shown in Fig. 7.3, the independent contrasts would be $(\bar{z}_b - \bar{z}_c)$ and $(\bar{z}_a - \bar{z}_{bc})$, where \bar{z}_{bc} is the trait mean of the common ancestor of b and c , which we will call bc . In data analysis, under the assumption of Brownian motion, these two contrasts are treated as independent observations of evolutionary outcome and scaled so that they have same expected trait variance. An alternative approach avoids the process assumption of independent contrasts as well as the need to reconstruct the means of common ancestors. In this alternative approach, we reconstruct the trait variance-covariance matrix for taxa evolving on the tree. This matrix could be based on Brownian motion or any other well-characterized stochastic process. In other words, we derive expressions for the off-diagonal elements of this matrix rather than transforming to contrasts so that they are zero (independent). A general solution to those elements has been derived by Hansen & Martins (1996).

Hansen & Martins (1996) have shown that the expected covariance of the trait means of two taxa evolving according to a stochastic process on a tree is equal to the expected trait variance of their most recent common ancestor. Suppose that ancestor is t generations removed from the ancestor of the tree. Then, if the stochastic process is drift, the trait variance of the common ancestor is the variance among replicates of that common ancestor evolving by drift for t generations is given by (7.02). We can use this

result to calculate the trait variance-covariance matrix for all of the taxa represented on the tips of the branches of a phylogeny.

To get an expression for that variance-covariance matrix, which we will call \mathbf{A} , we proceed in two steps. First, we represent the phylogeny as a matrix that gives the elapsed time from the tree ancestor to the tip taxa and to all of their common ancestors. Suppose there are r tip taxa. We can represent the phylogeny of those taxa with an $r \times r$ matrix \mathbf{T} , in which the off-diagonal ij th element is the time (in generations) from the root of the tree to the most recent common ancestor of taxon i and taxon j , and the diagonal elements are the time from the root to extant taxon i . \mathbf{T} is a matrix of shared ancestry in the sense that the elements represent the number of elapsed generations in which pairs of taxa experienced shared ancestry (Martins 1995, Hansen & Martins 1996). For example, the phylogeny of three taxa (a, b, c) in Fig. 7.3 can be represented by

$$\mathbf{T} = \begin{bmatrix} T_{aa} & T_{ab} & T_{ac} \\ T_{ba} & T_{bb} & T_{bc} \\ T_{ca} & T_{cb} & T_{cc} \end{bmatrix} = \begin{bmatrix} 1000 & 200 & 200 \\ 200 & 1000 & 800 \\ 200 & 800 & 1000 \end{bmatrix} \quad (7.04)$$

In the second step, we describe the outcome of drift on a phylogeny from generation 0 to present time, at the tips of the branches, by producing an $r \times r$ matrix analog of (7.03), viz.

$$\mathbf{A} = \frac{1}{N_e} G \mathbf{T}, \quad (7.05)$$

where G represents the average additive genetic variance for the trait across time and across populations, N_e is a similar average for effective population size, and \mathbf{T} is the $r \times r$ matrix of shared ancestry (Hansen & Martins 1996, Hohenlohe & Arnold 2008). The off-diagonal, ij th element of \mathbf{A} is the expected covariance in mean trait values between two taxa, i and j , with a shared coancestry equal to the ij th element of \mathbf{T} . Similarly, the diagonal, ii th element of \mathbf{A} is the expected variances in mean trait values after an elapsed time given by the ii th element of \mathbf{T} . For example, the \mathbf{A} -matrix for drift of a trait on the three taxon phylogeny shown in Fig. 7.3, with $N_e=100$, is

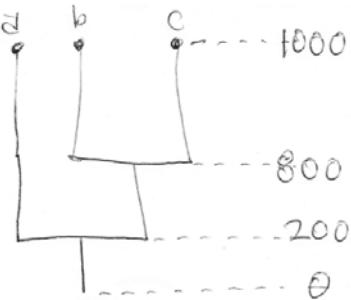
$$\mathbf{A} = \begin{bmatrix} (1000/100)G & (200/100)G & (200/100)G \\ (200/100)G & (1000/100)G & (800/100)G \\ (200/100)G & (800/100)G & (1000/100)G \end{bmatrix} = \begin{bmatrix} 10G & 2G & 2G \\ 2G & 10G & 8G \\ 2G & 8G & 10G \end{bmatrix} \quad (7.06)$$

In other words, if a series of replicate populations of taxon a evolved by drift on the phylogeny, the expected variance among their trait means would be ten times the genetic variance of the trait, $10G$. If a series of replicate pairs of taxa a and b evolved by drift on the phylogeny, the expected covariance of their trait means would be $2G$.

With replicate taxa that have evolved to the tips of a phylogeny given by \mathbf{T} , the distribution of traits means, \bar{z} , is normal,

$$\Phi(\bar{z}) = \frac{\exp[-\frac{1}{2}(\bar{z} - \bar{z}_0)^2 A^{-1}]}{\sqrt{(2\pi)^r |A|}}, \quad (7.07)$$

with a mean the same as in the ancestral population, \bar{z}_0 , and a variance-covariance matrix given by A .



7.6 Tests of neutral evolution on trees

... large literature from comparative studies {the problem we confront here is that nearly all of these studies adopt a model of neutral evolution (Brownian motion) without testing it; might be able to use published values of the fitted drift variance parameter to show that they are unrealistic; see Harmon et al. 2010} ... {the issue of rate parameter varying across branches; perhaps a summary table for this section}

7.7 Mutation-migration-drift balance in a metapopulation.

The complications of phylogeny disappear when we consider a metapopulation subdivided into a series of populations that exchange migrants each generation. The ancestry of the populations becomes immaterial because we assume that the input of variation from mutation and the tendency to homogenize population differences by migration have reached a steady state. Once the metapopulation has equilibrated, how will the total genetic variation in a selectively neutral trait in this metapopulation be apportioned among its constituent populations? We can get a useful answer this question, by making a series of simplifying assumptions. Let us assume that there are n populations each of effective size N and that the rate of migration between populations (the fraction of individuals that move from one population to another population each generation), m , is the same in both directions between all pairs of populations. We will also assume our standard additive genetic model for inheritance and that mutation contributes a constant amount, U , to the genetic variance in each population each generation. Under these conditions, Lande (1992; reconcile the results that follow with Whitlock 1999 Genet Res which are general for migration and case in terms of F_{ST}) has shown that the genetic variance within (G) and among populations (G_a) equilibrates so that

$$\hat{G} = 2nNU \quad (7.08)$$

and

$$\hat{G}_a = \frac{n-1}{m}U \quad (7.09)$$

Twice the genetic variance within populations is converted into genetic variance among populations by the process of random drift (Wright 1951, 1969). Consequently, the total genetic variance in the metapopulation is $G_a + 2G$ rather than $G_a + G$. Several important conclusions can be drawn from these equilibrium results. First, the migration rate, m , does not affect the level of genetic variance within populations. This results contradicts our intuition that migration should enhance variation within populations, but we have assumed that equal numbers of migrants enter and leave each population. Second, the level of genetic variation maintained among populations is enhanced by increasing the number of populations, eroded by increasing migration rate but is not affected by population size, N . The rates of approach to these equilibria (not shown) are negative exponential functions, so that the approach is rapid at first, following by a slow, asymptotic period. With a large number of populations, that final approach takes a few times $n/2m$ or $2nN$ generations, whichever is larger (Lande 1992). So, for 20 populations of effective size 500, exchanging $mN=2$ migrants per generation ($m=1/250$), several thousand generations might be required for the system to equilibrate in mutation-migration-drift balance {would be easy to tailor these numbers to fit the elegans metapopulation of Manier et al 2007!}.

It is useful to express the among-population genetic variance as a fraction of the total,

$$Q_{ST} = G_a / (G_a + 2G), \quad (7.10)$$

a fraction comparable to Wright's F_{ST} , a measure of population differentiation in gene frequencies (Wright 1969, Lande 1992, Spitzke 1993). ... (Fig. 7.4) = Multiple panel figure showing within- and among-population distributions of genetic values as a function of Qst . When the two parts of variation in a quantitative trait are estimated by a two-level analysis of variance, G_a and G are, respectively, the among- and within-population components of genetic variance.

{add material from more recent paper by Whitlock}

7.8 Tests for neutral evolution in metapopulations

.... tests for neutral evolution of garter snake coloration and vertebral numbers (Manier et al. 2007) – report in text just the results of the Fst/Qst comparison without going into the 3-level testing, don't attempt to produce a table, again too complicated; present and discuss Fig. 7.5 = Distribution of Fst/Qst results, produced using summary table in review(s) by Merila et al. = just Fst/Qst ... review of Fst/Qst studies; need to find most recent review, see Manier et al. 2007 for ref to review by Merila et al ;

Chapter 11: Evolution of a Single Trait on a Stationary Adaptive Landscape

© Stevan J. Arnold 2016

Overview.- The adaptive landscape is a powerful conceptual tool for modeling long-term responses to selection. In the simplest case, we imagine a set of replicate populations that descend all at once from a common ancestor and thereafter evolve on identical adaptive landscapes. As a general rule, variation among replicate means is inversely proportional to effective population size. The smaller the size of the replicates, the more variable their response to selection. A \cap -shaped adaptive landscape results in a stable equilibrium distribution of replicate means. Stabilizing selection tends to pull means towards an optimum, but that tendency is balanced by drift away from the optimum. Tests with a large data set indicate that this Ornstein-Uhlenbeck (OU) model is undoubtedly an element in the general explanation for stasis, but other factors are also in play.

11.0 Tendency to evolve uphill on the adaptive landscape

$$\Delta\bar{z} = G\beta = G \frac{\partial \ln \bar{W}}{\partial \bar{z}}$$

... Fig. 11.0 ...

11.1 Stochastic dynamics and equilibrium on a landscape with a single, stationary adaptive peak
 Gaussian adaptive landscape ... linear restoring force of selection towards the optimum proportional to the distance from the optimum Fig. 11.1 = Rubberband cartoon of the OU process **before this

figure need to specify that restoring force as $\frac{(\bar{z} - \theta)}{\omega + P}$ and identify it as the force of directional

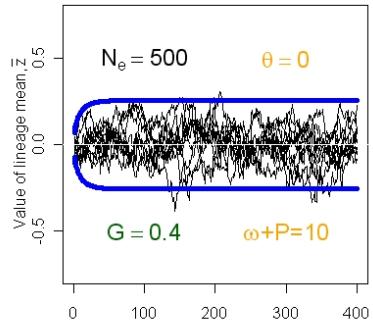
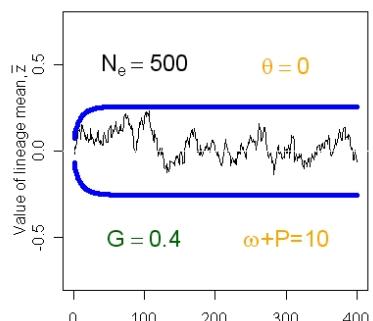
selection back towards the optimum, θ , is $\beta = \frac{(\bar{z} - \theta)}{\omega + P}$ **... replicate populations of size N_e ... Ornstein-

Uhlenbeck (OU) process (Ornstein & Uhlenbeck 1930) ... each generation the trait mean drifts by an amount equivalent to a draw from a normal distribution with mean zero and variance G/N_e and is pulled back towards the optimum with a force, β , that is proportional to the deviation from the optimum. In other words,

$$\bar{z}(t+1) = \bar{z}(t) + \frac{(\bar{z}(t) - \theta)}{\omega + P} G + N(0, G/N_e)$$

... let the position of the optimum be zero ... after t elapsed generations, the expected value of the mean of lineage means is zero with lineage means normally-distributed about that expectation with a variance of

$$Var(\bar{z}_t) = \frac{\omega + P}{2N_e} \left\{ 1 - \exp \left[-2 \left(\frac{G}{\omega + P} \right) t \right] \right\}.$$



An example of a lineage drifting about its optimum is shown in Fig. 11.3a. Each time the trait mean drifts away from the optimum, it is

pulled back, so ,as a long term average, the mean resides at the optimum. An ensemble of 10 replicate populations, buzzing about identical optima are shown in Fig. 11.3b. The outer limits of their paths are accurately described by the 99% confidence limits (shown in blue).

Fig. 11.3. Simulations of replicate populations evolving about a stationary intermediate optimum. (a) A single lineage mean in a population of finite size evolving according to an O-U process. (b) The lineage means of 10 replicate populations evolving according to the same process.

At equilibrium, trait means are normally distributed with a mean of means at the optimum, θ , and a variance of lineage means given by

$$\text{Var}(\bar{z})_{\infty} = \frac{\omega + P}{2N_e},$$

so that the equilibrium distribution of means is given by

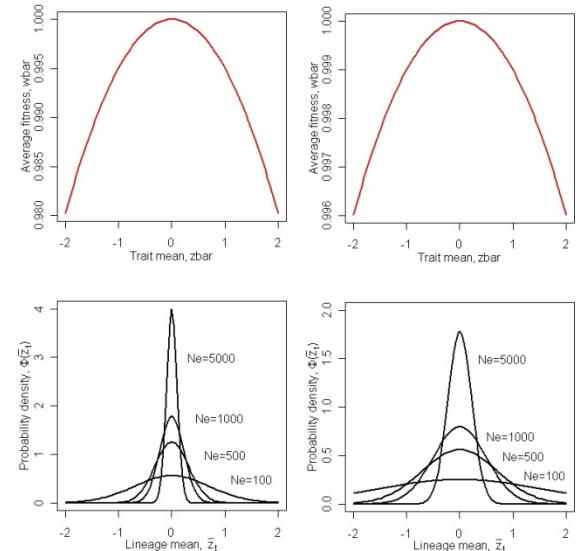
$$\Phi(\bar{z})_{\infty} = \frac{1}{\sqrt{2\pi\text{Var}(\bar{z})_{\infty}}} \exp\left\{-\frac{(\bar{z}-\theta)^2}{2\text{Var}(\bar{z})_{\infty}}\right\}.$$

Lande (1976). In other words, variation in lineage means is greater if stabilizing selection is weak (large $\omega + P$) and populations effective size is small (Fig. 11.2). Smaller populations can drift further from the optimum, especially if the restraining force of selection is weak.

Figure 11.2 Equilibrium distributions of lineage means under weak stabilizing selection. The upper panels show the adaptive landscape (note change in vertical scale from left to right). The lower panels show equilibrium distributions of lineage means, $\Phi(\bar{z}_{\infty})$, as a function of effective population size, N_e . (a) Weak stabilizing selection, $\omega + P = 100$. (b) Very weak stabilizing selection, $\omega + P = 500$.

On a more quantitative level, these equations show that stabilizing selection must be extremely weak to produce appreciable variation in lineage means at equilibrium (Fig. 11.2). For example, even a restraining force as weak at $\omega + P = 500$, dispersion of lineage means is small, so that even with an effective size of 500, the standard deviation of lineage means is less than 1 (compare with the drift distributions in Fig. 7.1).

... early-burst model (Blomberg et al. 2003), also known as the ACDC model ... variance among lineages decreases through time to account for bursts of diversification early in adaptive radiations ... at rate r ($r < 0$) ...



$$\bar{z}(t+1) = \bar{z}(t) + \frac{(\bar{z}(t) - \theta)}{\omega + P} G + N(0, \text{var})$$

where $\text{var} = (G/N_e)\exp\{rt\}$ and $r < 0$ is a parameter describing the diminution of drift variance with time (Harmon et al. 2010). The expected value of the lineage means is θ with normally-distributed variation about that value so that variance at generation t is

$$\text{Var}(\bar{z}_t) \approx \frac{\omega + P}{2N_e} \left\{ 1 - \exp \left[-2 \left(\frac{G}{\omega + P} \right) t \right] \right\} [\exp\{0.7rt\}] .$$

From this equation we see that the variance among replicate lineages decays exponentially. This equation differs from results given by Blomberg et al. (2003) and Harmon et al. (2010) by incorporating realistic parameters for drift and stabilizing selection. Taking the same parameters for population size, selection, and inheritance as in the last example, Fig. 11.8 portrays an example of results from the early-burst model. As expected, the model produces about as much among-lineage variation as the simple OU model from generations 0 to about 20, but after that variance rapidly decays so that after just a few hundred generations, all replicate lineage mean reside almost exactly at the optimum. These features persist so long as r is negative and different from zero, although as r approaches zero, (xEB var = last expression) converges on (xOU var).

In comparing the pattern in Fig. 11.8 with the pattern of data on long time scales (Fig. 7.x), we must remember that the onset of radiations are not synchronized to occur at the beginning of time intervals, as they are in the simulations. Instead, according to the model, we might expect bursts of diversification, followed by exponential decay, to periodically occur in the data. We must also realize that the early-burst model produces a radiation of modest proportions (i.e., far less than $\pm 6\sqrt{P}$) followed by collapse towards virtually no variation about the optimum. Thus, the early-burst model, unless supplemented by peak movement, cannot produce the full width of the long-term band apparent in Fig. 7.x and totally fails to produce the surge in diversification beyond $\pm 6\sqrt{P}$ that is apparent after long time intervals. Not surprisingly, Harmon et al. (2010) found that the early-burst model produced worse fits to data than Brownian movement and an OU process.

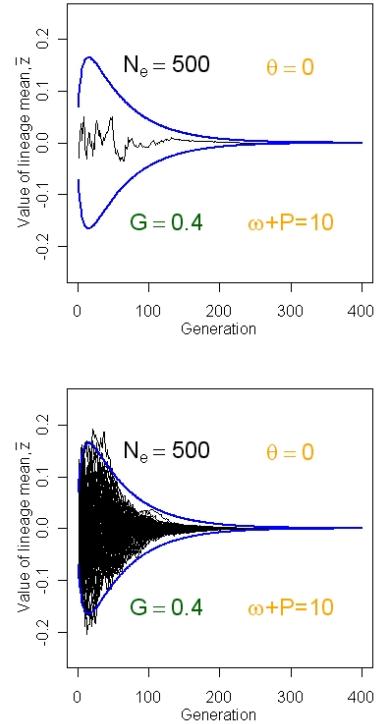


Fig. 11.8. Simulations of lineages evolving according to the early-burst model of Blomberg et al. (2003). Conventions as in Fig. 11.2. The parameter describing diminution of drift variance, r , is set to -0.05. (a) A simulation of one evolving lineage. (b) 100 replicate lineages.

11.2 Exploration of a stationary landscape by drift

... interaction of drift and selection ... modeling the change in the distribution with a diffusion equation ... the change in the distribution of replicate population means through time is given by

$$\frac{\partial \Phi}{\partial t} = -\frac{\partial}{\partial \bar{z}_t} \left\{ -\frac{G}{(\omega + P)} \bar{z}_t \right\} + \frac{\partial^2}{\partial \bar{z}_t^2} \left\{ \frac{G}{N_e} \right\}$$

where the terms in brackets are, respectively, the mean and variance of the transition probability (per generation) density function for $\Phi(\bar{z}_t)$ (Lande 1976).

... at equilibrium

$$\Phi(\bar{z})_\infty \propto \bar{W}^{2N_e}$$

... expected first passage time

Lande (1976).

{refer back to Fig. showing sample univariate paths of \bar{z} on a univariate Gaussian AL}.

11.3 Attempts to explain time series and adaptive radiations with a single stationary peak

... Estes & Arnold 2007 ... Gingerich data and a single peak (Fig. 11.4 shows tests from Estes & Arnold 2007 spreadsheet) ... tree-based data (Hansen 1997; Butler & King 2004, OUCH; Harmon et al 2010; the necessity of assuming very small N_e and/or extremely weak stabilizing selection;

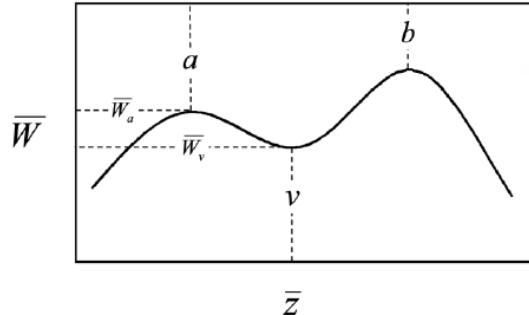
11.4 Shifting of the trait mean between two adaptive peaks

... Fig. 11.5 from Lande 1985 or 6

$$T \cong \frac{2\pi(\omega + P)}{G} \left(\frac{\bar{W}_a}{\bar{W}_v} \right)^{2N_e}$$

$$E[\bar{z}_t] = d(1-p)$$

$$Var(\bar{z}_t) = p\bar{z}_t^2 + (1-p)(d - \bar{z}_t)^2 + \frac{\omega + P}{2N_e}$$



... the first term on the right ... the second term ... the third term ... (Lande 1985, 1986).

... Estes & Arnold (2007) tests with the Gingerich data
... need to invoke unrealistically small N_e

Figure 5: Adaptive landscape in the peak shift model, which has two adaptive peaks, a and b . Mean fitness \bar{W} is a function of the average phenotype in the population, \bar{z} . Two critical parameters in the peak shift model that affect the probability that the population mean will shift from peak a (left) to peak b (right) are the height \bar{W}_a of the first adaptive peak and the height \bar{W}_v of the valley v separating the first adaptive peak from the second. The distance between the two peaks (between points a and b) is d .

11.5 Geographic variation

Geographic variation refers to among-population variation within a species. Such variation is an almost ubiquitous feature of most species, although the extent of geographic variation varies from trait to trait (Mayr 1963). ... example plots from James' study of body size in NA birds ... an intermediate optimum that varies in space results for clines

Chapter 12: Evolution of Multiple Traits on a Stationary Adaptive Landscape

© Stevan J. Arnold 2016

Overview.- The theoretical evolution of the multivariate mean has been most explored in the case of a hill-shaped adaptive landscape that has a consistent position, shape and orientation. Under certain conditions, the mean tends to evolve uphill on this landscape, towards the adaptive peak. ... evolution of two genetically correlated traits Theory is especially well developed for systems in which a single male trait evolves in response to sexual selection exerted by a second, female trait (sexual preference). Under these conditions, the equilibrium of the bivariate mean may be a stable point, a line or an elliptical cycle. Unstable dynamics are also possible, but probably unlikely. Finite population size adds an elements of uncertainty to the outcome, changing stable points into clouds, for example. Predictions from these models are consistent with various features of populations with sexually-selected traits, but discriminating, quantitative tests have not yet been accomplished.

{in this introduction need to bridge from the generally pessimistic results of empirical tests with single traits, preceding chapter, to this chapter ... single peaks are an element in the explanation of adaptive radiations but not the whole picture ... }

12.0 The tendency to evolve uphill on the adaptive landscape

$$\Delta \bar{z} = G\beta = G \frac{\partial \ln \bar{W}}{\partial \bar{z}} \quad (12.00)$$

... Mahalanobis distance in phenotypic space ... $[(\bar{z}_a - \bar{z}_b)^T P^{-1} (\bar{z}_a - \bar{z}_b)]^{\frac{1}{2}}$... *generalized phenotypic distance* between the means of two populations, a and b , ... accounts for covariances among traits and differences in variance ... for generalized *genetic* distance, replace P with G ... the square of generalized genetic distance is always positive and equivalent to the change in log mean fitness (Lande 1979),

$$\Delta \ln \bar{W} \cong \Delta \bar{z}^T G^{-1} \Delta z \geq 0 \quad (12.01)$$

... consequently we can conclude that ... (Fig. 12.0 = figure from chapter in the Boake book).

12.1 The spaceship model

The case of two traits evolving on a stationary adaptive landscape illustrates many properties of the general multivariate case, so we will use it as a starting point. In the case of a single trait, the process under discussion consists of the boring march of the frequency distribution, disparaged by Luria et al. 1981 (*A View of Life*), but in the two trait case, the process is much more interesting. In particular, we shall consider the process when the bivariate means lies some considerable distance from the optimum. Under this circumstance we will see that the process consists of two distinct phases: an initial phase during which the mean evolves rapidly towards the optimum, and a second phase, often with a distinctly different trajectory, during which the mean slowly approaches the optimum. ... deterministic response The per generation change in the trait mean when the AL is Gaussian is a special case of the general expression, $G\beta$,

$$\Delta \bar{z} = G\beta = G(\omega + P)^{-1}[\theta - \bar{z}_t]$$

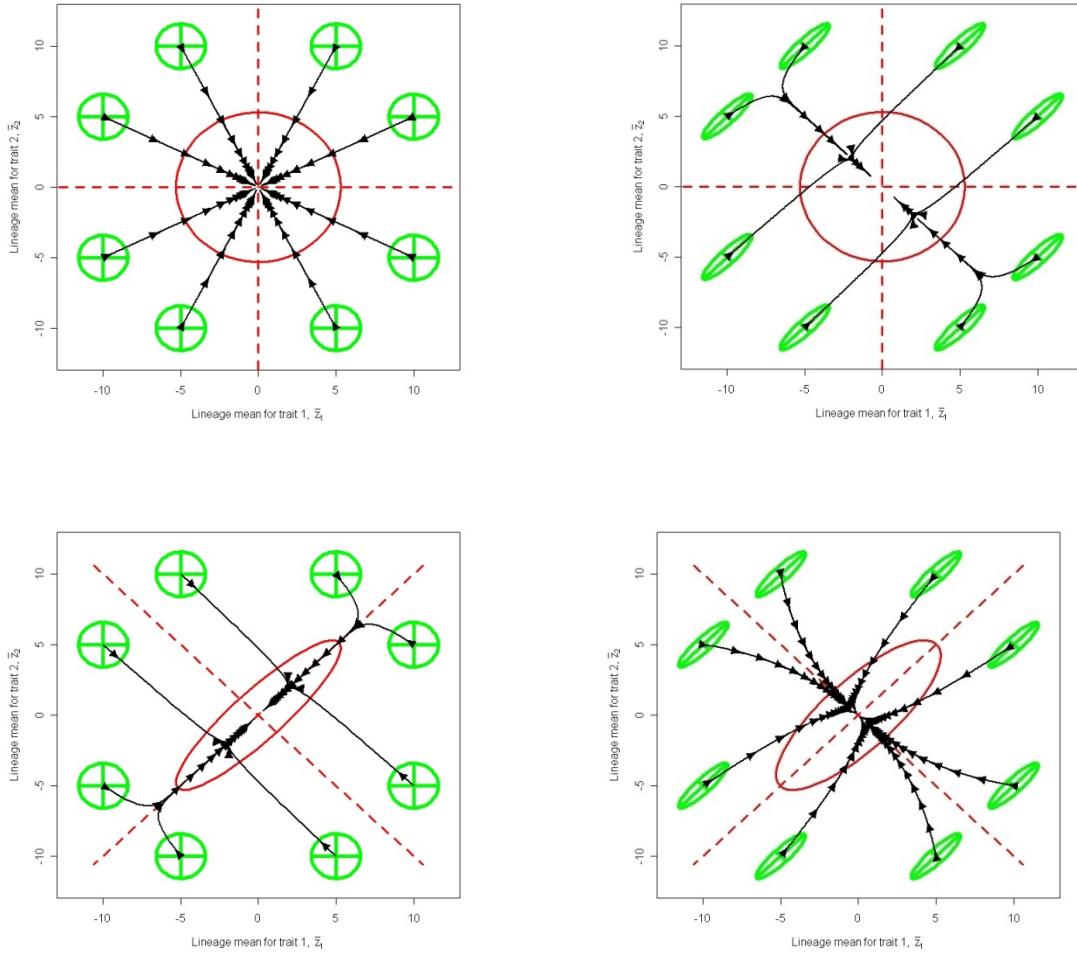


Figure 12.1 Bivariate evolution on a Gaussian adaptive landscape, the spaceship model. The adaptive landscape is shown in red: the ellipse represents the equivalent 50% confidence region, with eigenvectors shown as dashed lines. Stabilizing selection of equal magnitude acts on each trait: $(\omega + P)_{11} = (\omega + P)_{22} = 50$. G-matrices (represented as 95% confidence ellipses) are shown in green. In each case, the genetic variance of both traits is 0.4. Evolutionary trajectories of the bivariate mean are shown in black, with arrows at regular intervals of elapsed generations. (a) $r_g = 0, r_s = 0$. Each evolutionary trajectory is 500 generations in duration, with arrowheads every 50 generations. (b) $r_g = 0.9, r_s = 0$. Each evolutionary trajectory is 1500 generations in duration, with arrowheads every 300 generations. (c) $r_g = 0, r_s = 0.9$. Each evolutionary trajectory is 500 generation in duration, with arrowheads every 50 generations. (d) $r_g = 0.9, r_s = 0.8$. Each evolutionary trajectory is 700 generations in duration, with arrowheads every 50 generations.

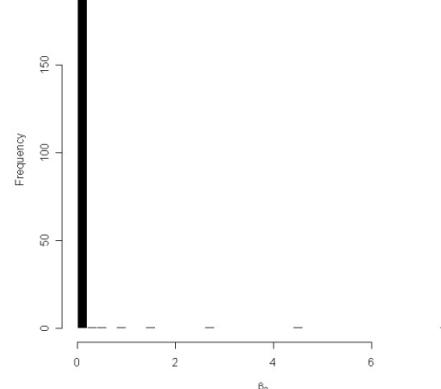
It will be useful to remember the two trait version of this last equation as the ‘spaceship model’. Think of a bivariate mean located some considerable distance from a single, stationary adaptive peak. The mean is surrounded by a cloud of genetic values that can be visualized as a green ellipse, our spaceship. In a thought experiment, we keep the overall configuration of the cloud constant as the bivariate mean responds to selection exerted by the distant peak and moves toward it. The particular trajectory that the spaceship takes on its journey is a function of its starting position, as well as the configurations of the genetic cloud and the AL (Fig. 12.1)

Fig. 12.1 illustrates three important properties of the lineage mean’s response to an distant adaptive peak. First, responses are greater when the mean is further from the adaptive peak. This greater response reflects the fact that directional selection is stronger, the further away the mean is from the optimum. Second, genetic covariance and correlational selection can each cause evolutionary trajectories to be curved. Third, the evolving mean can over- or under-shoot the optimum so that the population is subjected to long periods of maladaptation (Fig. 12.1b, c). This phenomenon of maladaptation is likewise a consequence of genetic covariance and correlational selection. We will now examine these aspects of the spaceship’s trajectory and journey in greater detail.

We also see in Fig. 12.1 that the shapes of the G-matrix and the AL can dramatically affect the evolutionary trajectory of the mean during both the initial and secondary phases of approach to the optimum. In the absence of both genetic correlation and correlational selection, both the cloud of genetic values and the AL are circular in cross-section. Under this circumstance (Fig. 12.1a [currently named 12.1c]), the mean evolves directly towards the adaptive peak. Evolution is rapid at first and then gradually decelerates, until the mean ceases its evolution at the peak. If the AL is circular but the G-matrix is elongate, the rapid initial phase is in the direction of g_{max} , the long axis of the G-matrix (Fig. 12.1b [12.1b]). However, when the mean approaches g_{min} , the short axis of the G-matrix, the trajectory turns in that direction and evolution decelerates dramatically as the mean completes its slow approach to the optimum. The difference in initial and secondary phases makes intuitive sense: a rapid initial phase fueled by abundant genetic variance, then a slow secondary phase in the direction of least genetic variance. This intuition is expanded when we consider a third case, in which the G-matrix and the AL have reverse configurations (Fig. 12.1c [12.1a]). Now the genetic cloud is circular and impotent, and the elongate AL governs the trajectories of approach to the optimum. The rapid initial phase is along ω_{min} , the short axis of the AL, and the slow, second phase is along ω_{max} , the long axis of the AL. When genetic variance is the same in all directions, the shape of the AL dictates the difference in trajectories during the initial and secondary phases of evolution. Finally, in a revealing but unillustrated case, consider the outcome when the G-matrix and the AL are perfectly aligned ($g_{max} = \omega_{max}$) and proportional (i.e., the corresponding eigenvalues of the two matrices differ by the same constant of proportionality). Under this circumstance of alignment and proportionality, the effects of the two process cancel and the trajectories of the mean assume the star-like pattern seen in Fig. 12.1a. In retrospect, we could have anticipated this result from (12.x): since $cG = (\omega + P)$, where c is a scalar constant, $cG(\omega + P)^{-1}$ is cI , where I is the identity matrix. Because this special case requires a precise combination of parameters, a more general case is illustrated in Fig. 12.1d. Here, G and the AL are aligned but not exactly the same shape, with the result that both inheritance and selection affect the trajectories of the mean.

... {some quantitative statements about how long the first and second phases last} ...

A arbitrary feature of these illustrations of the spaceship model is that we began each trajectory with the

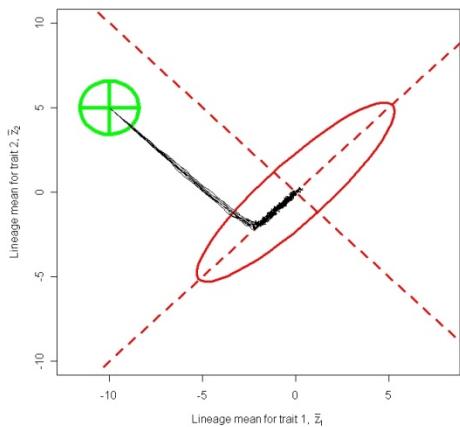


mean situated far from the optimum. How commonly does the trait mean reside 5-10 standard deviation away from the optimum? Under the Gaussian assumptions used in Fig. 12.1, the magnitude of the selection gradient, β , is directly proportional to the distance to the optimum and inversely proportional to the curvature of the AL, $(\omega+P)$. Consequently, we can produce a histogram of β -values for any of the trajectories shown in Fig. 12.1 and compare that histogram to the histogram of β -values estimated in natural populations (Fig. 2.4c). A representative histogram from a trajectory in Fig. 12.1 is shown in Fig. 12.2. That trajectory produces a few values of β that are much larger than those typically seen in natural populations and of course these values were produced during the first few generations of the first phase, when the mean was far from the optimum. The histogram also shows an overabundance of values very close to the optimum. In other words, the process of deterministic approach to an optimum from far away, as in Fig. 12.1, is not consistent with the picture of evolution implied by direct estimates of directional selection. ... this result highlights a problem with evolution along genetic lines of least resistance as an explanation for adaptive radiations ...

So far we have considered the deterministic evolution of the bivariate mean. Finite population size will cause stochasticity in the sampling of parents each generation and produce a distribution of trajectories. One such case is shown in Fig. 12.3. In general, as in this case, stochasticity is tempered by directional and stabilizing selection, so that the variation among lineage means is considerably less than in the neutral case.

Figure 12.3 Stochastic response of the bivariate mean to a Gaussian adaptive landscape for 500 generations in 8 replicate lineages. No genetic correlation, $r_s=0.9$ and $N_e=500$. Other conventions as in Figure 12.1 except that elapsed intervals are not indicated with arrowheads.

{a final paragraph for this section pointing out that even the 2-trait portrayal has implications for multiple trait evolution that go far beyond the specifics of sexual dimorphism ... general portrayal of the role of genetic covariance in evolution ... the real possibility of temporary but long lasting maladaptation ...}



12.2 The evolution of sexual dimorphism

Following Lande's (1980b) treatment of this problem, we will assume that multiple autosomal loci affect expression of the trait differently in the two sexes. Let z and y be normally distributed traits in males and females, respectively, of a dioecious species with within-sex genetic variances G_m and G_f , and a between-sex genetic covariance of B . The per generation evolution of the means, \bar{z} and \bar{y} , is a special case of (10.01),

$$\begin{bmatrix} \Delta\bar{z} \\ \Delta\bar{y} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} G_m & B \\ B & G_f \end{bmatrix} \begin{bmatrix} \beta_m \\ \beta_f \end{bmatrix} = \frac{1}{2} \begin{bmatrix} (G_m\beta_m + B\beta_f) \\ (B\beta_m + G_f\beta_f) \end{bmatrix}, \quad (12.02)$$

where the factor of $\frac{1}{2}$ accounts for sex-limited expression; β_m and β_f are the selection gradients acting on the two sexes. { diagram, instead of equations for the two-sex adaptive landscape, illustrating the two thetas, two omegas + Ps } We assume that the two traits are under Gaussian viability selection towards intermediate optima (θ_m and θ_f). Each generation viability selection is followed by a process of

assortative mating and frequency-dependent sexual selection. The total selection gradients in (12.02) are composed of two terms, representing the directional forces of viability and sexual selection.

$$\begin{bmatrix} \beta_m \\ \beta_f \end{bmatrix} = \begin{bmatrix} \frac{(\theta_m - \bar{z})}{(\omega_m + P_m)} + c_m \\ \frac{(\theta_f - \bar{y})}{(\omega_f + P_f)} + c_f \end{bmatrix} \quad (12.03)$$

If sexual selection does not affect mean fitness and only affects the distribution of offspring among the individuals of each sex, the adaptive landscape for the two sexes is defined solely by natural selection. For simplicity, we will assume that the forces of sexual selection acting on the two sexes, c_m and c_f , are constant from generation to generation.

Evolution ceases at an equilibrium given by

$$\bar{z} = \theta_m + (\omega_m + P_m)c_m \quad (12.04a)$$

$$\bar{y} = \theta_f + (\omega_f + P_f)c_f, \quad (12.04b)$$

where P_m and P_f are the within-sex phenotypic variances, while $(\omega_m + P_m)$ and $(\omega_f + P_f)$ are the ‘variances’ of the Gaussian adaptive landscape function in the \bar{z} and \bar{y} dimensions. Note that if no individual expresses both z and y , the phenotypic covariance between the two traits is undefined, and there is correlational selection between the two traits. Consequently, the corresponding adaptive landscape lacks tilt because $\omega_{mf}=0$ (Fig. 12.1). We see from (12.04) that the two traits will equilibrate on their adaptive peaks if there is no sexual selection ($c_m=0$ and $c_f=0$) (Fig. 12.1a, b). On the other hand, if sexual selection acts only on males ($c_f=0$), males will be pulled off their adaptive peak at equilibrium, but females will not (Fig. 12.1c,d).

To analyze the evolutionary approach to the equilibrium, it will be useful to consider a special case. For simplicity, let us focus on the stage early in the process of evolving sexual dimorphism, when the variances the two sexes are equal, $P_m = P_f = P$ and $G_m = G_f = G$, and the two sexes experience similar strengths of stabilizing selection, $\omega_m = \omega_f = \omega$. It will also be useful to define the sexual average as $a = \frac{1}{2}(\bar{z} + \bar{y})$ and sexual dimorphism as $d = \bar{z} - \bar{y}$. These new traits, a and d , constitute a 90 deg rotation of the axes in Fig. 12.x. Substituting these definitions of a and d into (12.03) and (12.02), we obtain equations for the evolution of the sexual average and sexual dimorphism

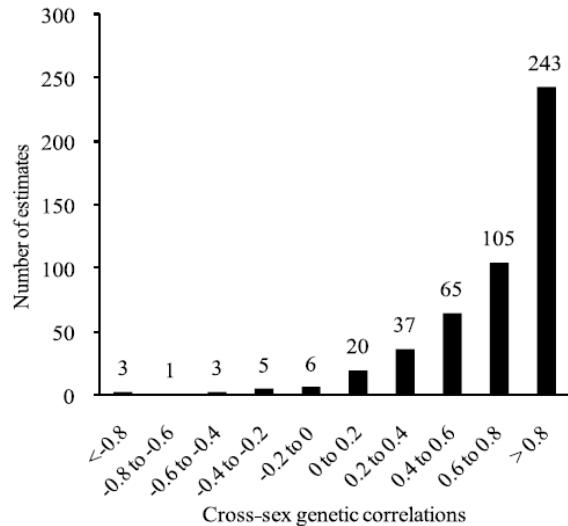
$$\begin{aligned} \Delta a &= \frac{1}{2}(G + B) \left\{ \frac{\frac{1}{2}(\theta_m + \theta_f) - a}{\omega + P} + \frac{1}{2}(c_m + c_f) \right\} \\ \Delta d &= \frac{1}{2}(G - B) \left\{ \frac{(\theta_m - \theta_f) - d}{\omega + P} + (c_m - c_f) \right\}. \end{aligned} \quad (12.05)$$

We see from (12.05) that the genetic covariance between the sexes, B , amplifies evolution of the sexual average but impedes evolution of sexual dimorphism ... {refer to rapid evolution of a followed by slow evolution of d ... departure from optima ... long period of maladaptation}

Empirical studies reveal a striking pattern of high genetic correlations between the sexes for homologous traits ... {Fig 12.4 = Figure from Poissant et al. 2009, n=488 estimates}

{need an empirical example of evolving sex dimorphism; check out Fairbairn & Prestesoi water strider papers ...}

{paragraph pointing out that above is a simplified, 2-trait portrayal of a model that allows multiple traits in each sex ... summarize some of the main conclusions from that more general model}



12.3 The balance between drift and stabilizing selection on a Gaussian landscape with a stationary peak.

We now consider the effect of finite population size in opposing the effects of stabilizing selection ...

... multivariate version of drift-selection balance with a Gaussian surface imposing the selection (section 11.1) ... Hansen & Martins (1997)

Each generation, the overall change in the lineage mean is the sum of two contributions: drift, which tends to move the trait mean away from the optimum, and stabilizing selection, which tends to pull the trait mean back towards the optimum,

$$\bar{z}_{t+1} = \bar{z}_t + G(\omega + P)^{-1}[\theta - \bar{z}_t] + N(0, G/N_e),$$

where the last term represents a random draw from a normal distribution with mean of zero and a variance-covariance matrix of G/N_e . We recognize the second term on the right as a response to directional selection, $G\beta$. At generation t , the lineage means are normally distributed about a mean vector of zeros with a variance-covariance matrix given by

$$Var(\bar{z}_t) = \frac{(\omega + P)}{2N_e} \left\{ 1 - \exp[-2tG(\omega + P)^{-1}] \right\}.$$

The first term involving $\omega + P$ represents the balance between drift and stabilizing selection, while the second represents responses to directional selection. In the last term, {}, we see exponential decay in the influence of the G -matrix, so that the dispersion matrix achieves a limiting value given by

$$Var(\bar{z}_\infty) = \frac{(\omega + P)}{2N_e}$$

Consequently, on long time scales, the probability distribution of lineage means is normal and converges on,

$$\Phi(\bar{z}_\infty) = \sqrt{(2\pi)^{-n} |Var(\bar{z}_\infty)^{-1}|} \exp\{-(\bar{z} - \theta)Var(\bar{z}_\infty)^{-1}(\bar{z} - \theta)\}$$

Notice that while the AL appears in the last two equations, the G-matrix does not, indicating that the long term pattern of lineage means is shaped by the AL, not by the G-matrix.

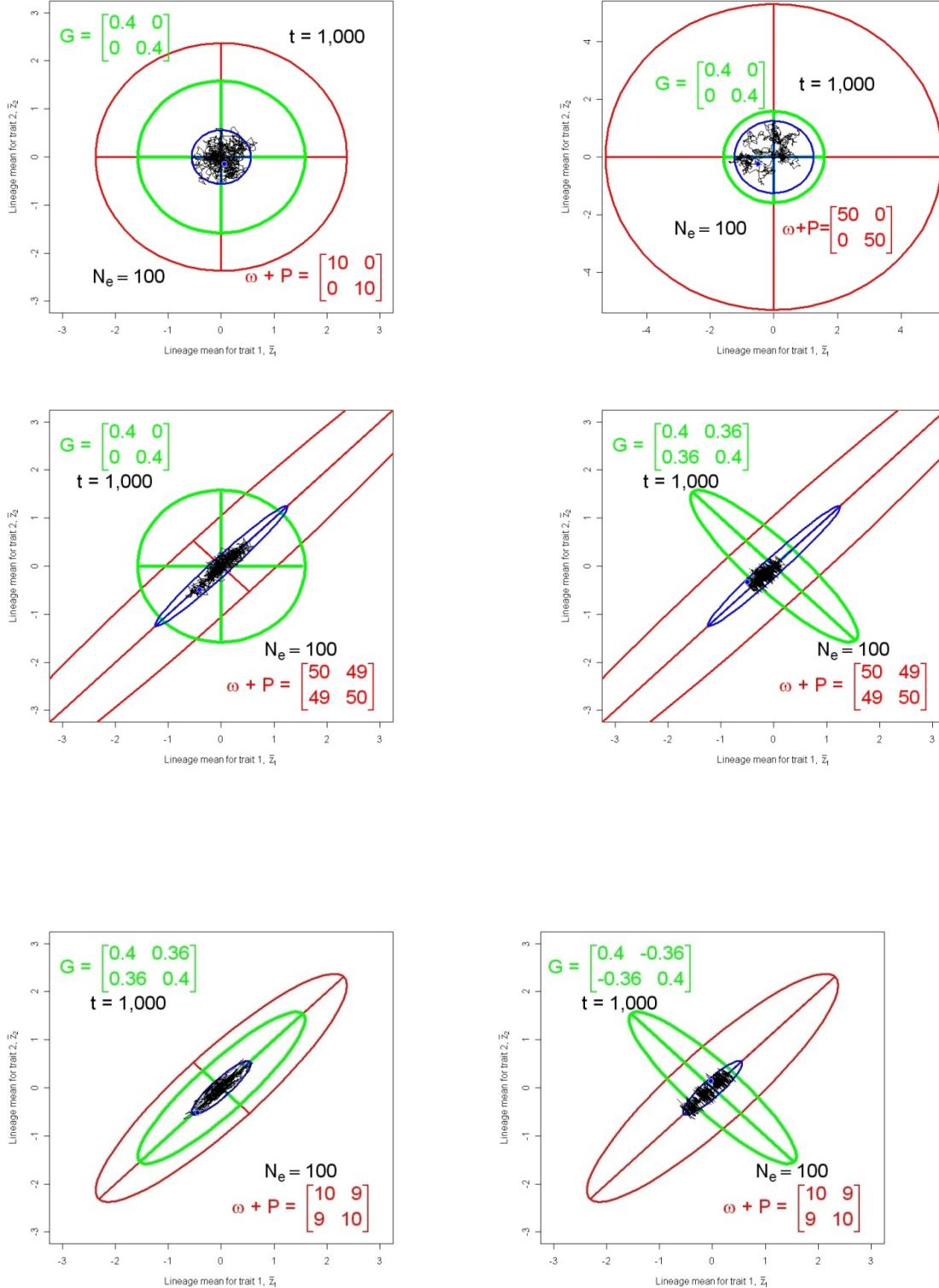
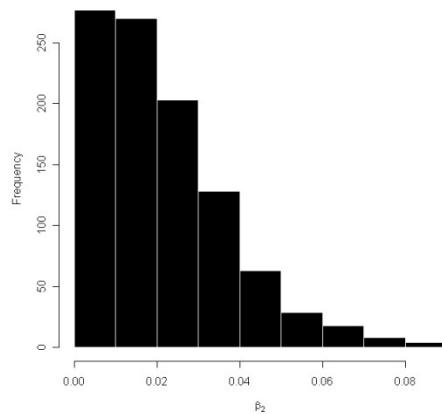


Figure 12.5 {show only the variable parameters on the panels, not the constant ones; ie, no t and N_e } The role of the adaptive landscape in shaping the results of drift-selection balance. The 95% confidence ellipse for genetic values is shown in green and represents the G-matrix. The adaptive landscape is bivariate stabilizing in all cases and is represented by its 25% confidence ellipse in red. Such a low level ellipse must be employed because, the limits of the 95% ellipse would be far outside the scale limits of the figure. The limiting pattern for dispersion of lineage means, $Var(\bar{z}_\infty)$, is represented by its 95% confidence ellipse in blue. In each panel the black line shows the results of 1000 generations of simulated evolution of the lineage mean under the conditions specified in each panel. (e [should be a]) Weak stabilizing selection with no selectional correlation and no genetic correlation. (f [should be b]). Very weak stabilizing selection with no selectional correlation and no genetic correlation. Notice the change in scale. (a [should be c]). Very weak stabilizing selection with strong positive selectional correlation and no genetic correlation. (b [should be d]) Very weak stabilizing selection with strong positive selectional correlation and a strong negative genetic correlation. (c [should be e]) Weak stabilizing selection with strong positive selectional correlation and a strong positive genetic correlation. (d [should be f]) Weak stabilizing selection with strong positive selectional correlation and a strong negative genetic correlation.

The adaptive landscape exerts a strong influence on the outcome of drift-selection balance. Figure 12.2 illustrates the relevant case of bivariate stabilizing selection, showing that even very weak stabilizing selection strongly constrains the extent to which lineage means depart from the immediate vicinity of the adaptive peak. To appreciate the remarkable constraint exerted by the AL compare Fig. 12.2 with Fig. 8.1 which shows the 95% confidence ellipse for lineage means evolving by drift alone. Those ellipses are considerably larger than the comparable ellipse for the G-matrix. In contrast, even very weak stabilizing selection constrains evolving lineage means so that they occupy a bivariate area that is a fraction of the size of the ellipse that represents genetic values dispersed around the bivariate mean (Fig. 12.2a, b). The AL also strongly constrains the shape of the lineage dispersion. Those dispersion patterns (shown as blue ellipses), always mirror the shape of the AL and are completely independent of the shape of the G-matrix, as shown by (12.x). Finally, notice that a particularly favorable case for lineage dispersion is presented in Fig. 12.2 by using a very small effective population size ($N_e = 100$). For more realistic sizes, say N_e in the range 500 to 5,000, the lineage dispersion ellipses would be minuscule and the paths of simulated lineage paths could not be visually resolved. It is no exaggeration to say that, in the absence of peak movement, even very weak stabilizing selection eliminates the possibility of appreciable radiation among lineages.

Is the process of drift-selection balance capable of producing the kind of distribution of directional selection gradients that we see in nature? ...

Fig. 12.6 ... the process produces the right range of β -values and the distribution is almost the right shape ... a shortage of values very close to zero, corresponding to mean very close to the optimum ... despite this discrepancy, the overall distribution is much closer to observation than one produced by approach to the optimum from afar (Fig. 12.2) ...



12.4 Hill climbing with two adaptive peaks

... another role for genetic correlation ... can affect which of two peaks is climbed ... {pull in illustration from Kirkpatrick or Schluter or create one = Fig. 12.7}

12.5 The joint evolution of sexual preference and sexually-selected traits

Darwin (1859, 1874) proposed a special process, sexual selection, to account for the evolution of traits that aid males in obtaining mates by sexual persuasion of females or by physical offense and defense in contests with sexual rivals. In Darwin's view, ordinary viability selection could not account for the elaboration of such traits, and indeed that selective agency would often act in the opposite direction, toward diminution. While sexual preference of females for elaborate male traits seemed inevitable to Darwin, it remained controversial for decades into the 20th century. Fisher (1915, 1930), however, sketched a model for Darwin's process of mate choice that became famous for its prediction of runaway evolution of female preference and male trait. In this section, we summarize Lande's (1981) model of the Darwin-Fisher proposal, with some simplifications in the interest of getting quickly to the main conclusions.

Lande (1981) modeled the joint evolution of two traits with sex-limited expression: a male trait, that we will call the ornament (e.g., tail size in peacocks), and a female trait, sexual preference for the male ornament (Fig. 12.8). This model has the interesting feature that evolution results in a line of equilibria, rather than a single unique outcome. The line of equilibria is a consequence of specifying two types of selection on the male ornament: viability selection represented by the tiger in Fig. 12.9, and mating preferences, represented by the peahen in the same figure.



Let z be the phenotypic value of a normally distributed male ornament. Before selection the mean of z is \bar{z} and its variance is P_m . The ornament is affected by many autosomal genes of small effect so that its breeding value is normally distributed with genetic variance G . Each generation Gaussian viability selection acts on the ornament in a stabilizing mode with an intermediate optimum θ_m . This viability selection changes the distribution of z so that after selection its mean and variance are

$$\bar{z}^* = (\bar{z}\omega + \theta P_m) / (\omega + P_m) \quad (12.06)$$

$$P_m^* = \omega P_m / (\omega + P_m), \quad (12.07)$$

where ω is the ‘variance’ of the viability function, so that the selection differential due to viability selection is

$$\bar{z}^* - \bar{z} = (\theta - \bar{z}) / (\omega + P_m)$$

Each generation, following viability selection, the population of surviving males is subjected to sexual selection by females that choose their mates on the basis of the male ornament. Female preference, y , is a normally distributed trait with phenotypic mean \bar{y} and variance τ^2 . We suppose that female preference is affected by many genes, so that breeding values are normally-distributed with variance H . Female choice

of males occurs in the following way. Each female is characterized by a Gaussian mate preference function with an optimum at y and a ‘variance’ of v^2 ,

$$\psi(z|y) \propto \exp\{-(z-y)^2/2v^2\}.$$

In other words, a female is most likely to mate with an encountered male if his ornament matches her preference, so that $z = y$, and her tendency to mate falls off as a Gaussian function as his ornament deviates from that optimum, y . Averaging these preference functions over the population of females, we obtain a Gaussian function that gives us the overall probability of mating as a function of male ornament value, z ,

$$\psi(z) = b \exp\{-(z - \bar{y})^2 / 2(\tau^2 + v^2)\}$$

where $b = \dots$. After sexual selection by a subset of female with preference y , the ornament mean \bar{z}^* will be shifted by an amount

$$(y - \bar{z}^*)P_m*/(v^2 + P_m^*) .$$

Averaging these shifts over the entire female population yields the selection differential due to sexual preference, $\bar{z}^{**} - \bar{z}^*$. The total selection differential is the sum of viability and sexual differentials, giving a total selection gradient for the ornament

$$P_m^{-1}s = P_m^{-1}[(\bar{z}^{**} - \bar{z}^*) + (\bar{z}^* - \bar{z})] = P_m^{-1}(\bar{z}^{**} - \bar{z})$$

$$P_m^{-1}s \approx \frac{\bar{y}/\alpha - (1+1/\alpha)\bar{z} + \theta}{\omega},$$

where $\alpha \equiv v^2/\omega$.

Turning to females, let us assume that a female’s fitness (progeny count) is unaffected by her mate choice, with the consequence that no selection acts on female preference, y . As a result, average preference, \bar{y} , will evolve only as a correlated response to selection on the ornament. We realize from earlier results (xxxx), that such a correlated response will depend on a genetic covariance between the two traits, B , a genetic covariance between the sexes. This genetic covariance is unlikely to be a consequence of pleiotropy, because in general it is difficult to imagine a contributing locus that affects a male ornament will also have a pleiotropic effect on female preference, or vice versa. On the other hand, B could reflect linkage disequilibrium. Mate choice in our model leads to assortative mating between the z and y phenotypes as well as selection that favors particular combinations of z and y . Lande (1981) reports a submodel showing that linkage disequilibrium does result from this combination of assortative mating and sexual selection and accumulates across loci to constitute a genetic covariance between the sexes, B . Furthermore, the primary determinate of the magnitude of B is ratio of mutation rates affecting female preferences versus the male ornament {a little more elaboration here, see Lande section}. Putting these results together with an expectation that G and H will equilibrate in mutation-selection balance, we can conclude that the G-matrix for our two characters,

$$\begin{bmatrix} G & B \\ B & H \end{bmatrix},$$

is likely to be stable if stabilizing selection on the preference is weak ($\omega \gg P_m$).

Using our standard expression for the response to two traits to selection (e.g., 12.00) and recalling that we have not allowed direct selection on y , we have

$$\Delta\bar{z} = \frac{1}{2}GP_m^{-1}s \quad (12.05a)$$

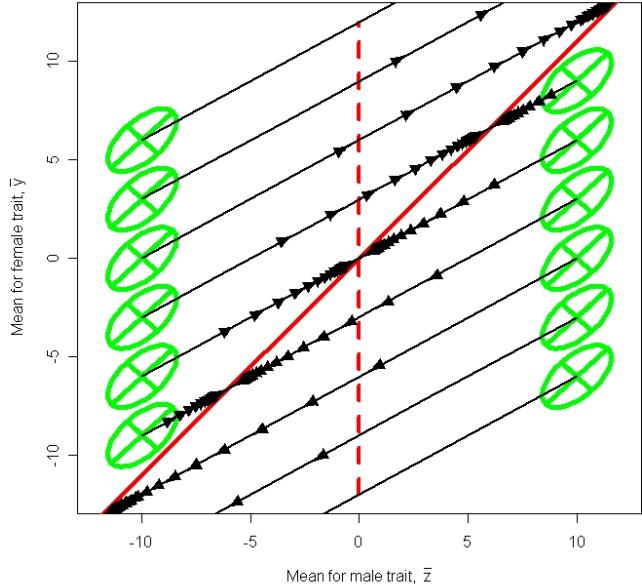
$$\Delta\bar{y} = \frac{1}{2}BP_m^{-1}s \quad (12.05b),$$

...
To find the equilibrium, we set the total selection gradient equal to zero and find that the equilibrium is a line

$$\bar{y} = (\alpha + 1)\bar{z} - \alpha\theta. \quad (12.13)$$

In other words, at any position along the equilibrium line, the force of viability selection pulling \bar{z} towards the optimum θ is exactly balanced by a force of sexual selection in the opposite direction (Fig. 12.5).

Figure 12.10. Lande's 1981 model for the joint evolution of a male ornament and female mating preferences based on that ornament. The stable case is illustrated. The solid red line is the line of equilibria. The dashed red line shows the position of the viability selection optimum, θ , for the male trait. Solid black lines show evolutionary trajectories for 12 starting positions, with arrows spaced every generation. The green ellipses at the starting points are G-matrices. $G=H=0.4$; $B=0.24$; $\omega=4$, $\alpha=0.1$, $r^2=P_m=1$. The similarity between g_{max} and the line of equilibria in this figure is a coincidence.



To analyze the stability properties of this equilibrium, it is useful to shift the coordinate system by a constant in the \bar{z} dimension {illustrate this shifted coord system on a fig} by defining

$$\tilde{z} = \bar{z} - \theta/(1+1/\alpha) \quad (12.14a)$$

$$\tilde{y} = \bar{y} \quad (12.14b)$$

In this new coordinate system, the selection gradient is

$$\beta = P_m^{-1}s = \frac{\tilde{y}/\alpha - \tilde{z}(1+1/\alpha)}{\omega} \quad (12.15)$$

and consequently the line of equilibria is

$$\tilde{y} = (\alpha + 1)\tilde{z} \quad (12.16)$$

Substituting (12.15) into (12.05), we obtain a simple expression for evolution in our new coordinate frame,

$$\begin{bmatrix} \Delta\tilde{z} \\ \Delta\tilde{y} \end{bmatrix} = \frac{1}{2\alpha\omega} \begin{bmatrix} -(\alpha+1)G & G \\ -(\alpha+1)B & B \end{bmatrix} \begin{bmatrix} \tilde{z} \\ \tilde{y} \end{bmatrix} = \frac{1}{2\alpha\omega} \begin{bmatrix} -(\alpha+1)G\tilde{z} + G\tilde{y} \\ -(\alpha+1)B\tilde{z} + B\tilde{y} \end{bmatrix}. \quad (12.17)$$

From (12.05) we see that $\Delta\bar{y}/\Delta\bar{z} = B/G$, indicating that evolution will occur along straight line trajectories with a slope given by the genetic regression B/G . From (12.17), we obtain the same result. Taking the eigenvalues of the matrix on the left side of (12.17), we find that one eigenvector corresponds to B/G , and its eigenvalue is

$$\lambda = [B - (\alpha+1)G]/(2\alpha\omega). \quad (12.18)$$

Consequently, we can conclude that after t elapsed generations, the vector of trait means evolving along a B/G trajectory will have changed by an amount $(1+\lambda)^t$. If $B/G < (\alpha+1)$, then λ is negative, and evolving populations approach the line of equilibria with ever decreasing speed; the equilibrium is stable (Fig. 12.5). But if $B/G > (\alpha+1)$, λ is positive, and populations evolve away from the line of equilibrium at ever increasing speed (as proposed by Fisher 1958); the equilibrium is unstable. The other eigenvector is the line of equilibrium itself with an associated eigenvalue equal to zero, indicating responses to selection cease on that line. The simplest change in ecology that could trigger the unstable, runaway case is a relaxation in stabilizing selection (an increase in ω) so that $B/G > (\nu^2/\omega) + 1$.

So far, we have considered only deterministic evolution of preference and ornament, ignoring the possibility of drift. The existence of a line of equilibria invites us to consider drift because, especially in the stable case, the bivariate mean might drift up and down that line. We need to consider situations in which effective population size is relatively large ($N_e \geq 500$ or 1000), so that the G-matrix is likely to be relatively constant. Under these conditions results from section (xx) tell us that random sampling of parents will increases the variance and covariance of replicate populations means by a constant amount each generation,

$$Cov(\bar{z}, \bar{y}) = \frac{1}{N_e} \begin{bmatrix} G & B \\ B & H \end{bmatrix}. \quad (12.19)$$

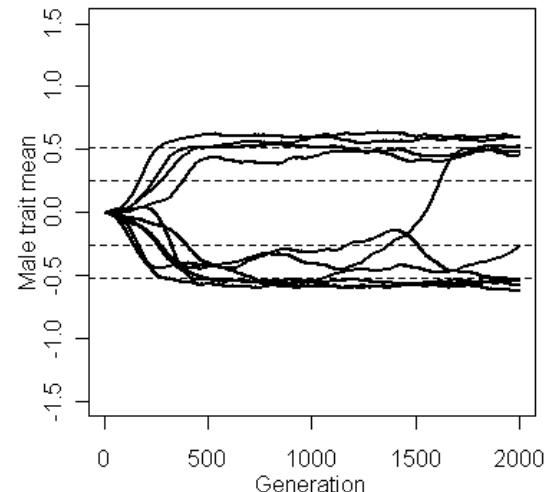
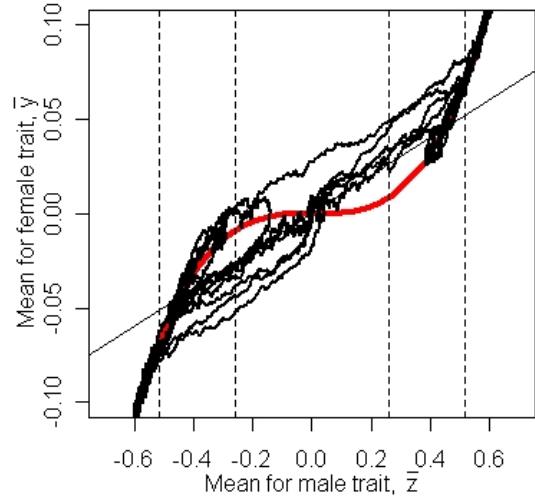
After t generations have elapsed, Lande (1981) shows that in the stable case the dispersion among replicate means will be

$$D(t) \approx K + \frac{H(1-r_g^2)t}{N_e(\alpha+1-B/G)^2} \begin{bmatrix} 1 & \alpha+1 \\ \alpha+1 & (\alpha+1)^2 \end{bmatrix} = \begin{bmatrix} Var(\bar{z})_t & Cov(\bar{z}, \bar{y})_t \\ Cov(\bar{z}, \bar{y})_t & Var(\bar{y})_t \end{bmatrix}, \quad (12.20)$$

where K is a constant matrix and $r_g = B/\sqrt{GH}$. Thus, genetic variance for preference, H , which played no role in deterministic responses to selection, plays a major role in the drift of both preference and ornament. This result makes intuitive sense. The preference distribution is not under direct selection and is free to drift along the line of equilibria. As it drifts, that preference distribution exerts directional selection on the male ornament distribution, so that the drift of \bar{z} is paced by the inheritance of y , i.e., by H .

Simulations by Uyeda et al. (2009) confirm Lande's (1981) interpretation that drift along the line of equilibria can produce rapid evolutionary diversification, even with N_e as large as 5,000-10,000. Those same simulations reveal that \bar{z} and \bar{y} are extremely close together as a lineage drifts along the equilibrium line, so that departures from the line are nonobvious, and that the two means often switch positions, so that sometimes $\bar{y} > \bar{z}$ and sometimes $\bar{z} > \bar{y}$.

Iwasa & Pomiankowski (1995) have described a process that yields perpetual evolution of a male ornament and female mating preference ... differs from Lande's (1981) model in a couple of important respects ... viability selection on the male trait is a fourth-power function ... $\exp\{c\bar{z}^4\}$, where c is a constant, so that the viability function is more flat-topped than a Gaussian or quadratic function. The optimum trait value with respect to viability selection is arbitrarily set to 0. ... female mating preference for a particular male is an exponential function of his trait value ... $\exp\{a(z - \bar{z})\bar{y}\}$, where a is a constant ... selection acts on female mating preferences, so that female fitness with respect to her preference value y is $\exp\{-by^2\}$. Ordinarily, the effect of selection on female preferences is to collapse a line of equilibria, as in Fig. 12.x, to a point, so that in the stable case, the joint evolution of ornament and preference has a single unique outcome (Pomiankowski et al. 1991a, b). In the present model, however, when the assumption of selection on preferences is combined with a flat-topped, viability selection function for the male trait, the combined effect is to produce a curve of equilibrium that is a cubic function. This cubic function has the remarkable property of producing a stable limit cycle in a population of infinite size (Iwasa & Pomiankowski 1995). This regular cyclical behavior of the model disappears in populations of finite size even if they are very large (Fig. 12.11), but nevertheless evolution appears to consist of perpetual swings between alternative stable zones on either side of the viability optimum, $\bar{z} = 0$.



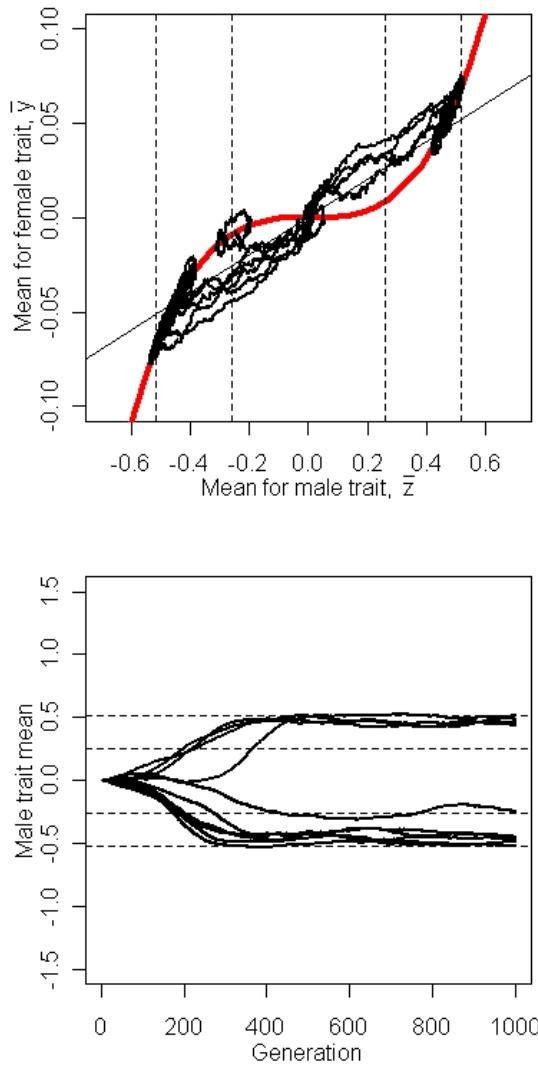


Figure 12.11. Perpetual evolution of a male ornament and female mating preferences. In this model the line of equilibria, the red curves in the upper panels, is a cubic function. This line of equilibria consists of stable and unstable zones. The unstable zone around the inflection of the red curve is denoted by the inner pair of dashed lines. The transition to flanking stable zones is located between the pairs of dashed lines. The evolutionary trajectories of two sets of 10 replicate lineages are shown with origination at $\bar{z} = 0, \bar{y} = 0$. Populations off the line of equilibria evolve deterministically along trajectories with a slope of B/G ; one such line (black) is shown through the origin. (a) On a 1,000 generation time scale, the tendency of lineages to stall within the zones of stability is apparent. (b) On a 2,000 generation time scale, diversification within the stable zones has continued, and one lineage has evolved across the unstable zone, from one stable zone to the other. Parameter values are: $N_e=10^6$, $G=H=0.5$, $B=0.05$, $a=0.4$, $b=0.001$, $c=0.05$, $u=0$, where u represents biased mutation.

In Fig. 12.11b, a single lineage makes the transit between alternative stable zones, but on a longer time scale, such transits are relatively common. Alternatively, with N_e smaller but still large (e.g., 10,000), transits are common on a time scale $\geq 1,000$ generations.

In several respects the Iwasa & Pomiankowski 1995 model should be considered a work in progress. For one thing, because of the model's parameterization, we do not know whether its assumptions about forms and intensity of selection on the two traits match empirical reality. Secondly, we do not know whether its signature feature, limit cycles, depends on a narrow choice of parameter values or persists over a broad range. Finally, the function that specifies the distribution of trait means as a function of time and N_e is unknown. Nevertheless, the model emphasizes the essential point, discussed below, that a variety of evolutionary outcomes are possible when two traits have interactive effects on fitness.

The 30+ models that have been based on the framework established by Lande (1981) have explored a variety of alternative assumptions and established some important generalizations (Mead & Arnold 2004) ... a variety of outcomes are possible, depending mainly on assumptions about selection (Fig. 12.6) ... alternative assumptions have been used to capture the essential features of 'good genes', sexual conflict, costly female choice, ...despite the variety of possible outcomes, all models reveal both stable and unstable outcomes ... most models have not explored drift ... refer to explorations described by Uyeda et al 2009 ... end with a statement about implications for speciation.

{In a final paragraph for this section, connect up these modeling efforts with the empirical literature on sexual selection ... estimates of the heritability of male traits and female preference (find the most recent review) ... estimates of the genetic correlation between preference and trait (find the most recent review) ... tests for various kinds of selection ... finally, comment on the unsatisfactory state of tests for alternative models}

12.9 The evolution of sexual isolation

Lines of equilibria and stable limit cycles imply that diversification by sexual selection could lead to speciation, perhaps easily and rapidly. To verify those implications and predict rates of speciation, we need a model that connects sexual selection to sexual isolation, and speciation. One way to proceed is to make those connections using Lande's (1981) model, outlined in the last section (Arnold et al. 1996). In particular, we can use (12.0x) to specify the probability that a randomly drawn female from one population, A , will mate with a randomly drawn male from another population, B ,

$$\pi_{AB} = c \exp\{-(\bar{z}_B - \bar{y}_A)^2 / 2\sigma^2\}, \quad (12.21)$$

where c is a constant less than 1 but greater than 0, and the variation parameter $\sigma^2 = \nu^2 + \tau^2 + P_m$ is assumed to be the same in both populations. This same expression can be used to specify the probabilities of mating, π_{AA} and π_{BB} , when mating partners are drawn from the same population or from different populations in opposite order, female from B and male from A , π_{BA} . These same four probabilities can be estimated in an experimental assay of sexual isolation. In such an assay, mating partners are drawn at random from two different populations, four kinds of sexual encounter are staged and replicated, mating success is scored for each kind, and the four probabilities of mating are estimated directly from those data. One common and useful measure of sexual isolation that summarizes such data is called the *index of joint isolation*,

$$JI = \pi_{AA} + \pi_{BB} - \pi_{AB} - \pi_{BA}.$$

This index ranges from +2 when ... to -2 when ...

To model the evolution of Jl , we need to consider a representative pair of populations, A and B , that have evolved to the line of stable equilibrium in Lande's (1981) model and are now drifting along it. To proceed, it will help to make some simplifying assumptions. First, we note (and confirm by simulation) that the difference between preference and ornament means within a population will be negligible relative to the expected between-population difference in means. Under this assumption, from (12.xx), $\pi_{AA} = \pi_{BB} = 1$ and $\pi_{AB} = \pi_{BA}$. The distribution of $\pi_{AB} = \pi_{BA}$ among replicate pairs of populations is then a function of the corresponding distribution of pairs of ornament means (Fig. 12.7), which we can determine using (12.xx). Pursuing this course, we find that the function giving the expected value of Jl as a function of time is not simple, but it can be evaluated numerically or the process can be simulated (Uyeda et al. 2009). Some sample paths are shown in

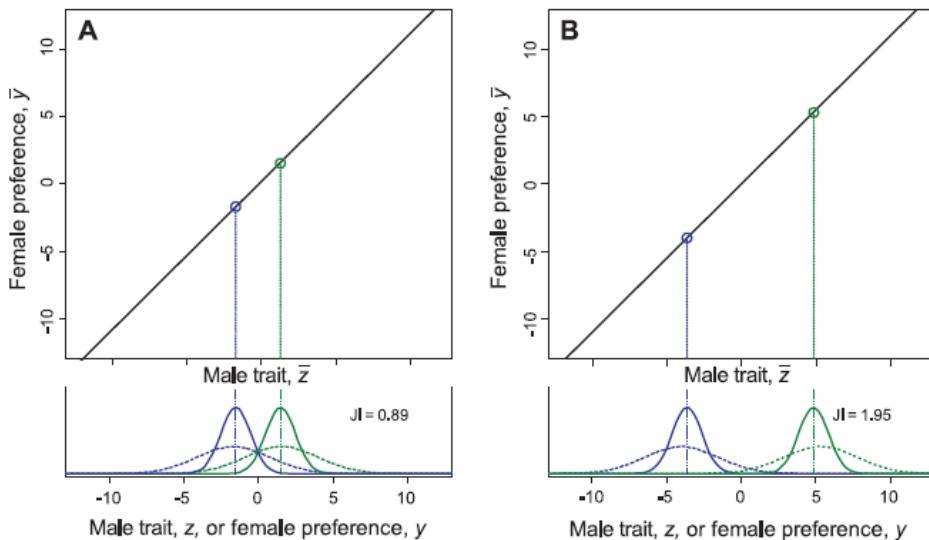
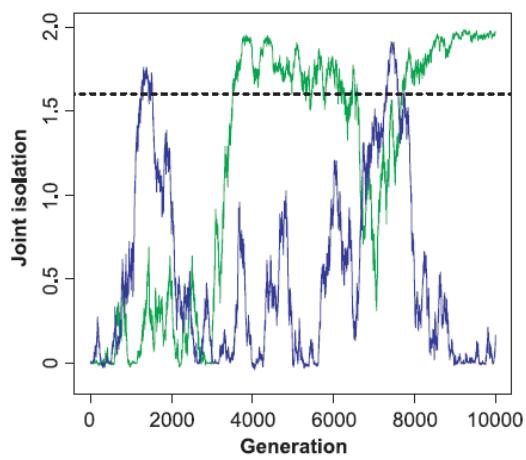
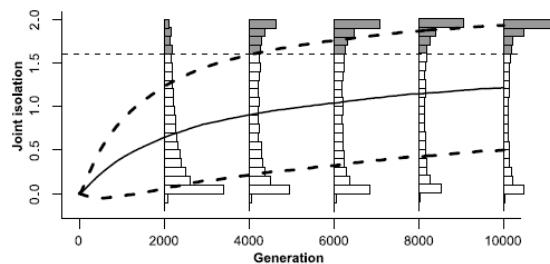
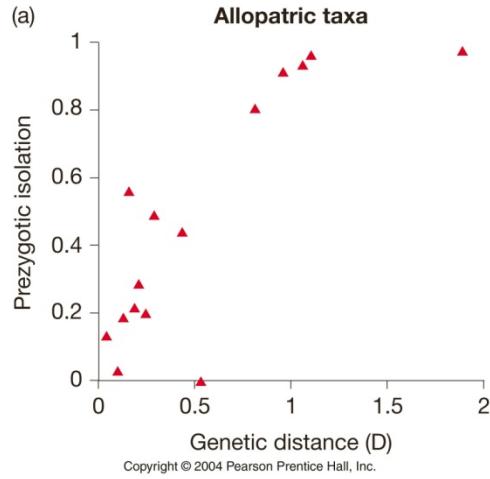


Fig 12.7. Simulations reveal that the evolutionary trajectory of Jl for any pair of populations is often chaotic, as we might expect from the overwhelming influence of drift. Instead of progressing steadily towards increasing isolation, individual trajectories often alternate between episodes in which isolation waxes and wanes (Fig. 12.8). On the other hand, when we follow the histories of many replicate pairs of populations, we do see a trend towards increasing isolation (Fig. 12.9). Nevertheless, the distribution of Jl is characteristically bimodal during intermediate stages, until finally the mode becomes strong sexual isolation ($Jl \approx 2$), typically after some thousands of generations.



{go on to discuss influence of N_e , etc. ...; relate this model to the evolution of Jl in drosophila and desmogs, perhaps producing histograms of Jl that can be compared to the fig on the left; bring in a figure from Coyne & Orr 1997 showing Jl as a function of

genetic distance and comment especially on the great spread in J_1 as short genetic distances, changing asymptotically to high values at long Nei distance}



12.10 The dimensionality of mate choice and sexual isolation

... from a variety of perspectives it seems likely that mate choice is based on multiple traits ... the possibility that sexual selection involves multiple traits sits in stark contrast to the modeling preoccupation with single traits ... Pomiankowski & Iwasa ... Iwasa & Pomiankowski (1993, 1994) are the exception ... they found that ... Hohenlohe & Arnold (2010) took another approach by ... evidence that sexual isolation based on two or more trait dimensions in drosophila, cichlid fishes and plethodontid salamanders ...

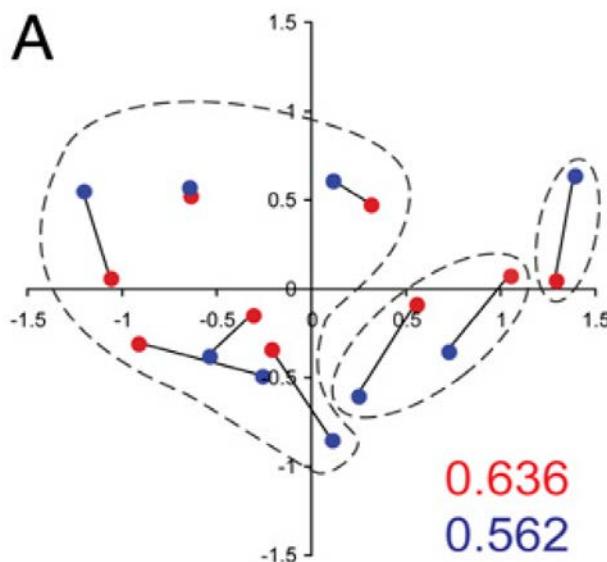
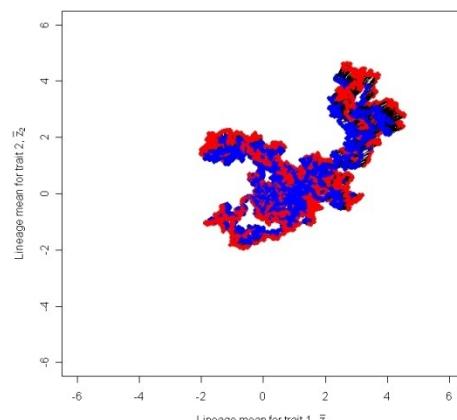


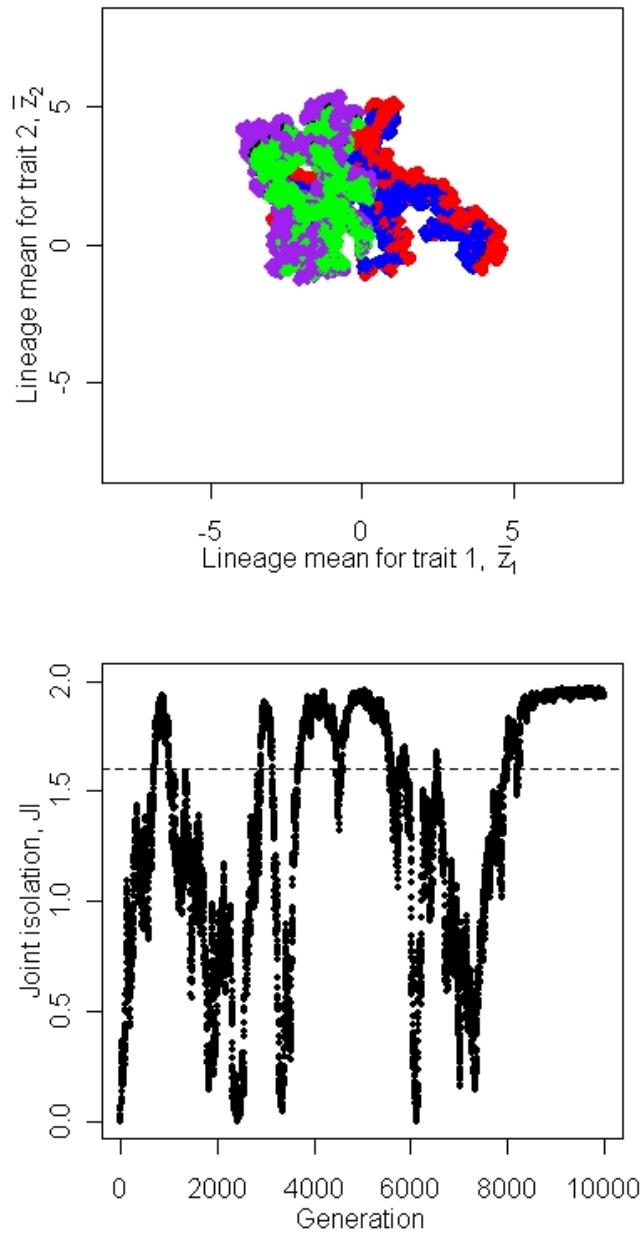
Fig. 12.11. Latent axes for male and female traits in *Desmognathus* salamanders, inferred from a survey of sexual isolation. Male mean (blue) and female means (red) from the same population are connected by solid lines. Hohenlohe & Arnold 2010.

Figure 12.12. Joint evolution of two male traits and female mating preferences based on them on a stable plane of equilibria. Male means are shown in blue and female means are shown in red. Mean in the same generation are connected by solid black lines, but are often not visible because the



means of the two sexes are close together. 10,000 generation of interactions between drift and selection. $N_e=500$. The genetic variance of male and female traits is 0.4. Traits are genetically uncorrelated within sexes. Genetic correlations between the sexes are 0.6 for traits with the same subscript and 0 otherwise. For both male traits, $\omega=4$ (with no correlational selection and $P_m=1$; for both both female traits, $v^2=0.4$ (with no correlational preferences), $\tau^2=1$, and $\alpha=0.1$. The optimum or the male traits is situated at 0, 0.

Figure 12.13 Sexual isolation evolving as a function of the joint evolution of two male traits and female mating preferences based on them in two, independent lineages. (a) Joint evolution of bivariate male and female traits in two lineages. Same conventions and parameter values as in Fig. 12.12 except that one pair of means is colored blue and red, and the other pair is colored purple and green. The second lineage partially obscures the history of the first lineage. (b) Time course for joint isolation, J_I . The horizontal dashed line shows the lower level of J_I that is characteristic of sympatric, isolated species.



Chapter 13: Trait Evolution on Dynamic Adaptive Landscapes: Stochastic Models

© Stevan J. Arnold 2016

Overview.- Although a variety of models have been proposed to explain adaptive radiations in a single trait, a common feature of relatively successful models is that the peak of a hill-shaped adaptive landscape moves, but the shape of the hill does not change. Although many of these models fail to explain temporal patterns in large data set assembled by Philip Gingerich and others, models in which the optimum fluctuates with Brownian motion can approximate data from adaptive radiations. Models that come closest to accounting for the data are ones in which the optimum moves rarely, but is capable of substantial excursions (0-6 phenotypic standard deviations). These results direct our attention to the problem of formulating and testing alternative models for the peak controller process (PCP).

{introductory paragraph here, sketching the rationale for the approach (see section in Estes & Arnold 2007), anticipating the main results that follow, and using Fig. 13.0 = cartoon of the models of peak movement as a guide ...}

13.0 Brownian motion of the adaptive peak

In Chapter 11 we showed that the drift of populations of finite size on an adaptive landscape with a single stationary adaptive peak does not explain the diversification commonly seen in evolving clades. The basic problem is that a stationary peak does not permit enough variation among lineage means. We can imagine solving this problem by allowing the peak to move in some pattern. In this section we explore Brownian motion of the peak and evaluate the diversification produced by that motion.

By Brownian motion of the optimum we mean that the current position of the optimum deviates from its position in the preceding time interval and by an amount that is a normally distributed random variable: $\theta_{t+1} = \theta_t + \varepsilon_\theta$, where the last term is random variable drawn from a normal distribution with a mean of zero and variance of σ_θ^2 . We want to determine how a set of replicate populations will evolve in response to this mode of peak movement. If we focus on a single evolving lineage, the change in the trait mean in a particular generation ($t+1$) has two components

$$\Delta\bar{z}(t+1) = \left(\frac{\theta_t - \bar{z}_t}{\omega + P} \right) G + N(0, G/N_e),$$

a deterministic response (the first term on the right) that is proportional to the distance to the optimum and a stochastic, drift response that is exacerbated by small population size. The deviation of the optimum from its position in the previous generation is a draw from a normal distribution with a mean of zero and a variance of σ_θ^2 ,

$$\Delta\theta(t+1) = \theta_t + N(0, \sigma_\theta^2).$$

The distribution of the means of replicate lineages is normally distributed about the long term average position of the optimum with an ever-increasing variance (Fig. 13.1a),

$$Var[\bar{z}(t)] = \frac{\sigma_\theta^2 + \frac{G}{N_e}}{2a} \{1 - exp[-2at]\} + \sigma_\theta^2 t \left\{ 1 - 2 \frac{(1 - exp[-at])}{at} \right\},$$

where $a = \frac{G}{\omega+P}$ (Hansen et al. 2008). In contrast, Hansen et al. (2008) find that the variance among replicates in the position of the optimum at generation t is

$$\text{Var}[\theta(t)] = \sigma_\theta^2 t.$$

We see that the expression for the variance for the variance among replicate means rapidly converges on $\sigma_\theta^2 t$, which is the variance in the position of the optimum. An example of lineage responses to Brownian motion of adaptive peaks is provided in Fig. 13.3. Using reasonable values for selection and inheritance, we see – as expected – that the dispersion of lineage means is largely a reflection of variation in the position of optima.

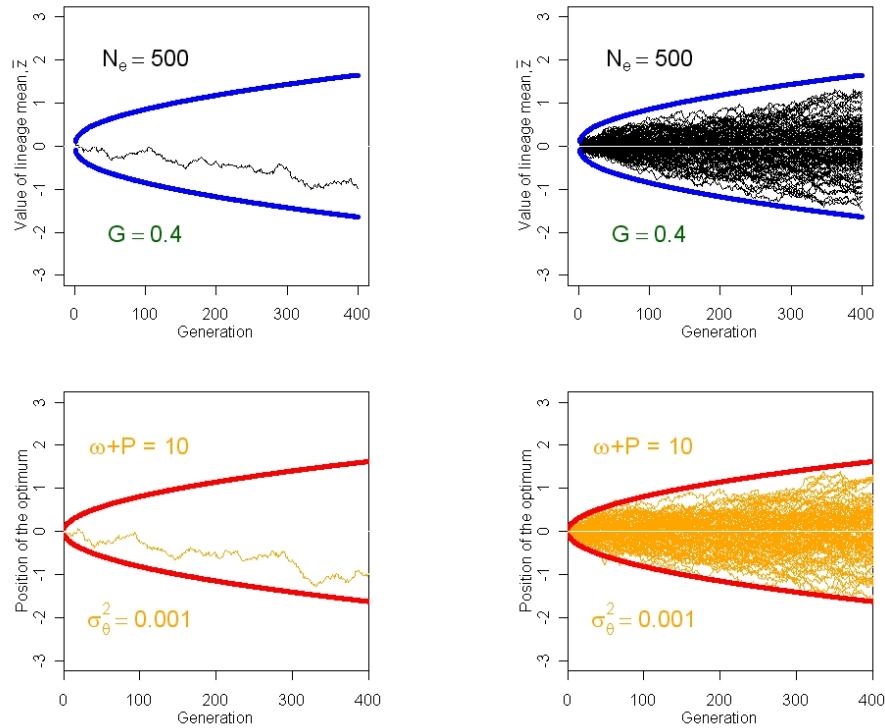


Figure 13.3 Response of the lineage mean to an intermediate optimum that undergoes a random walk. Lower panels show the path of the optimum, θ , through time. The 99% confidence limits of the optimum are shown with red lines. The upper panels show the paths of lineage means, \bar{z} . 99% confidence limits are shown with blue lines. (a) A single lineage mean closely tracks the moving optimum of its AL. (b) 100 replicate lineages track the moving optima of their ALs.

An example of lineage responses to Brownian motion of adaptive peaks is provided in Fig. 13.3. Using reasonable values for selection and inheritance, we see that the dispersion of lineage means is largely a reflection of variation in the position of optima.

13.1 White noise motion of the adaptive peak

We now turn to a more dramatic mode of peak movement, white noise motion. In this mode our model of the position of the optimum is $\theta_{t+1} = \varepsilon_\theta$, where, as before, the last term is random variable drawn from a normal distribution with a mean of zero and variance of σ_θ^2 . Notice that the optimum carries none of its

past history, as it does in the case of Brownian motion. Instead, the optimum undergoes more erratic motion. As in the last model, each generation of the trait mean of a lineage is affected by two factors, recent movement of the optimum and genetic drift of the trait mean. The trait means of replicate lineages are normally distributed with a expected value that corresponds to the expected value for the optimum, which we have assumed is zero. The variance among the trait means of replicate lineages at any particular generation t is

$$Var(\bar{z}_t) = \left[\frac{G\sigma_\theta^2}{2(\omega+P)} + \frac{\omega+P}{2N_e} \right] \left\{ 1 - \exp \left[-2 \left(\frac{G}{\omega+P} \right) t \right] \right\}$$

... the new term is the first one on the right, which represents the contribution from the trait mean as it tracks peak movement ... facilitated by genetic variance ... inhibited if stabilizing selection is weak ...

$$Var(\bar{z}_\infty) = \frac{G\sigma_\theta^2}{2(\omega+P)} + \frac{\omega+P}{2N_e}$$

... notice that the variance takes a constant value at equilibrium, in contrast to the Brownian motion case ... (Fig. 13.1b). A particular example of white noise motion and the evolution it evokes is shown in Fig. 13.5, which employs intermediate values for genetic variance and population size and imposes moderately weak stabilizing. Notice in Fig. 13.5a that the evolving mean fails to track the rapid, erratic motion of the optimum. As a consequence, the ensemble properties of many replicate populations are a much damped version of their moving optima (Fig. 13.5b). Another revealing contrast with Brownian motion is the general result that variance among replicate means rapidly achieves its asymptotic value instead of expanding endlessly. The consequence of this stationary variance is bounded evolution that reflects the principal features observed in large data sets (Fig.xxxxx) ... the problem of too little diversification ... how does the imposed variance in theta compare with observations in nature?

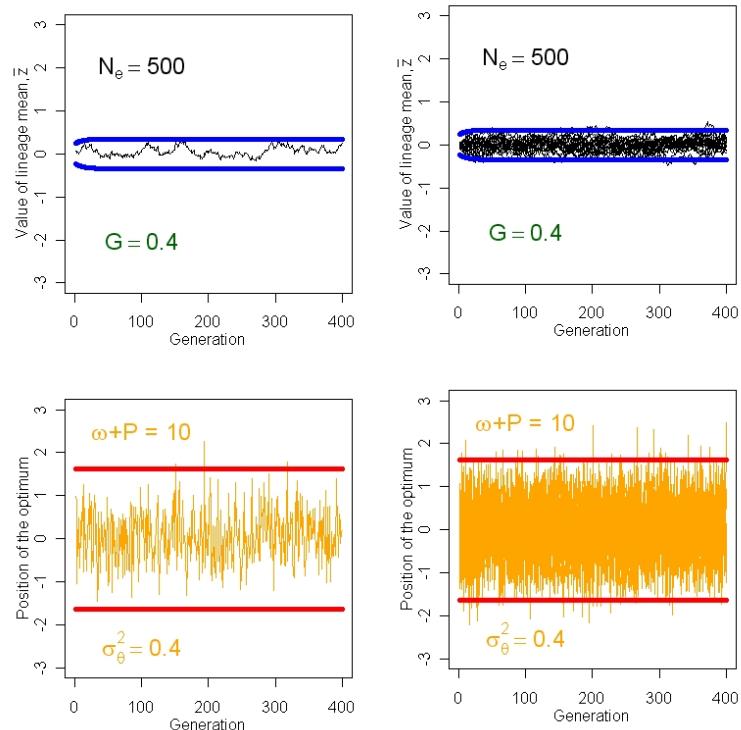


Figure 13.5. Simulations of lineages evolving in response to white noise motion of the optimum of the AL. The long term average position of the optimum is zero. The blue lines in the upper panels and red lines in the lower panels show the upper and lower 99% confidence limits (a) The trait mean of a single lineage (above) evolving in response the movement of the optimum of its AL. (b) The trait means of 20 replicate lineages (a) evolving in response to the movement of their ALs

13.2 Steady movement of the adaptive peak

In this model, in contrast to the models discussed so far, the trait mean shows a long term trend ... Our model of the position of the optimum is $\theta_t = kt + \varepsilon_\theta$, where k is a constant denoting the rate of deterministic change in the optimum. (Fig. 13.2a = showing both θ and \bar{z} as a function of time) ... Notice that white noise motion of a stationary peak is a special case of this model ...

$$E[\bar{z}_t] = kt - k\left(\frac{\omega + P}{G}\right)\left\{1 - \exp\left[-\left(\frac{G}{\omega + P}\right)t\right]\right\}$$

... first term on the right represents response to directional selection elicited by peak movement ... the second term represents the negative contribution of stabilizing selection balance ... last term ... approach to equilibrium ... (Fig. 13.2b = based on Fig. 4 of Estes & Arnold 2007)

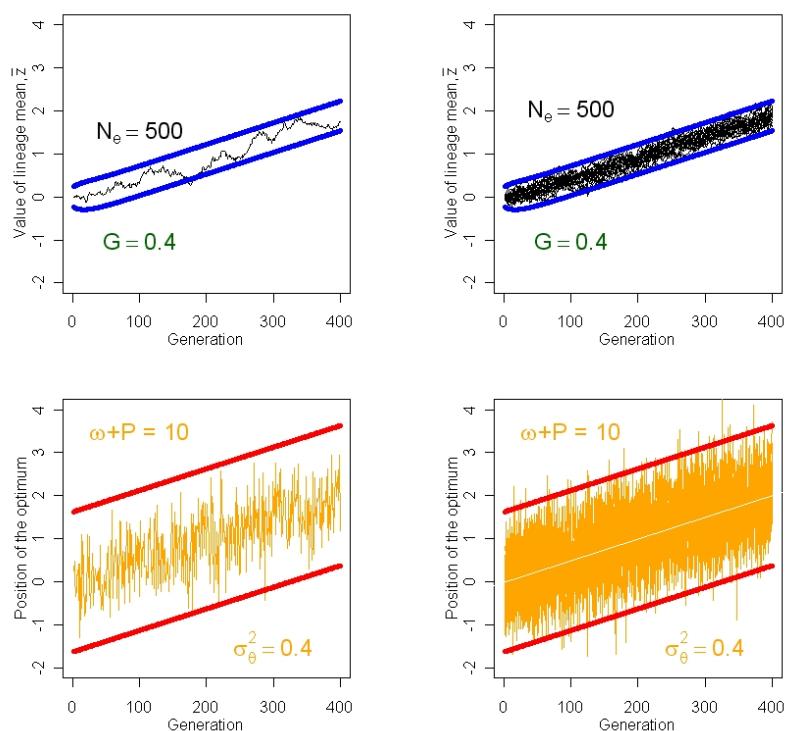
$$E[\bar{z}_\infty] = kt - k\left(\frac{\omega + P}{G}\right)$$

$$Var(\bar{z}_t) = \left[\frac{G\sigma_\theta^2}{2(\omega + P)} + \frac{\omega + P}{2N_e} \right] \left\{ 1 - \exp\left[-2\left(\frac{G}{\omega + P}\right)t\right] \right\}$$

$$Var(\bar{z}_\infty) = \frac{G\sigma_\theta^2}{2(\omega + P)} + \frac{\omega + P}{2N_e}$$

An example of evolving populations evolving in response to steady movement of the adaptive peak is provided in Fig. 13.6. Notice that, as before, substantial white noise movement of the peak evokes only moderate excursions in the evolving trait mean.

Figure 13.6 Simulations of lineages evolving in response to a steadily moving intermediate optimum with superimposed white noise fluctuations in position. Conventions as in Fig. 13.5. The optimum increases in value at a rate of 0.005 within-population standard deviation per generation ($k =$



0.005). (a) The trait mean of a single lineage evolving in response to the moving optimum. (b) The trait means of 20 replicate lineages evolving in response to the same selection regime.

The Achilles heel of the steadily moving optimum model is that in the fullness of time the trait mean is bound to evolve outside the observed boundaries for divergence. For example, even if the optimum moves at a rate that would be undetectable in microevolutionary analyses ($k=0.0002$), the average lineage will have diverged about 6 standard deviations after 30,000 generations (Fig. 13.7). Although this excursion past the boundary may not seem egregious, steadily moving optima lead to impossible levels of divergence on a time scales that exceed 100,000 generations, under all combinations of realistic parameter values (Estes & Arnold 2007).

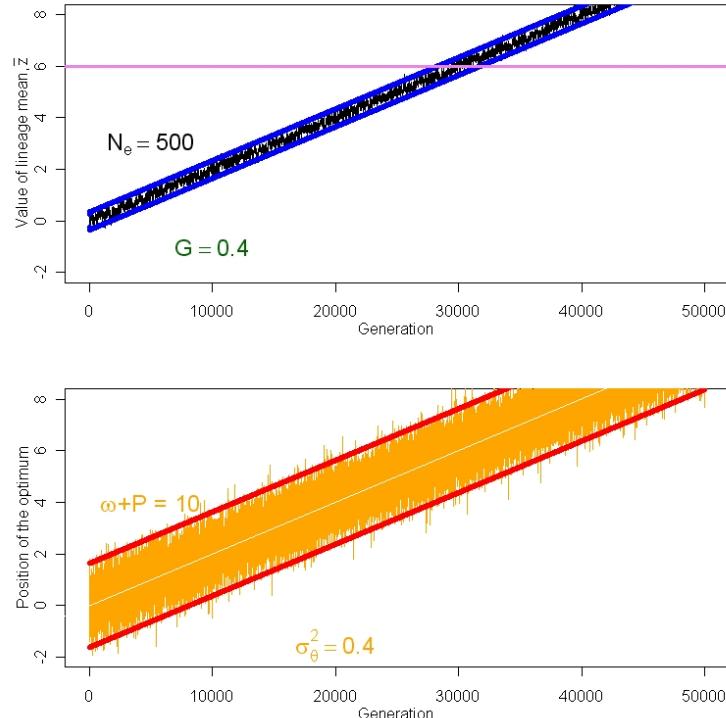
Figure 13.7 Simulations of 20 replicate populations evolving in response to a very slowly moving intermediate optimum. The rate of motion of the optimum is $k=0.0002$, 25 times slower than in Fig. 13.6.

Displacement of the optimum will cause a lag, a temporary exaggeration of distance between mean and the adaptive peak. If the optimum moves at a steady rate, this lag equilibrates at a value that is larger when stabilizing selection is weak and genetic variance is small,

$$L \equiv \theta_t - \bar{z}_t = (\omega + P)G^{-1}\Delta\theta,$$

where $\Delta\theta$ is the per generation change in the position of the optimum

(Lynch & Lande 1993, Jones et al. 2004). Lag can have deleterious demographic consequences for the population. The further the mean is displaced from the adaptive peak, the greater the decline in the population's mean fitness. The resulting impairment in population growth rate can lead to population extinction. Gomulkiewicz & Houle (2009) show how the resulting risk to the population can be expressed in critical values of genetic parameters.



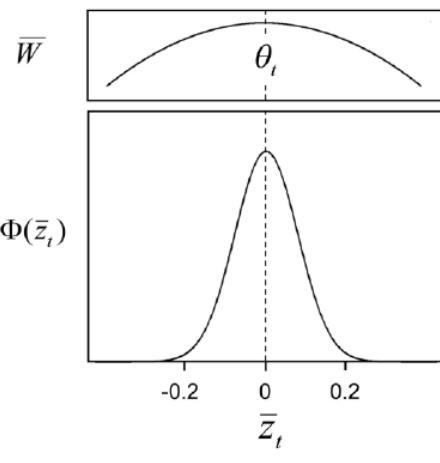


Figure 4: Diversification among replicate populations at generation t under the moving optimum model. The top panel shows the expected position of the optimum, θ_t . The bottom panel shows the distribution of phenotypic means, $\Phi(\bar{z}_t)$. The expected phenotypic mean lags so very slightly behind the optimum at generation t that it appears to be superimposed on the optimum. Heritability is 0.4, $\omega^2 = 10$, $N_e = 1,000$, $k = 0.001$, and $\sigma_\theta^2 = 0.001$. Other conventions as in figure 2.

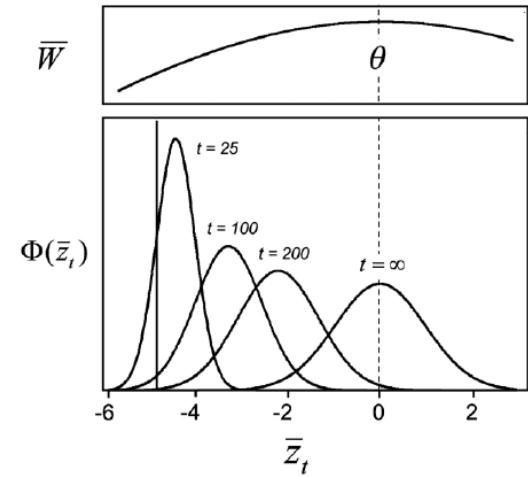


Figure 3: Divergence in response to a displaced optimum. The top panel shows an adaptive landscape with an optimum that has been displaced 5 phenotypic standard deviations from the phenotypic mean, shown as a vertical line. The bottom panel shows the distribution of phenotypic means, $\Phi(\bar{z}_t)$, after different intervals of time have elapsed, as populations evolve in response to the displaced optimum. Heritability is 0.4, $\omega^2 = 99$, and $N_e = 50$. Other conventions as in figure 2.

13.3 A single displacement of the adaptive peak

... drift-stabilizing selection balance about a stationary peak (section 11.z) is a special case of this model

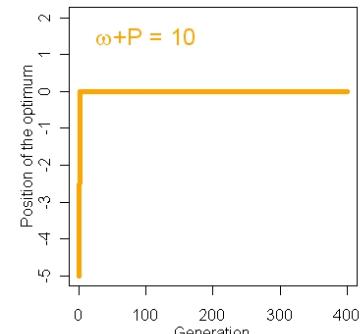
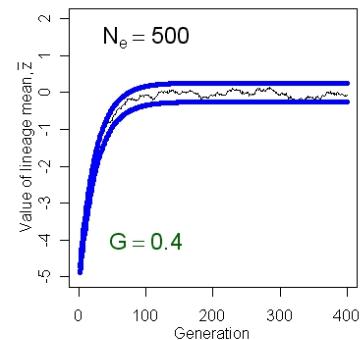
$$E(\bar{z}_t) = \bar{z}_0 \exp\left[-\left(\frac{G}{\omega + P}\right)t\right]$$

$$\text{Var}(\bar{z}_t) = \frac{\omega + P}{2N_e} \left\{ 1 - \exp\left[-2\left(\frac{G}{\omega + P}\right)t\right] \right\}$$

$$\text{Var}(\bar{z}_\infty) = \frac{\omega + P}{2N_e}$$

... (Lande 1976) (Fig. 13.3).

Figure 13.8. The trait mean of a single population evolving in response to a single displacement of its adaptive peak. The displacement of the adaptive peak by $5\sqrt{P}$ occurs instantaneously in generation 1 (lower panel). The evolving trait mean is shown in black in the upper panel. 99% confidence limits under these parameter values are shown in blue.



{Fig. 13.4 = based on Figure 1 from Hunt et al. 2007} ... loss of anti-predator structures in a fossil lineage of stickleback ... good fit to the displaced optimum model ... suggests that an essentially instantaneous escape from predators was followed by a loss of structures over a period of 2,000 generations *run some simulations; could this really be an instantaneous change in the optimum?**

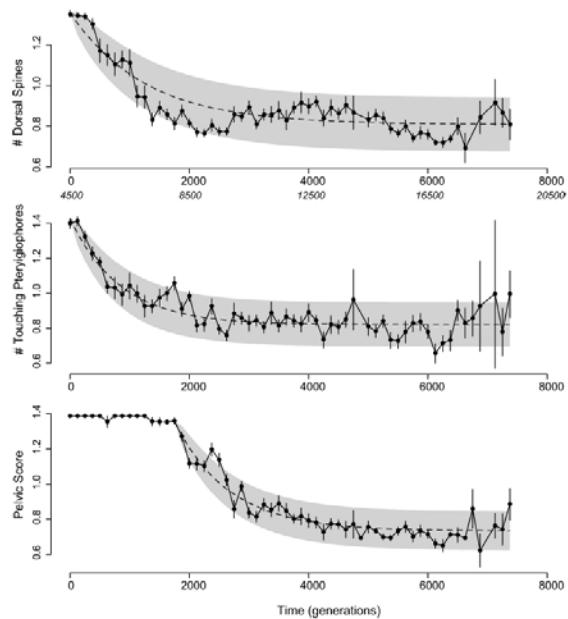
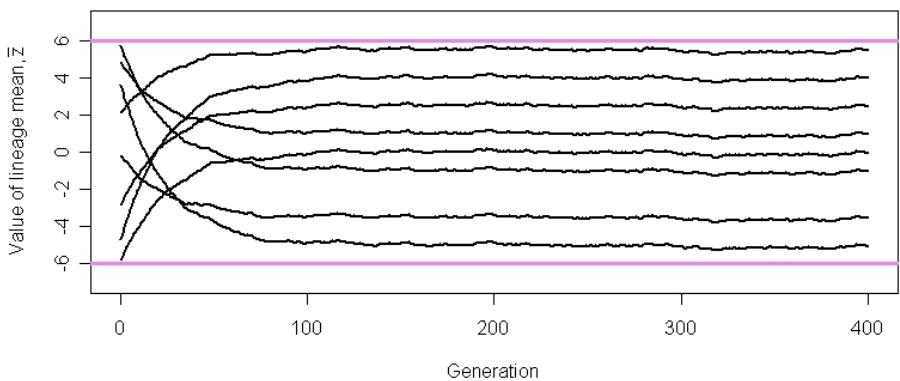


Figure 1. Plots of mean trait values over time, starting with the first appearance of the highly armored stickleback lineage. Time is measured in organismal generations (= 2 years/generation), with generation zero set at the first sample marking the invasion of the lake by this particular lineage (smaller italicized numbers below the time axis in the top panel show the time scale from Bell et al. [2006], which is in years). Vertical error bars denote one standard error; samples with fewer than five individuals have been omitted. For each trait, the expected evolutionary trajectory of the best-fit adaptive model is shown as a dotted line, with the 95% probability envelope around this solution in gray. There is a delay before the pelvic score character begins to evolve to a lower armored form (see text).

{add sections here as necessary to cover additional models in the literature ... e.g., the single- and multiple-burst models of Uyeda et al. 2011}

The potential of the displaced optimum model to account for the bounded evolution that is observed in large, long term data sets is illustrated in Fig. 13.10.

Figure 13.10. Simulations of 8 lineages evolving in response to different displacements of their intermediate optima. At generation zero, the position of the optima ranges from -5 to 6 and at generation 1 those optima are displaced by $3.5 \pm 9\sqrt{P}$.



Here we relax the assumption that the optimum is zero before displacement and in addition allow displacements of different magnitudes in different lineages. These adjustments to the model allow evolution outcomes that produce a band of evolving means that reside of indefinitely within specified boundaries, matching a pattern seen in long term data sets (Estes & Arnold 2007). Nevertheless, even with these adjustments, the model is unsatisfactory in a few respects. Allowing optima displacements only in generation one is obviously artificial. Secondly, we would like both the timing and magnitude of displacements to be generated by a stochastic process, rather than impose them by fiat. The next model grants both of those wishes.

13.4 Single burst model

This model corrects the deficiencies of the displaced optimum model with the aim of producing the kind of evolutionary pattern observed in Fig. xx.x (Uyeda et al. 2011). For simplicity in this and the next model, we will equate the position of the trait mean with the position of its intermediate optimum. In other words, we will introduce additional realism into the movement of the optimum, but we will assume that the trait mean track those movements with no lag. We will consider the consequences of this modeling short cut at the end of this section.

A key innovation of this model is that, instead of restricting occurrence of displacements at the beginning of time intervals, we will incorporate stochastic variation in the timing of displacements. The timing of the displacements is generated at random by a Poisson process, but as in the displaced optimum model, only a single displacement is allowed (Fig. 13.14).

Figure 13.14 Simulation of the single burst model. The orange line shows the position of the optimum. Parameter values are $\sigma_p^2 = 5$, $\sigma_d^2 = 37$, and $\lambda = 0.005$. The average waiting time for a displacement of the optimum is $1/\lambda = 200$ generations. The 99% confidence limits for a normal distribution with mean zero and variance $\sigma_p^2 = 5$ are shown in violet. In this run, a single displacement of the optimum occurs at about generation 770.

A Poisson process is one in which rare occurrences of a certain event, e.g., displacement, are both rare and independent of time. Thus, for a single lineage the probability that no displacement has occurred in elapsed time t is $\exp(-\lambda t)$, where λ , the average rate of occurrence, is a constant. The probability that a first displacement occurs by time t is $1 - \exp(-\lambda t)$. Once a displacement happens, its magnitude, d , is determined by a draw from a normal distribution with a mean of zero and a variance of σ_d^2 . Even in the absence of a displacement, the position of the optimum varies from generation to generation, so that the steady state variance among trait means is normally distributed with a mean of zero and a variance of σ_p^2 . The resulting distribution of trait means for a set of replicate lineages is the sum of two normal distributions, each weighted the probability that a displacement does or does not occur, and is given by

$$\Phi(\bar{z}_t) = \frac{[1 - \exp(-\lambda t)]}{\sqrt{2\pi(\sigma_p^2 + \sigma_d^2)}} \exp\left(\frac{-\bar{z}^2}{2(\sigma_p^2 + \sigma_d^2)}\right) + \frac{[\exp(-\lambda t)]}{\sqrt{2\pi\sigma_p^2}} \exp\left(\frac{-\bar{z}^2}{2\sigma_p^2}\right)$$

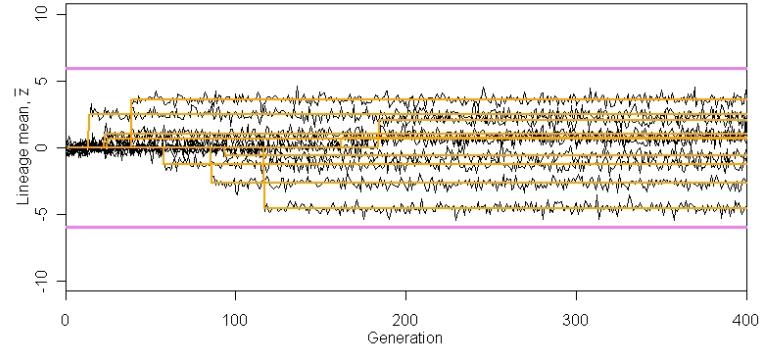
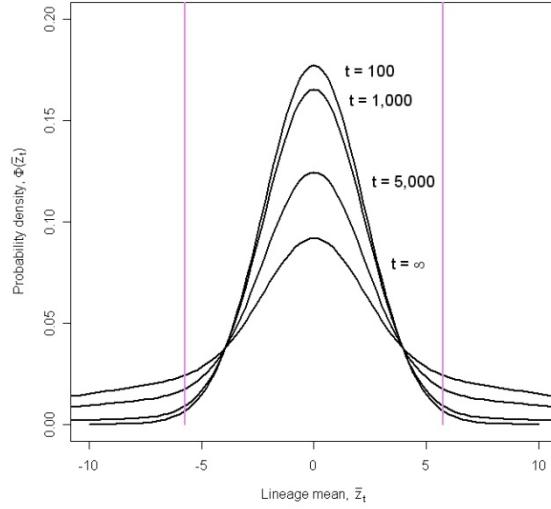
(Uyeda et al. 2011). As $t \rightarrow \infty$, this distribution converges on a normal distribution with mean zero and a variance of $\sigma_p^2 + \sigma_d^2$.

Figure 13.11 Distributions of lineage means after various elapsed times according to the single burst model. Parameter values are $\sigma_p^2 = 5$, $\sigma_d^2 = 87.5$, and $\lambda = 0.0001$. The average waiting time for a displacement of the optimum is $1/\lambda = 10,000$ generations. The 99% confidence limits for a normal distribution with mean zero and variance $\sigma_p^2 = 5$ are shown in violet.

When the rate of occurrence, λ , is very small, lineages evolve within a bounded region for thousands of generations. The width of that bounded region is determined by the confidence limits of a normal distribution with a mean at zero and a variance of σ_p^2 (Fig. 13.11). On a longer time frame, the bounded region expands so that its width corresponds to the confidence limits of a normal distribution with a mean at zero and a variance of $\sigma_p^2 + \sigma_d^2$.

Figure 13.15. Simulation of 10 replicate lineages evolving according to the single burst model. The orange lines show the position of the optima, which undergo a single displacement in each lineage.

Parameter values are $\sigma_p^2 = 0.1$, $\sigma_d^2 = 10$, and $\lambda = 0.01$. The average waiting time for a displacement of the optimum is $1/\lambda = 100$ generations. Typical bounds for actual data are shown in violet.



In the preceding section we used the single burst model to explain departures from stasis that persists on a long time scale (for hundreds of thousands to millions of generations) but the model can also be used to represent evolution on a much shorter time scales. For example, just as we used the displaced optimum model to explain diversification within but not outside the normal bounds of data (Fig. 13.10), the single burst model can also be employed to the same end. In Fig. 13.15 we have chosen a small value for σ_p^2 to represent small amplitude fluctuations in the position of the optimum, and a modest value for

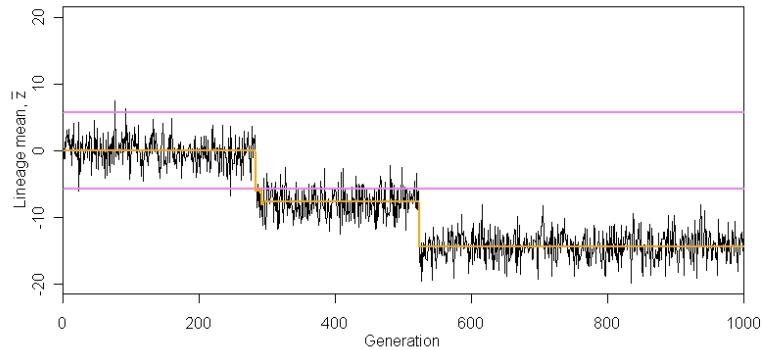
σ_d^2 so that single displacements of the optimum are unlikely to carry the trait mean more than about 5 standard deviations away from the clade's long term average.

Although this model makes important and useful innovations in how the optimum moves, it takes some short cuts in modeling the evolution of the trait mean. In the first place, we have simply equated the movement of the trait mean with the movement of the optimum. This economy in modeling is obviously a simplification, because we know from earlier models (e.g., section 13.3) that the trait mean is likely to take dozens of generations to track a sudden, large displacement of the optimum. Secondly, we have modeled relatively large displacements of the optimum, but ignored the possibility of fluctuations in position on short time scales. Instead, we employed a white noise process to generate short term stochastic variation in the trait mean via the variance parameter σ_p^2 . Again, we know that this is an unrealistic simplifying assumption, because we discovered in sections 13.1 and 13.2 that even white noise movement of the optimum does not evoke white noise movement of the trait mean. In a white noise process, the deviation that is drawn from a normal distribution is from the long term average, not just from the trait mean in the last generation (as in Brownian motion). The consequence is that under a white noise process, the mean would be rapidly whipped from one position to the next every generation. The normal rules of quantitative inheritance prevent this from happening in the real world. A real mean meanders through time (Fig. 13.5a), it does not whip back and forth. In other words, σ_p^2 should be taken as a surrogate of variation produced by some process that needs to be more realistically modeled. White noise is an assumption of convenience that produces a desired result. The desired result is that variance produced by the process does not increase with time, enabling us to account for the long term band of stasis that is observed in long term data sets (Fig. 7.x). The upshot is that while the single burst model produces some key features that characterize data (long term stasis or long delays in the appearance of substantial diversification), the model is still a work in process that awaits realistic modeling for short term behavior of the mean. The same realization applies to the model that we shall consider next.

13.5 Multiple bursts model

This model corrects another deficiency of the displaced optimum model that was not addressed by the single burst model. The present model allows multiple displacements of the optimum. ‘Multiple bursts’ refers to this key feature of the model. Multiple displacements of an intermediate optimum can occur at random during the lifespan of a lineage. Each of these displacements, may evoke a burst of evolution, but these bursts are interspersed by potentially long intervals during which the optimum is stationary and evolution is static (Fig. 13.13). As in the single burst model, the occurrence of the displacements is generated

Figure 13.13. Simulation of the multiple burst model. The orange line shows the position of the optimum. Parameter values are $\sigma_p^2 = 5$, $\sigma_d^2 = 37$, and $\lambda = 0.005$. The average waiting time for a displacement of the optimum is $1/\lambda = 200$ generations. The 99% confidence limits for a normal distribution with mean zero and variance $\sigma_p^2 = 5$ are shown in violet. In this example, three displacements (at about 290, 300, 520 generations) move the lineage mean in the same direction, but that will not be the case in most runs.



generated at random intervals by a Poisson process. The general characteristic of the process is that the probability of observing n displacement events in some time interval of length t is given by

$$p_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!},$$

where λ is a constant, the long term average rate of displacement (Bailey 1964). The expected number of displacements in an interval of length t is λt . Once a displacement occurs, its magnitude, d , is determined by a draw from a normal distribution with a mean of zero and a variance of σ_d^2 . Let the trait mean be zero at the start of a time interval of length t . At the end of that interval the probability distribution for the magnitude of divergence, \bar{z} , is a compound distribution given by

$$\Phi(\bar{z}_t) = \sum_{n=0}^{\infty} \frac{\exp\left(\frac{\bar{z}^2}{2n\sigma_d^2}\right)}{\sqrt{2\pi n\sigma_d^2}} e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

(Uyeda et al. 2011). As before, we add an element of white noise variation, σ_p^2 , to the trait mean that represents, for example, variance arising from the balance between drift and stabilizing selection, as well as contributions from other sources. Our probability distribution becomes

$$\Phi(\bar{z}_t) = \sum_{n=0}^{\infty} \frac{\exp\left(\frac{\bar{z}^2}{2(\sigma_p^2 + n\sigma_d^2)}\right)}{\sqrt{2\pi(\sigma_p^2 + n\sigma_d^2)}} e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

(Uyeda et al. 2011). The behavior of this model differs from the single burst model in an important way. Instead of the distribution converging on a limiting variance, the variance expands perpetually. Nevertheless, this model can capture the essential features of the data summarized in Fig. 7.x, if displacements are extremely rare. Under that condition, lineage means can reside for tens thousands of generations within boundaries approximated by the confidence limits of a normal distribution with

variance σ_p^2 . On longer time scales, lineage means can diverge beyond those limits. Figure 13.12 shows the distributions of lineage means under the same parameters as in Fig. 13.12, but here the average waiting time for displacement of the optimum is much longer, 100,000 generations, rather than 200. If displacements of the optimum are appreciably rarer than in Fig. 13.12, a initial period of relative stasis can last hundreds of thousands or millions of generations, followed by a period of more substantial divergence that can closely approximate the pattern in Fig. 7.x.

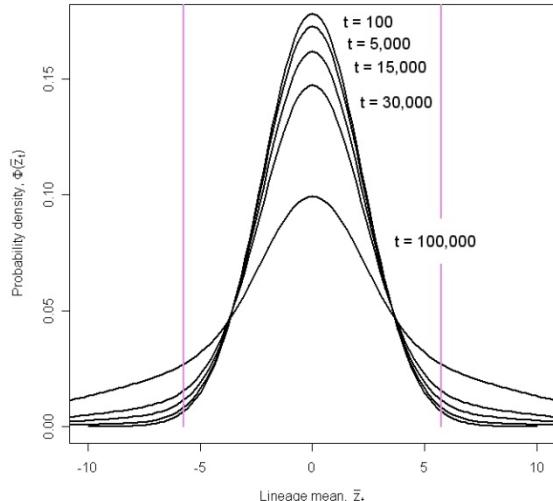


Figure 13.12 Distributions of lineage means after three different elapsed times according to the multiple burst model. Parameter values are $\sigma_p^2 = 5$, $\sigma_d^2 = 37$, and $\lambda = 0.00001$. The average waiting time for a displacement of the optimum is $1/\lambda = 100,000$ generations. The 99% confidence limits for a normal distribution with mean zero and variance $\sigma_p^2 = 5$ are shown in violet.

Uyeda et al. (2011) used maximum likelihood to fit this model to the data shown in Fig. 7.x and estimated that $1/\hat{\lambda} = 10^{7.3976}$. In other words, the average waiting time to a displacement of the optimum was nearly 25 million years, the average magnitude of displacement was only xxx *need value on standard scale*. Thus, displacements of the optimum appear to be very rare indeed. Substantial divergence of the magnitude observed in time intervals longer than 10 million years, apparently arise as consequence of repeated instances of rare movements of the optimum, each displacement being of modest magnitude. Substantial macroevolutionary divergence arises from the accumulation of repeated, rare, modest bursts of evolution.

The multiple bursts model was the best performing of three models compared by Uyeda et al. 2011 ... * more discussion here*

Before leaving the multiple burst model, we need to recall that it shares certain simplifying features with its more constrained sibling, the single burst model. Both of these models employ a white noise process because of the property that its contribution of among-lineage variance will be time invariant, even though its characterization of short term behavior of the mean is unrealistic. Secondly, as before, we have assumed that the response of the trait mean to the displaced optimum is instantaneous rather than asymptotic. Both of these features need to be replaced by more realistic assumptions, which will inevitably mean more complicated probability expressions to replace 13.x and 13.y *equations for the two prob distributions*.

13.5 Synapomorphy resulting from a rare displacement of an adaptive peak

A character that changes just once in the phylogenetic history of a clade constitutes an ideal indicator of phylogenetic relationships and is known as a synapomorphy (Hennig 1966) ... The most plausible model of the process that underlies synapomorphy is the one discussed in 13.3, a single displacement of an adaptive peak ... rapidity of adjustment in phenotypic mean ... the most puzzling aspect of synapomorphic peak movement is why the displacement should occur just once ... {direct reader to ecological discussion in Chap 19} ... But whatever the causes of synapomorphic peak movement, the phenomenon is sufficiently common that systematists are routinely able to find characters that adhere to this mode of evolution.

13.6 Testing models of peak movement

... {a summary of the Estes & Arnold 2007 results using the Gingerich data = Fig. 13.5 = based on summary figure in Estes & Arnold 2007}

13.7 Double OU process

... {present and discuss here depending on what is published by Schweder et al}

13.8 Other peak controller processes

13.0 Evidence for fluctuation and trends in the position of intermediate optima

A variety of sources suggest that even though the movement of the adaptive peak is conservative enough to be responsible for stasis in the fossil record, small scale movements can be detected in ecological

studies. At the finest temporal scale, we can estimate the speed of movement from studies that record trait means generation-to-generation, on the basis of either of two suppositions: that means closely tracks the optimum or that lag has stabilized so that the position of the mean mirrors the position of the optimum. In a sample of 4 such studies (Table 13.x), trait means moved an average of 0.09 phenotypic standard deviations per generation over an average of 11 generations. The average per generation variance in movement of the optimum, σ_θ^2 , estimated from the same studies, was 0.10.

Table 13.x. Estimates of the average rate ($\Delta\theta$) and sample variance (σ_θ^2) of per-generation shifts in the selective optimum, in units of phenotypic standard deviation. The number of change-time intervals in each study is shown in the n column. Absolute values of average rates are presented in the $\Delta\theta$ column. From Jones et al. 2011.

Species	Trait ^a	Number of generations	n	Average rate ($\Delta\theta$)	Variance (σ_θ^2)	Ref
Grayling (<i>Thymallus thymallus</i>)	Age-specific body length	13	6	0.184	0.582	Haugen and Vøllestad (2001), Fig. 3
	at five different ages	13	7	0.092	0.118	
		13	6	0.010	0.036	
		13	6	0.021	0.060	
		13	6	0.045	0.079	
Darwin's finch (<i>Geospiza fortis</i>)	Body size	6	24	0.068	0.067	Grant and Grant (2002), Fig. 1
	Beak size	6	24	0.000	0.061	
	Beak shape	6	25	0.078	0.040	
Darwin's finch (<i>Geospiza scandens</i>)	Body size	5	24	0.197	0.062	
	Beak size	5	25	0.137	0.100	
	Beak shape	5	24	0.290	0.025	
Great tit (<i>Parus major</i>)	Fledgling mass	19	34	0.003	0.0003	Garant et al. (2004), Fig. 3
Mean (SE)		10.8		0.094 (0.026)	0.103 (0.045)	

^aPhenotypic means were used for all estimates except the great tit dataset, where breeding values were used.

Another perspective is provided by studies that estimate selection coefficients from the same population over a period of years. A sample of time series from studies of bird populations,

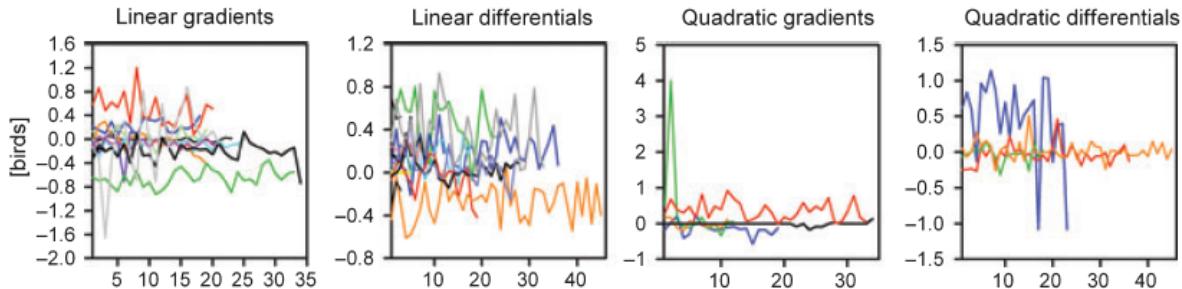


Figure 13.y. Estimates of selection coefficients year-by-year in long term studies of bird populations. Estimates of 4 different coefficients are plotted as a function of consecutive year in the study. Time series for 6 different studies are plotted in different colors. Quadratic coefficients were estimated in only 4 studies (Siepielski et al. 2009).

among the best studied taxa in this regard, is shown in Fig. 13.y. Fluctuation in selection coefficients is a conspicuous feature of these and other time series compiled by Siepielski et al. 2009. In some case, the fluctuations in directional selection (β) are about a long term trend of zero. This pattern is consistent with fluctuation of an adaptive peak (θ) about a stationary position ($\bar{\theta} = 0$). In other cases, the fluctuations are about a more less constant value of β , suggesting that the peak is moving at a constant rate, with the trait mean lagging behind that moving optimum. Estimates of the variation in β , which include contributions from both temporal fluctuation in direction and strength (Siepielski et al. 2009, Fig. 4a), suggest a overall median value of about $\sigma_\beta^2 = 0.01$, with most values falling in the range 0-0.25. These same studies also indicate that the configuration (shape and orientation) of the AL fluctuates through time. Although the available data are meager and difficult to interpret, they suggest that fluctuations of this sort are modest compared to fluctuations and trends in θ . We shall refer to these results in the sections that follow, for they provide a means of validating choices of parameter values in modeling and simulations studies ****need to systematically do that cross-checking****.

The challenge that faces us is to find a model for peak movement capable of accounting for the ecological results that we have just reviewed, but that can also account for the evidence of stasis and evolutionary bursts that springs from data on a paleontological time scale. In the sections that follow, we review a set progressively more complicated models that takes to this goal.