

Resolving the Paradox of Stasis: Models with Stabilizing Selection Explain Evolutionary Divergence on All Timescales

Suzanne Estes^{*} and Stevan J. Arnold[†]

Department of Zoology, Oregon State University, Corvallis, Oregon 97331

Submitted July 21, 2006; Accepted August 22, 2006;
Electronically published January 4, 2007

Online enhancement: zip archive.

ABSTRACT: We tested the ability of six quantitative genetic models to explain the evolution of phenotypic means using an extensive database compiled by Gingerich. Our approach differs from past efforts in that we use explicit models of evolutionary process, with parameters estimated from contemporary populations, to analyze a large sample of divergence data on many different timescales. We show that one quantitative genetic model yields a good fit to data on phenotypic divergence across timescales ranging from a few generations to 10 million generations. The key feature of this model is a fitness optimum that moves within fixed limits. Conversely, a model of neutral evolution, models with a stationary optimum that undergoes Brownian or white noise motion, a model with a moving optimum, and a peak shift model all fail to account for the data on most or all timescales. We discuss our results within the framework of Simpson's concept of adaptive landscapes and zones. Our analysis suggests that the underlying process causing phenotypic stasis is adaptation to an optimum that moves within an adaptive zone with stable boundaries. We discuss the implication of our results for comparative studies and phylogeny inference based on phenotypic characters.

Keywords: adaptive landscape, macroevolution, microevolution, phenotypic evolution, quantitative genetic models, comparative methods.

The common observation of evolutionary stasis (persis-

tence of morphospecies over geological time) seems a paradox when juxtaposed with the observation that abundant genetic variation is available for most traits in contemporary populations (Hansen and Houle 2004). The existence of prolonged stasis has been appreciated since Darwin (1859), who considered its prevalence, coupled with the abrupt appearance of new species in fossil sequences, problematic for his theory of evolution by gradual modification. The problem of stasis has been tackled from both theoretical and empirical standpoints over the past 60 years (e.g., Simpson 1944, 1953; Haldane 1949; Eldredge and Gould 1972; Lande 1976; Gingerich 1983; Lynch 1990). Recent renewed interest in evolutionary rates has highlighted both the prominence of evolutionary stasis and its paradoxical nature in light of the increasing number of cases that document rapid adaptation on short timescales (e.g., Hendry and Kinnison 1999; Gingerich 2001; Kinnison and Hendry 2001; Sheets and Mitchell 2001; Hansen and Houle 2004).

Despite this long history of interest, no generally accepted consensus has been reached regarding the likely cause of stasis. Underlying factors proposed to account for stasis on a geological timescale include protracted periods of stabilizing selection (Charlesworth et al. 1982; Lynch 1990), genetic and developmental constraints (Hansen and Houle 2004; Blows and Hoffmann 2005), competition for resources, selective constraints due to coevolution (reviewed in Mayr 2001, chap. 10), canceling of "positive" and "negative" evolutionary trajectories over time (Stanley and Yang 1987; Gingerich 2001), mathematical artifact (Bookstein 1987; Roopnarine 2003), habitat selection (Partridge 1978), and complexities involved with evolution in metapopulations (Eldredge et al. 2005). Despite this diversity of proposals, we accept the arguments of Charlesworth et al. (1982) that stabilizing selection is the most likely explanation for stasis. Attempts to use genetic constraints, for example, as a universal explanation for stasis confront several difficulties. First, genetic studies of natural populations have usually revealed ample quantitative ge-

^{*} Present address: Department of Biology, Portland State University, Portland, Oregon 97201; e-mail: estess@pdx.edu.

[†] E-mail: arnolds@science.oregonstate.edu.

netic variation for most kinds of traits, although much of this variation may be deleterious in any given environment (Lynch and Walsh 1999, chap. 12; Blows and Hoffmann 2005). Second, contemporary populations routinely mount rapid evolutionary responses to changing environmental conditions (reviewed in Kinnison and Hendry 2001; but see Bradshaw 1991; for a fossil example, see Bunce et al. 2005). Third, while genetic constraints on evolution exist and can affect evolution in the short to medium term, these constraints can evolve and so are unlikely to account for the lengthy periods of stasis that characterize many fossil lineages (Charlesworth et al. 1982). In other words, populations are often well equipped genetically to respond to at least short-term ecological challenges, refuting the notion of omnipresent, internal constraints on evolution. For these and other reasons (for discussion, see Charlesworth et al. 1982), stabilizing selection emerges as the leading contender among explanations for stasis. Although stabilizing selection is undoubtedly an ingredient in the production of stasis, we need to move to the next stage and examine the power of alternative processes of stabilizing selection to explain divergence data.

Most studies of stasis have focused on the evolutionary pattern without investigating the processes that produce that pattern. Most data on evolutionary rates have been summarized without reference to models, so little insight is gained into the causes of stasis. In those instances in which models are employed, they are often ad hoc models with arbitrary parameters that have no direct connection to studies of evolutionary processes (e.g., Bookstein 1987; Gingerich 2001; Kinnison and Hendry 2001). Such models are capable of producing stasis, but one cannot determine whether the parameters employed are consistent with empirical studies of inheritance, selection, and population size. Furthermore, the quantification and study of evolutionary rates have been fraught with both semantic and statistical difficulties (Brock 2000, chap. 18; Kinnison and Hendry 2001; Roopnarine 2003; Hansen and Houle 2004; Eldredge et al. 2005), culminating in a recent suggestion that few, if any, useful inferences regarding long-term evolutionary pattern or process have yet been drawn from even extensive compilations of rate data (Roopnarine 2003). This pessimistic view arises in part from a general failure to connect empirical studies of evolutionary rates of divergence with the most meaningful models of evolutionary change—those with ties to evolutionary genetics. Ironically, the empirical concentration on evolutionary rates, rather than on divergence, has thrown data analysis out of register with the most informative models of evolutionary process. In contrast to these classes of failure, quantitative genetic models have been successfully used to analyze particular instances of phenotypic evolution

(Lande 1976; Lynch 1990; Hansen 1997), often rejecting neutrality in favor of stabilizing selection. Nevertheless, the quantitative genetic models that have been produced over the past 30 years have generally been applied piecemeal to one or a few case studies. It is fair to say that most models have not experienced serious encounters with data. Consequently, our goal is to compare the utility of the entire range of quantitative genetic models so far produced by moving beyond the analysis of individual cases. In particular, we use Gingerich's (2001) extensive compilation of divergence data to explicitly test alternative explanations of stasis.

Surveys of phenotypic evolution often reveal rapid change on short timescales but stasis on long timescales. While net evolutionary divergence within lineages generally increases with time, the average evolutionary rate of change is remarkably slow (Kinnison and Hendry 2001). In nature, many lineages show little or no net divergence in numerous phenotypic traits over timescales ranging from 10^6 to 10^8 generations. Extreme examples are provided by the so-called living fossils (e.g., coelacanths, bracken ferns, millipedes, horseshoe crabs). In other cases, directional change occurs over shorter timescales but is later canceled by evolution in the opposite direction, so a relatively stationary mean phenotype is maintained in the long term. An example of this phenomenon, termed "zig-zag evolution," is provided by Stanley and Yang's (1987) observations of morphological change for three lineages of Pliocene bivalves. While an overall pattern of stasis is a common finding in analyses of fossil data series, as discussed by Simpson (1944, 1953), a universal attribute of such time series is that they reveal considerable fluctuations in average phenotypes on the smallest resolvable timescales. We shall discuss this phenomenon in light of our findings and in terms of Simpson's (1944) concept of adaptive landscapes and zones.

Our general aim is to focus on the broad picture of net evolutionary divergence accumulating in lineages over time and determine whether a single model of evolutionary process can account for the overall pattern of divergence on all timescales. By using explicit genetic models of divergence, we hope to reveal the microevolutionary underpinnings of stasis. This article has two immediate goals. First, we wish to determine whether a single model of evolutionary process can account for the tempo of evolution on all timescales. Second, we wish to determine whether such a best-fitting model can be found in which parameters are in register with estimates of inheritance, selection, and population size from contemporary populations.

Data

We used the database of 2,639 "microevolutionary," "historical," and paleontological evolutionary rates compiled

by Gingerich (2001; plotted in his figs. 1, 8B–8D) for our analyses. This database consists of a sampling of divergence for various characters in a wide range of taxa (e.g., planktonic foraminiferans, ceratopsid dinosaurs) over timescales ranging from 1 to 10 million generations from 44 sources. These historical and paleontological data represent divergence calculated for various morphometric characters (e.g., shell size and shape); microevolutionary divergence was calculated for a mix of morphological, behavioral, and life-history characters (e.g., see table 1 in Hendry and Kinnison 1999). We omitted rates estimated from laboratory experimental evolution studies (i.e., those from Gingerich's [2001] fig. 8A) so that all divergence rates represent instances of evolution in nature. The Gingerich (2001) data were expressed as evolutionary rates, reported in Haldane units and corresponding time intervals (in generations). The Haldane rate is D/I , where D is the difference between two sample means divided by the pooled standard deviation of the samples and I is the time that has elapsed between the samples in generations (Gingerich 1983). Using the Gingerich data, we calculated total morphological divergence, D , and plotted this against the corresponding time interval, I (fig. 1A). Because of the compression in divergence values on the raw scale, we followed Gingerich (2001) and Kinnison and Hendry (2001) and expressed the divergence data on the \log_{10} scale (fig. 1B). For each of these plots, we estimated a 99% confidence ellipse for the bivariate mean (fig. 1). These ellipses define a band of acceptability for assessing the performance of each model. In other words, when the predictions of a model fall outside this band, we can be reasonably sure that those predictions are rarely met in nature.

As in most paleontological data, the Gingerich data set contains a mix of both allochronic (within-lineage) and synchronic (between-lineage) data. Contamination with between-lineage values inevitably arises when an undetected second lineage invades the site of sampling and replaces the first lineage. The effects of this kind of contamination on our analyses are probably minor for two reasons. First, dramatic instances of one lineage replacing another will appear as a sudden, large spurt in divergence and will be excluded from the confidence ellipse on that basis. Second, when less dramatic instances of lineage replacements are inadvertently included, the effect will be to average divergence across lineages. This latter effect may eliminate some extreme values from the total sample, but it is not likely to change our band of acceptability. We also note that the individual data points in figure 1 are not independent. When a long time series was available, Gingerich (2001) took multiple but nonoverlapping samples from that series. Nevertheless, the resulting sampling covariance should not complicate our results because the

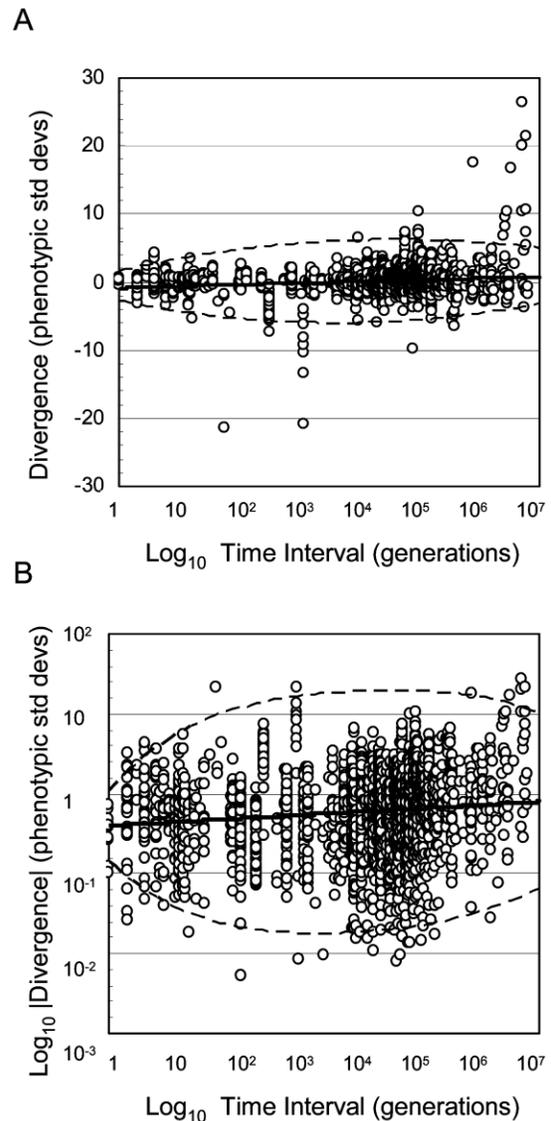


Figure 1: A, Plot of divergence as a function of time interval. Divergence is measured as the difference between the average trait values of an ancestral and a descendant population, expressed in units of pooled phenotypic standard deviation. The slope of the fitted regression line is 0.2407 (SE = 0.0268), $P < .0001$. 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.

confidence ellipses in figure 1 are used only for graphical comparisons with model predictions.

We also calculated the effect of errors of measurement on our mode of analysis. Error is especially likely in the

estimation of both phenotypic divergence and time interval on a geological timescale. Errors of these kinds arise because of small sample sizes and methodological errors in dating. As upper bounds on the \log_{10} scales, we assumed that all such sources of error might have inflated the variances in both divergence and time interval by as much as 50%. We recomputed the 99% confidence ellipse after subtracting the inflated portion of the variance on both axes and obtained a new ellipse (without inflation) that was nested inside the original 99% ellipse. This new ellipse is almost exactly comparable to a 95% confidence ellipse calculated without assuming inflation of bivariate variance (see zip archive in the online edition of the *American Naturalist*). In other words, even appreciable measurement error would have a minor effect on the boundary of the confidence ellipse that we used in assessing the fits of the various process models. Consequently, error of measurement is unlikely to have affected our conclusions.

Theoretical Background

We have two aims in the sections that follow. First, we introduce the measures of inheritance and selection (h^2 , θ , ω^2) that are used in the models and show how these selection parameters are related to commonly estimated parameters of selection (β , γ). The connection between θ and ω^2 , on the one hand, and β and γ , on the other, is embodied in the concept of the adaptive landscape, which we briefly review. Second, we introduce the models themselves and show how their predictions can be brought into register with Gingerich's (2001) data.

We use a series of quantitative genetic models developed by Lande and colleagues to analyze and interpret data on divergence in morphological characters (Lande 1976, 1985, 1986; Lynch and Lande 1995; Lande and Shannon 1996). These models are based on a theory of stochastic processes originally devised to account for Brownian motion and the diffusion of particles through a permeable membrane (Karlin and Taylor 1981, chap. 15). The models predict how much the mean of a phenotypic trait is expected to change over a specified number of generations. To make this prediction, aspects of inheritance, selection, and population size are held constant. Stochasticity arises in the models from finite population size. Because the population is finite, accidents of parentage introduce random variation in the trait mean so that it tends to fluctuate or drift from generation to generation. The statistical properties of this random variation can be specified. Imagine a set of replicate populations derived from the same ancestral population and maintaining the same parameters of inheritance, selection, and size. Stochastic process theory allows us to specify the variance among populations in their trait means after some number of generations have elapsed.

Under some models, this variance in means is ever increasing, but in others, it reaches and then maintains a stable value. In either case, we can use the variance formula to compute a standard deviation, and hence a confidence limit, about the expected mean. In summary, each genetic model predicts how much divergence can be expected after t generations and how much variation can be expected about that average. Those expectations can then be compared with the observed amounts of divergence to determine whether the model and a particular set of parameter values are capable of accounting for the data.

All of the models that we will use rely on a common set of simplifying assumptions. We restrict our focus to the case of a single, continuously distributed phenotypic character, because virtually all of the relevant empirical data are in a univariate format. The character z is assumed to be normally distributed (or to have been transformed to normality) with mean \bar{z} and phenotypic variance σ^2 before selection. Inheritance is characterized by a heritability h^2 . We will assume that the additive genetic variance of the character, $h^2\sigma^2$, remains constant even over long stretches of evolutionary time because of a balance among the forces of selection, drift, mutation, and/or migration (Jones et al. 2003, 2004). We also assume that selection favors an intermediate optimum phenotype θ and can be characterized by a Gaussian function such that the expected fitness of an individual with phenotype z is given by

$$W(z) = \exp\left[-\frac{(z - \theta)^2}{2\omega^2}\right], \quad (1)$$

where ω is the width of the Gaussian function; ω^2 is analogous to a variance. Corresponding to this individual selection function is an adaptive landscape that relates average fitness in the population to the mean of the phenotypic character. This landscape is also a Gaussian function with an optimum at θ and a "variance" of $\omega^2 + \sigma^2$:

$$\bar{W} \propto \exp\left[-\frac{(\bar{z} - \theta)^2}{2(\omega^2 + \sigma^2)}\right] \quad (2)$$

(Lande 1976, 1979). The strength of stabilizing selection is readily visualized with these formulations. For example, if we standardize a character so that it has unit standard deviation ($\sigma = 1$), then stabilizing selection is easily visualized as a normal curve with a width ("standard deviation") that is some multiple of σ . Thus, $\omega^2 = 99$ corresponds to very weak stabilizing selection in which the width of the adaptive landscape, $(\omega^2 + \sigma^2)^{1/2}$, is 10 times wider than the character distribution. At the other extreme, $\omega^2 = 1$

corresponds to very strong stabilizing in which the width of the adaptive landscape, $(\omega^2 + \sigma^2)^{1/2}$, is $2^{1/2} \approx 1.4$ times wider than the character distribution.

Standard measures of selection can be obtained by taking the first and second derivatives of the landscape with respect to the character mean. Thus, the first derivative, or slope of the landscape, evaluated at \bar{z} is

$$\beta \equiv \frac{\partial \ln \bar{W}}{\partial \bar{z}} = (\omega^2 + \sigma^2)^{-1}(\theta - \bar{z}), \quad (3)$$

a measure of directional selection (Lande 1979, 1980). The second derivative, or curvature of the landscape, evaluated at \bar{z} is

$$H \equiv \frac{\partial^2 \ln \bar{W}}{\partial \bar{z}^2} = -(\omega^2 + \sigma^2)^{-1}, \quad (4)$$

a measure of stabilizing selection (Lande 1979). An important consequence of assuming normality of the character distribution and a constant Gaussian form for the selection function is that, under most of the models we will use, one can specify the distribution $\Phi(\bar{z}_t)$ of character means for a set of replicate populations at any generation in the future.

The first and second derivatives of the adaptive landscape, β and H , can be estimated by making a quadratic approximation to the individual selection surface, which in the univariate case is

$$W(z) = 1 + \beta z + \left(\frac{1}{2}\right)\gamma z^2, \quad (5)$$

in which individual fitness $W(z)$ has been transformed so that its mean is 1 and trait values z have been transformed so that the mean is 0 and variance is 1 (Lande and Arnold 1983). The strength of the direct effects of quadratic selection is described by γ , the standardized quadratic selection gradient. Assuming that the individual selection surface is approximately Gaussian, one can estimate ω^2 using equation (4):

$$-(\omega^2 + \sigma^2)^{-1} = \gamma - \beta^2 \quad (6)$$

(Phillips and Arnold 1989; the derivation of this result is available in the zip archive). The distance from the character mean to the optimum can be estimated using the formula

$$|\bar{z} - \theta| = \left| \frac{\beta}{-\gamma} \right|, \quad (7)$$

assuming that a quadratic function gives a reasonable approximation to the individual selection surface (Phillips and Arnold 1989, their eq. [11]). We now turn to quantitative genetic models that are cast in terms of these parameters. The models are described in detail by the cited authors. The behavior of the models can be visualized using a PowerPoint presentation available in the zip archive. A spreadsheet showing how numerical evaluations of the models were compared with data is provided in the same zip archive. Our aim here is to review the expressions, derived from these models, for expected divergence as a function of time and the variation that is predicted about those expectations.

Neutrality

In a population of effective size N_e , in the absence of selection, the variance created each generation by drift among replicate populations is $h^2\sigma^2/N_e$. After t generations, the replicates derived from a common ancestor at $t = 0$ with an initial mean phenotype of $\bar{z}_0 = 0$ will be normally distributed about a mean of 0 with a variance of

$$\text{Var}(\bar{z}_t) = \frac{h^2\sigma^2 t}{N_e} \quad (8)$$

(Lande 1976; fig. 2). A different model of neutral evolution has been proposed by Lynch and Hill (1986), in which the

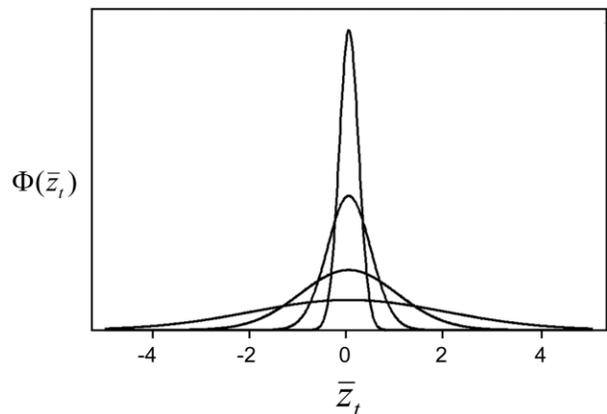


Figure 2: Divergence in phenotypic mean under neutrality. Beginning at time 0, a series of replicate populations diverge due to drift. The curves show the distribution of phenotypic means, $\Phi(\bar{z}_t)$, after different intervals of time have elapsed. The variance of such a distribution at generation t is $\text{Var}(\bar{z}_t)$. From the uppermost to the lowermost curve, the elapsed intervals are 200, 1,000, 5,000, and 20,000 generations. The scale on the X-axis is in units of phenotypic standard deviations. Heritability is 0.4, and $N_e = 1,000$.

variance given by equation (8) is the product of mutational variance, σ_m^2 , and t . With respect to testing, the models are comparable in the sense that one assumes that either $h^2\sigma^2/N_e$ or σ_m^2 is constant over time (Lynch 1990).

Displaced Optimum

Consider the case of a population of effective size N_e evolving in response to the adaptive landscape described in equation (2) (Lande 1976). Let the mean of the phenotypic character at generation 0 be \bar{z}_0 . Now suppose that the optimum is displaced some distance from \bar{z}_0 . The initial distance from the mean to the optimum is $(\bar{z}_0 - \theta)$, and for mathematical convenience, we can let the displaced optimum take a value of 0 (fig. 3). After some number of generations t , the expected value of population means, $E[\bar{z}_t]$, will have approached the optimum, so that

$$E[\bar{z}_t] = \bar{z}_0 \exp\left[-\left(\frac{h^2\sigma^2}{\omega^2 + \sigma^2}\right)t\right]. \quad (9)$$

If we consider a series of replicate populations with these same characteristics at generation t , they will deviate from this expectation because of the stochasticity induced by finite population size. The replicates will be normally distributed about $E[\bar{z}_t]$ with a variance given by

$$\text{Var}(\bar{z}_t) = \frac{\omega^2 + \sigma^2}{2N_e} \left[1 - \exp\left[-2\left(\frac{h^2\sigma^2}{\omega^2 + \sigma^2}\right)t\right]\right]. \quad (10)$$

After many generations, the expected value of the phenotypic mean will coincide with the optimum, and the dispersion of the replicate populations will achieve a constant variance of $(\omega^2 + \sigma^2)/2N_e$ (Lande 1976; fig. 3). This constant variance is a manifestation of an equilibrium between drift, which tends to cause divergence among replicates, and stabilizing selection, which tends to pull replicate means toward the optimum.

Brownian Motion of Optimum

Next, consider the situation in which the optimum, θ , undergoes Brownian motion but with no long-term change in the expected value of its mean position. In other words, the position of the optimum at generation $t + 1$ is normally distributed with a mean equal to θ , and a variance σ_θ^2 and with the variance among replicates accumulating through time, so that at generation t it is $t\sigma_\theta^2$. As in the preceding model (Lande 1976), the trait experiences stabilizing selection in any given generation as well as directional selection proportional to the distance between the trait mean, \bar{z}_t , and the optimum. Letting the trait mean

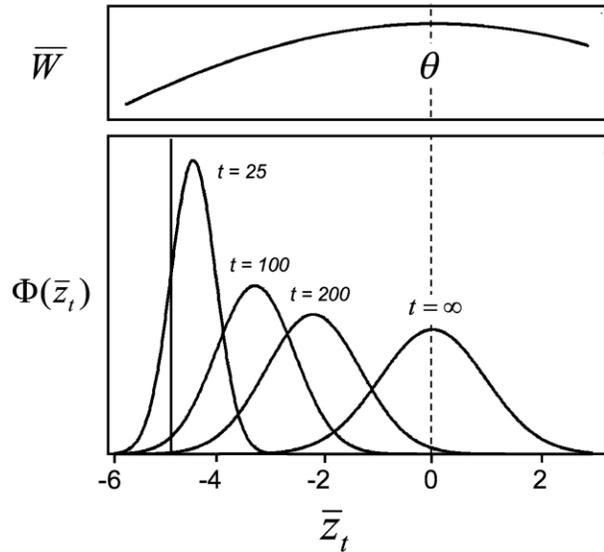


Figure 3: Divergence in response to a displaced optimum. The top panel shows an adaptive landscape with an optimum that has been displaced 5 phenotypic standard deviations from the phenotypic mean, shown as a vertical line. The bottom panel shows the distribution of phenotypic means, $\Phi(\bar{z}_t)$, 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.

initially coincide with the position of the optimum, so that $\bar{z}_0 = \theta_0 = 0$, after some number of generations, t , the expected value of the population mean will coincide with the long-term expectation for the optimum, so that $E[\bar{z}_t] = E[\theta_t] = 0$. The replicates will be normally distributed about 0 with a variance given by

$$\text{Var}(\bar{z}_t) = t\sigma_\theta^2 + \frac{\omega^2 + \sigma^2}{2N_e} \left[1 - \exp\left[-2\left(\frac{h^2\sigma^2}{\omega^2 + \sigma^2}\right)t\right]\right]. \quad (11)$$

Moving Optimum

A simple way to introduce perpetual directional selection is to assume that the optimum moves at a constant rate k . We will also allow for an element of stochasticity by allowing motion with variance σ_θ^2 each generation about the trend line for θ . In other words, the optimum at generation t is $\theta_t = kt + \varepsilon_\theta$, where ε_θ is normally distributed with no correlation in time and with mean 0 and variance σ_θ^2 . In contrast to the preceding model, the movement of the optimum about the trend line is a Gaussian white noise process, so that the variance that the process contributes to \bar{z}_t does not increase linearly with time (Bürger and Lynch 1995; Lande and Shannon 1996; Bürger 2000). Eventually,

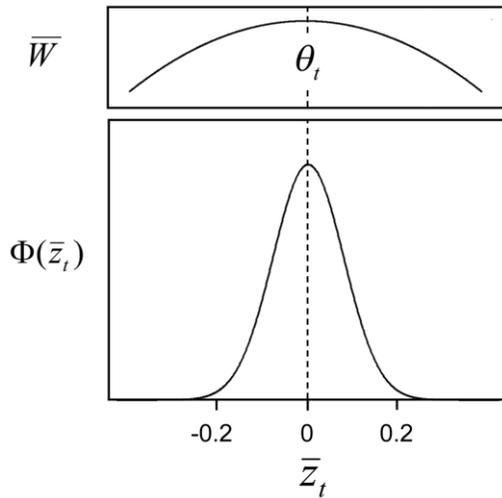


Figure 4: Diversification among replicate populations at generation t under the moving optimum model. The top panel shows the expected position of the optimum, θ_t . The bottom panel shows the distribution of phenotypic means, $\Phi(\bar{z}_t)$. The expected phenotypic mean lags so very slightly behind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

a balance is achieved between selection and drift, but in general, at generation t the expected character value in a series of replicate populations is normally distributed with a mean of

$$E[\bar{z}_t] = kt - k \left(\frac{\omega^2 + \sigma^2}{h^2 \sigma^2} \right) \left[1 - \exp \left[- \left(\frac{h^2 \sigma^2}{\omega^2 + \sigma^2} \right) t \right] \right] \quad (12)$$

and a variance of

$$\text{Var}(\bar{z}_t) = \left[\frac{\omega^2 + \sigma^2}{2N_c} + \frac{h^2 \sigma^2 \sigma_\theta^2}{2(\omega^2 + \sigma^2)} \right] \left[1 - \exp \left[-2 \left(\frac{h^2 \sigma^2}{\omega^2 + \sigma^2} \right) t \right] \right] \quad (13)$$

(Lynch and Lande 1995, their eqq. [14], [15]; fig. 4). When $k = 0$, this model describes the evolution of a character in response to an optimum that fluctuates in position according to a Gaussian white noise process.

Peak Shift

In all of the preceding selection models, the evolving mean, \bar{z} , tracked a single stationary or moving optimum. In peak shift models, the character mean evolves in response to two adaptive peaks (Barton and Charlesworth 1984; Lande 1985, 1986). The two peaks may differ in height and are

separated by an adaptive valley (fig. 5). A population of finite size spends most of its time in the vicinity of one adaptive peak, but it may drift into the valley, against the force of selection, and then be attracted to the second peak. Remarkably, the expected time for the population to evolve from the vicinity of the first Gaussian peak to the second Gaussian peak is virtually independent of the distance between the two peaks. The expected time T does, however, depend on the height of the first peak, \bar{W}_a , and the height of the valley, \bar{W}_v , so that

$$T \approx \frac{2\pi(\omega^2 + \sigma^2)}{h^2 \sigma^2} \left(\frac{\bar{W}_a}{\bar{W}_v} \right)^{2N_c}, \quad (14)$$

assuming that the curvature in the vicinity of the peak is comparable to the curvature at the valley (Lande 1985, 1986). In this model, $\omega^2 + \sigma^2$ is the “variance” of the first adaptive peak. The rate of peak shift is $1/T$, so in t generations, the expected number of replicates making the shift is t/T . We may treat peak shifting as a Poisson process (Bailey 1964). Thus, if we suppose that the two peaks are separated by a distance d , after t generations, a proportion $p = \exp(-t/T)$ of populations in a set of replicates will still reside on the first adaptive peak, while another proportion, $1 - p$, will have moved to the second peak. We assume that the second adaptive peak is at least 10% higher than the valley ($\bar{W}_b/\bar{W}_v \geq 1.10$), so that shifts from the second peak back to the first can be ignored. The expected amount of evolution will be

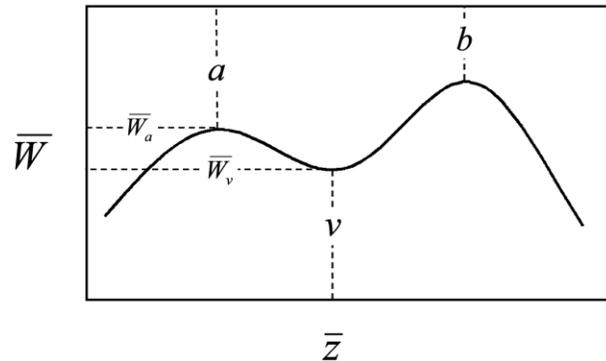


Figure 5: Adaptive landscape in the peak shift model, which has two adaptive peaks, a and b . Mean fitness \bar{W} is a function of the average phenotype in the population, \bar{z} . Two critical parameters in the peak shift model that affect the probability that the population mean will shift from peak 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 .

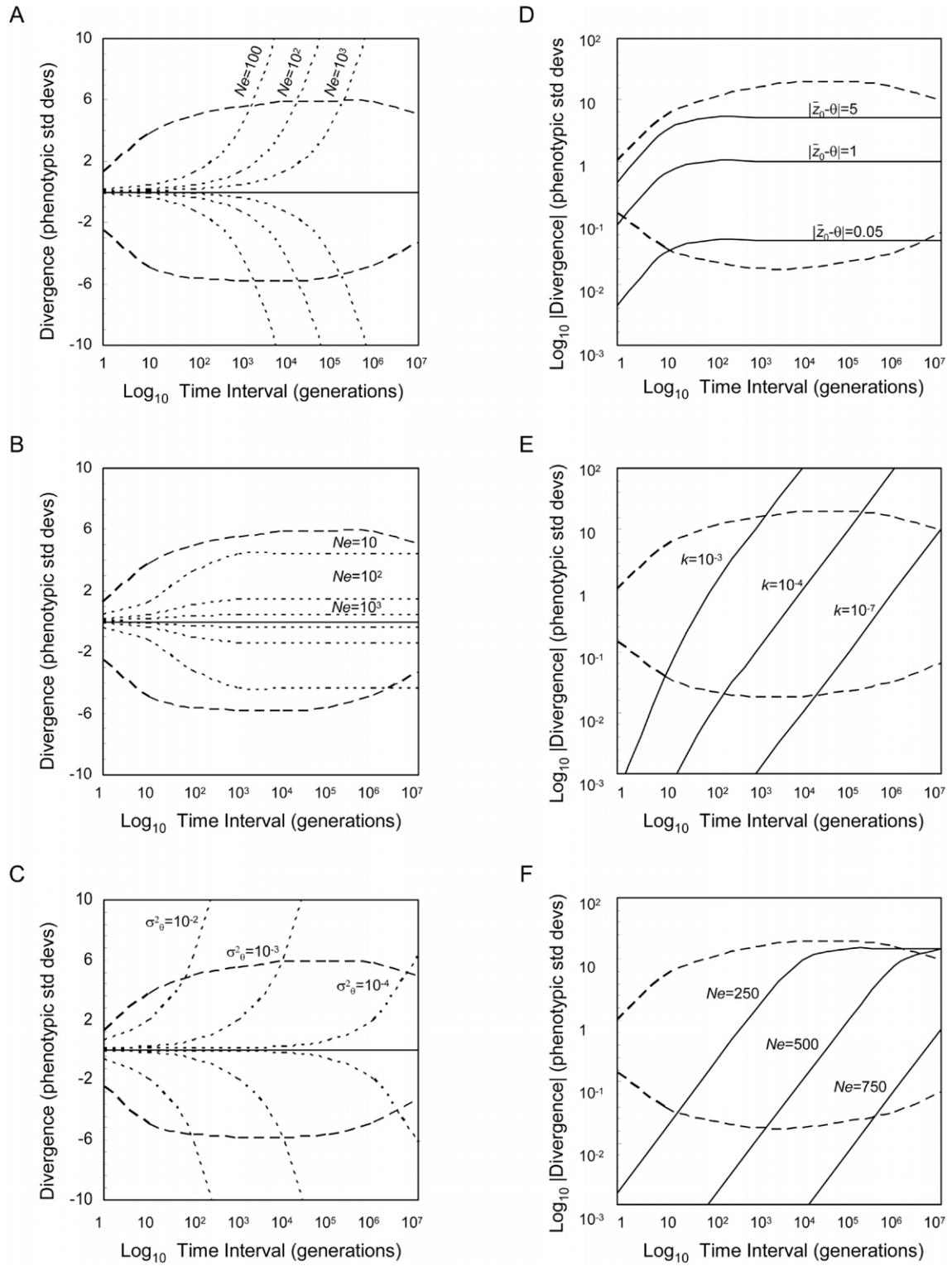


Figure 6: Example fits of neutral (A), white noise (B), Brownian motion of the optimum (C), displaced optimum (D), moving optimum (E), and peak shift (F) models to the divergence data. Heavy dashed lines in each panel represent the 99% confidence ellipse for the divergence data of Gingerich (2001). If model predictions fell within this ellipse, we considered the fit to the data acceptable. To illustrate how each model responds

$$E[\bar{z}_i] = d(1 - p), \quad (15)$$

assuming that the phenotypic value of a population mean directly under the first peak, at a , is 0. Let the “variance” of the second adaptive peak be the same as the first. The variance among replicate means at generation t , composed of a between-class variance and an average within-class variance, will be approximately

$$\text{Var}(\bar{z}_i) = p\bar{z}_i^2 + (1 - p)(d - \bar{z}_i)^2 + \frac{(\omega^2 + \sigma^2)}{2N_e}. \quad (16)$$

Methods

Implementation of Models

Stochastic models for evolution are often fitted to actual data by making calculations or simulations on a particular phylogeny for which data on phenotypic means are available at the tips of the tree (e.g., Hansen 1997; Schluter et al. 1997; Butler and King 2004). We did not take this approach. The data we used are for particular lineages that have been followed through time, obviating the need for a phylogeny. Furthermore, we treated the data en masse and defined the limits of observed divergence as a function of elapsed time (fig. 1). For each stochastic model, we simply asked whether predictions could fall within those observed limits under realistic parameters of inheritance, selection, and population size. The procedure can be thought of as a screening test. An advantage of our procedure is that we can screen models over a relatively large range of timescales (from 1 to 10,000,000 generations). Detailed model-fitting exercises usually examine model performance over a small fraction of this range.

To test the adequacy of each of the models, a realistic range of parameter values was systematically searched and the adequacy of each combination evaluated simply by assessing whether the predicted pattern of morphological divergence with time fell within the 99% confidence ellipse for the actual data (fig. 6). The search was conducted in a graphical way. We used a spreadsheet to calculate ex-

pected values for divergence as a function of elapsed time. To assess the fit of that curve to actual data, we superimposed the curve on the 99% confidence ellipse for the data, shown in figure 1 (spreadsheets used for model fitting are available in the zip archive). In nearly all instances, it was a simple matter to assess the fit. Either the curve fell inside the data ellipse over the entire timescale or it did not (fig. 6). To determine the sensitivity of the fit to variation in a particular parameter, we systematically varied the value of that parameter on the spreadsheet and observed the corresponding effect on the fit to the data. Note, however, that a model can fail for a second reason. Some models predict no divergence over a wide range of parameter values. Thus, under these conditions, the models fail because they fail to account for the wide range of divergence values shown by actual data (fig. 1).

Because we employed stochastic models of phenotypic evolution, each model predicts a range of values for divergence for any particular set of parameter values and divergence time. To visualize this variation about expected divergence, we computed the upper and lower 95% confidence limits using the standard formula, which assumes a normal distribution. To obtain confidence limits on the \log_{10} scale for expected morphological divergence, we first employed the approximation from Wright (1968, eq. [10.42]) to estimate the variance of morphological divergence on the \log_{10} scale: $\sigma'^2 = 0.4343 \log_{10}(1 + C^2)$, where $C = \sigma/\bar{z}$ is the observed coefficient of variation, the standard deviation divided by the mean, on the untransformed scale. This formula requires the assumption that the distribution of divergences is normal on the transformed scale, an assumption that is necessarily true by the nature of the models. Because the expected amount of divergence for the neutral model is 0 on the raw scale, we evaluated divergence under this model on the untransformed rather than the log scale. In our graphical tests of model performance, we used our spreadsheet to compute the 95% confidence limits for expected divergence as a function of time. These confidence limits were then superimposed on the data ellipse, along with the expected values predicted

to varying parameter values, a single parameter for which each model was found to be highly sensitive was varied within a realistic range, and three example fits are shown. Solid lines represent the model prediction, that is, amounts of expected evolutionary divergence (or \log_{10} divergence where noted) in units of phenotypic standard deviations (or \log_{10} phenotypic standard deviations) under the stated parameter values. For the neutral, white noise, and Brownian motion models (A–C), predicted net divergence is 0 across all timescales. The dotted lines in those panels represent the 95% confidence limits on either side of the prediction for each value of N_e shown. The 95% confidence limits for model prediction were used in assessing the fit to the data but are not shown in D–F. On the scales illustrated in this figure, the lines representing the confidence limits in these cases are, in most cases, extremely close to the model predictions. For the peak shift model (F), peak shifts rarely occur when $N_e > 200$, and as a consequence, there is virtually no divergence. Unless otherwise noted, central parameter values are used: $N_e = 1,000$, $h^2 = 0.4$, $\omega^2 = 3.214$. For the moving optimum model, $\sigma_b^2 = 0.0001$. For the peak shift model, the distance between the two adaptive peaks is $d = 10$, $W_a = 1.01$, and $W_v = 1.00$.

Table 1: Values of parameters of population size, inheritance, and selection that were used to assess model performance

Parameter	Symbol	Range evaluated	Reference(s)
Effective population size	N_e	10–100,000	Barrowclough 1980; Begon et al. 1980; Husband and Barrett 1992; Jorde and Ryman 1996; Storz et al. 2001; Turner et al. 2002
Trait heritability	h^2	.0001–.9	Mousseau and Roff 1987
Strength of stabilizing selection	ω^2	3–50	Kingsolver et al. 2001
Distance to the optimum	$ \bar{z} - \theta $.01–10	Kingsolver et al. 2001

Note: The range of values that was explored (range evaluated) is based on surveys of empirical studies conducted by authors indicated in the last column. The calculation of $|\bar{z} - \theta|$ is described in the text.

by the model (see the zip archive for spreadsheets and superimposed graphs).

Model Parameter Values

We employed realistic ranges of parameter values for inheritance, selection, and population size to evaluate each model (table 1). Values for narrow sense heritability (h^2) were chosen such that the majority of the range of such estimates ($n = 570$) for morphological characters reviewed by Mousseau and Roff (1987) was represented ($h^2 = 0.001$ – 0.9). The median heritability for morphological traits was 0.43, while that for life-history and behavioral traits was 0.25 and 0.28, respectively (table 1 in Mousseau and Roff 1987).

Constraints on the rate and direction of evolution can be produced by lack of appropriate genetic variation and by the effects of pleiotropic gene action, at least in the short term (e.g., Bradshaw 1991; Arnold 1992). By imposing a very small value for heritability ($h^2 = 0.001$), we evaluated the effect of strong genetic constraints on divergence in each model.

The strength of stabilizing selection (ω^2) and distance of the character mean from the optimum ($|\bar{z} - \theta|$) were estimated from data on β and γ given at the Web site associated with Kingsolver et al. (2001), using equations (4) and (6) (table 2). We used only those cases in which both β and γ had been estimated from the same data. We also omitted 17 cases in which the reported value for γ was 0 to avoid division by 0 in the calculation of ω^2 . Finally, we omitted two extreme outliers on either side of the distribution of ω^2 . Most values of ω^2 in contemporary populations are in the range -50 to 50 , with a strong mode at about 3 (fig. 7). Large negative values correspond to weak disruptive selection, and large positive values correspond to weak stabilizing selection. We note that a peculiar property of the ω^2 scale is that, in the vicinity of 0, going from positive to negative values, selection changes from strongly stabilizing to strongly disruptive. Nevertheless, more than 60% of the observed ω^2 values are positive, indicating stabilizing selection, and of these $\omega^2 < 20$ in

76% of cases. Consequently, we focused on $\omega^2 = 3$ – 50 , corresponding to $\omega \approx 1.7$ – 7.1σ in our calculations. In calculating $|\bar{z} - \theta|$ using equation (7), we used only those cases in which γ was negative (downward curvature) and excluded one large outlier, yielding a sample of 197 cases. In model implementation, we focused on the most frequently encountered range of values for $|\bar{z} - \theta|$, namely 0.01 – 10σ , centered about a modal value of <1 phenotypic standard deviation (fig. 8).

We chose a range of values for effective population size on the basis of a review of empirical estimates from a variety of organisms not considered to be rare or endangered (e.g., *Drosophila subobscura*, modern humans). Organisms that were sufficiently abundant in the fossil record to estimate character means and variances must have had large populations. Nevertheless, the effective size of a population can often be considerably smaller than the census size (e.g., Frankham 1995; Turner et al. 2002). We therefore used a central value of 1,000 but evaluated models over a broad range ($N_e = 10$ – $100,000$).

The peak shift model requires the ratio of two additional parameters, average population fitness at the original adaptive peak, \bar{W}_a , and in the valley between the two adjacent peaks, \bar{W}_v (fig. 5). No explicit estimates of \bar{W}_a/\bar{W}_v exist for any system. Following Lande (1985), we assumed that the adaptive peak was 1%–5% higher than the valley, so that $\bar{W}_a/\bar{W}_v = 1.01$ – 1.05 .

Table 2: Descriptive statistics for curvature of the adaptive landscape (H), the strength of stabilizing/disruptive selection (ω^2), and the absolute distance of the phenotypic mean from the optimum ($|\bar{z} - \theta|$)

	H	ω^2	$ \bar{z} - \theta $
Mean (SE)	-.108 (.022)	11.29 (6.462)	3.873 (1.293)
Median	-.043	1.676	1.100
Mode	.020	3.214	<1.000
Minimum	-2.576	-230	0
Maximum	1.916	1,735	238
N	355	355	197

Note: Statistics calculated from the database of Kingsolver et al. (2001). The mode of the ω^2 distribution was evaluated with bins 5 units wide.

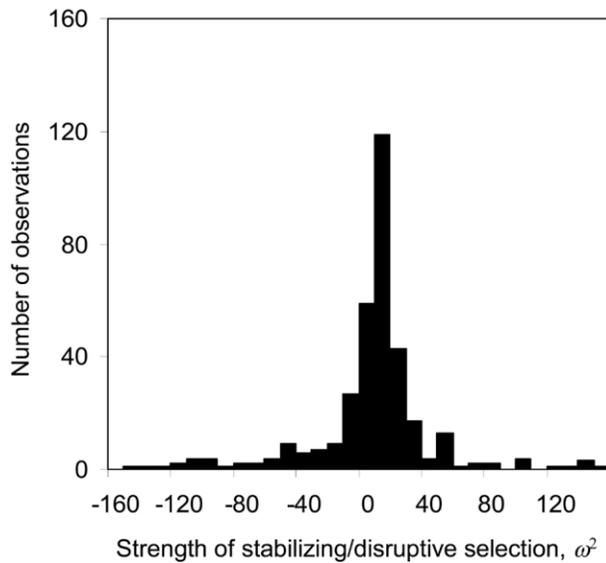


Figure 7: Strength of stabilizing/disruptive selection in natural populations ($n = 355$). Values of ω^2 were calculated from the database of Kingsolver et al. (2001). To improve visualization, in this graph the distribution was truncated at ± 200 (12 extreme values were omitted).

Under the moving optimum model, the rate of peak movement, k , could be equated with the rate of character evolution, but the degree of stochasticity in peak motion, σ_θ^2 , is an unknown parameter. A large value of σ_θ^2 is biologically untenable because it will promote extinction of the population (Lynch and Lande 1995). In particular, if $\sigma_\theta \geq \omega$, the population runs a high risk that the optimum will fluctuate so far from the population mean that the resulting decrement in mean fitness will result in extinction (Bürger and Lynch 1995). Viewing this argument on a geological timescale means that the condition $\sigma_\theta \ll \omega$ will be necessary for population persistence. We therefore employed a variety of values for both of these parameters in the ranges $k = 10^{-9}$ – 0.10 and $\sigma_\theta^2 = 10^{-9}$ – 0.10 , corresponding to moderate rates for peak movement and modest variation in the position of the optimum.

We began our evaluations of each model using a set of central values for common parameters ($N_e = 1,000$, $h^2 = 0.4$, $\omega^2 = 3.214$) and then systematically varied these and other parameters over the entire range of feasible values. Other methodological details are provided in the spreadsheet available in the zip archive.

Results

Pattern of Morphological Divergence

Although average morphological divergence increases with time (fig. 1), the net rate of evolution is remarkably slow,

as has been observed in previous studies (e.g., Lynch 1990; reviewed in Kinnison and Hendry 2001). In fact, the observed pattern is consistent with an average divergence of 8.42×10^{-7} phenotypic standard deviations per generation, or only 0.84σ over a million generations (fig. 1A). This rate is nearly six orders of magnitude slower than rates commonly observed in selection experiments (Lynch and Walsh 1999). This discrepancy shows that some sort of restraint acts on evolutionary divergence in the natural world (e.g., stabilizing selection, genetic or functional constraints). To explore the potential underlying causes of this result, we evaluated the ability of a series of quantitative genetic models to explain the pattern of divergence evident in the empirical data.

Neutrality

We find that a model of neutral evolution fails to account for the empirical data across all timescales. We evaluated divergence under this model on the raw scale (fig. 1A) because the expected average divergence is 0 for this model (fig. 2) and so is undefined on the log scale. Neutrality consistently fails for shorter time intervals by predicting too little divergence (fig. 6A). Conversely, neutrality consistently fails for longer timescales because the model pre-

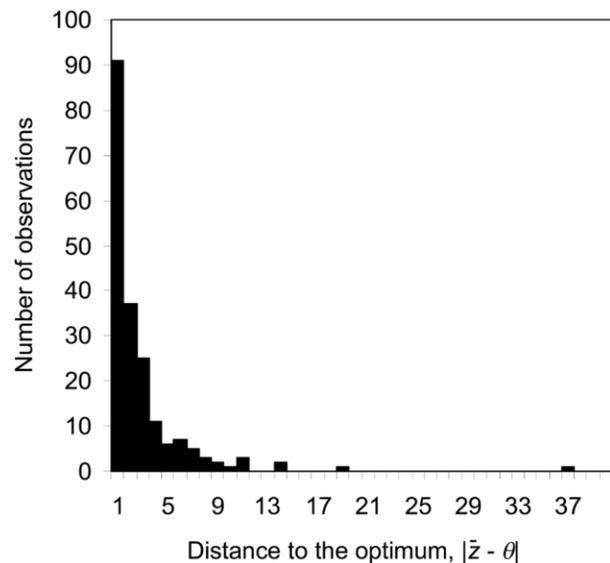


Figure 8: Distance of the phenotypic mean from the intermediate optimum, θ , in natural populations ($n = 197$). Values of $|\bar{z} - \theta|$, the absolute distance from the phenotypic mean to the optimum in units of phenotypic standard deviations, were calculated from the database of Kingsolver et al. (2001). The first bin shows the number of observations within 1 phenotypic standard deviation of the optimum, the second bin shows the number within 2 but ≥ 1 phenotypic standard deviation from the optimum, and so on. Two values >40 were dropped from the plot.

dicts far greater changes than those observed in the fossil record. The exact time intervals for failure depend on the combinations of values for h^2 and N_e employed in the model. As heritability increases, the confidence intervals about expected mean divergence quickly become very large. The opposite situation occurs with increasing N_e . As N_e increases, the model fails to predict the range of divergence values observed over relatively short intervals of time. While extremely low values for h^2 do serve as a constraint on evolution, substantial amounts of morphological divergence change can still occur over geological time—much greater than those observed in the fossil record.

Displaced Optimum

A model involving a displaced phenotypic optimum provides a good fit to Gingerich's (2001) data across all timescales for a range of parameter combinations. This model performs very well indeed when median parameter values from the literature are employed. For example, with small shifts in the optimum ($|\bar{z}_0 - \theta| < 6$) and strong stabilizing selection ($\omega^2 = 3.214$), predicted divergence almost exactly matches the observed mean change (figs. 6D, 1B). This model is most sensitive to variation in optimum displacement $|\bar{z}_0 - \theta|$ and, when employing central values for the other parameters, will provide a reasonable fit to the data only in the range $0.3 < |\bar{z}_0 - \theta| < 6$. The predicted dynamical pattern of divergence is almost always curvilinear such that if the model fails, it is because the model predicts too little divergence evolution on short timescales. When the optimum is displaced from the population mean much beyond 6σ (i.e., when $|\bar{z}_0 - \theta| > 6$), model failure due to overestimation of evolutionary divergence on nearly all timescales becomes common under a wide range of values for other parameters. Additionally, under this model, low values of heritability can act as effective constraints on evolutionary change. For any parameter combination, the model yields a poor fit to the data unless $h^2 \geq 0.01$. Finally, it is noteworthy that for any particular set of parameter values this model tends to explain only a modest portion of the total variance in observed rates of evolutionary divergence. (This limitation does not necessarily hold when N_e is very small.) However, reasonable parameter combinations can be found that sweep the curves for the expected mean divergence across the entire 99% data ellipse. Combinations that drive the expectation near to the upper boundary of the ellipse include those in which $|\bar{z}_0 - \theta|$ is "large" (around 6) and $h^2 > 0.1$. For the expected divergence to be near the lower boundary of the data ellipse, $|\bar{z}_0 - \theta|$ must be small (less than 0.1) regardless of the other parameter values employed.

Moving Optimum

In the case of a moving phenotypic optimum, the amount of predicted divergence is too little on short timescales, while that predicted for longer timescales far exceeds empirical observations for all reasonable parameter values (fig. 6E). The most favorable conditions are provided by central values for ω^2 and h^2 , providing that the rate at which the optimum moves is extremely slow ($k \approx 10^{-4} \sigma/\text{generation}$). Even under these most favorable conditions, however, the model fails to match empirical observations. The model is exceedingly sensitive to k , and unless the values employed for these parameters are minuscule, the model predicts evolutionary change that is orders of magnitude too large across all timescales. However, if k is much smaller than 10^{-4} , the model projects levels of divergence slower than observed values except over the longest time intervals. Conversely, this model is relatively insensitive to changes in N_e , h^2 , and σ_θ^2 , the amount of stochasticity in the movement of the optimum. In fact, k is the only parameter that affects expected values for divergence; the other parameters affect only the width of the confidence interval. Moreover, the function describing predicted divergence always maintains the same shape and slope under this model. Varying k simply shifts the predicted curve right (toward slower divergence with lower values of k) or left (toward faster divergence with higher values of k).

We also made additional calculations to describe the movement of the phenotypic optimum and the amount by which populations are expected to lag behind the optimum under particular parameter values for this model (see spreadsheet in the zip archive). For the parameter combinations most favorable for this model, the phenotypic mean closely tracks the optimum, exhibiting virtually no lag. The predicted location of the optimum (in terms of the expected amount of phenotypic divergence) therefore shows the same pattern as expected divergence, falling outside the bounds of empirical reality.

White Noise Motion of the Optimum

The special case of the preceding model in which populations experience a stationary optimum that undergoes Gaussian white noise motion ($k = 0$, but $\sigma_\theta^2 > 0$) was evaluated in the same manner as the neutral model. In other words, because the predicted mean divergence over time is 0, the model was evaluated on the raw scale. The white noise model generally predicts a range for divergence that is much smaller than the observed range across all timescales (fig. 6B). This model can provide a reasonable fit to the data with central values of h^2 , ω^2 , and N_e but only in combination with a degree of stochasticity in the position of the optimum ($\sigma_\theta^2 = 50\text{--}100$) that would result

in population extinction. This model differs profoundly from the one that follows in that the stochastic variation contributed by motion of the optimum is constant through time, whereas under Brownian motion, it increases linearly with time.

Brownian Motion of the Optimum

A model in which the optimum undergoes Brownian motion, with no long-term average change in position, suffers from the same problems as the neutral model (fig. 6C). The model is virtually insensitive to h^2 , ω^2 , and N_e ; only varying the value of σ_θ^2 has a material effect on its predictions. Consequently, whatever the values of h^2 , ω^2 , and N_e , the model predicts too little divergence on short timescales and too much divergence on long timescales throughout the realistic range of values for variance in movement of the optimum ($\sigma_\theta^2 = 10^{-5}$ – 10^{-1}). When σ_θ^2 takes on very small values (10^{-9} – 10^{-6}), the model predicts too little divergence over the timescale. No realistic values of heritability, effective population size, or strength of stabilizing selection can rescue the model from this predicament.

Peak Shift

A model in which the phenotypic mean shifts between adjacent optima yields a reasonably good fit to the data only under a highly restricted range of parameter combinations. With the central values for heritability and stabilizing selection ($h^2 = 0.4$, $\omega^2 = 3.214$), the model fits the data only if N_e is in the range 200–750 and the distance between the peaks is between 0.1 and 15σ (fig. 6F). Even then, the model tends to underestimate divergence on short timescales. Weaker stabilizing selection only makes the situation worse. The model is especially sensitive to the depth of the valley between the adjacent peaks, W_a/W_v . Notice that unless the valley is extremely shallow, such that $W_a/W_v \leq 1.01$, no divergence is predicted in most cases, and divergence thus becomes undefined on the \log_{10} scale (see supplementary spreadsheet). Otherwise, model failure is always a result of predicted divergence that is too slow on short timescales.

Discussion

In this article, we show that the paradox of morphological stasis can be resolved by applying models that incorporate elements of both directional and stabilizing selection. The failure of a neutral model, coupled with the success of models with stabilizing and directional selection, provides powerful evidence that selection is a crucial element in understanding broad patterns of morphological divergence (see details below). On the other hand, our results indicate

that genetic constraints play at most a minor role in the explanation of evolution patterns (i.e., fig. 1). We found that model performance was highly sensitive to variation in parameters of directional and stabilizing selection but remarkably insensitive to variation in genetic constraint. Indeed, in most models, heritability could be varied over its entire feasible range (0.001–0.999) without appreciable impact on predictions. This insensitivity is largely a consequence of the large timescale at the far end of the data range. Over a period of 10^4 – 10^7 generations, even a trait with miniscule heritability can show a considerable response to selection. Finally, we note that although genetic constraint does not appear to be a general explanation of stasis in the data set that we have analyzed, genetic variation may limit evolution in other contexts and for other kinds of characters. We now turn to the more detailed lessons than can be gleaned by considering each model in turn.

We found that a model of neutrality fails to account for both stasis and patterns of morphological divergence. Under the neutral model, huge evolutionary change can occur over geological time as a result of genetic drift in the phenotypic mean (Lynch and Hill 1986). Magnitudes of change as large as those predicted by drift over long time periods are simply not present in the data. Conversely, the neutral model predicts too little evolutionary change on shorter timescales (e.g., fig. 6A). One cannot compensate for these model failures by constraining genetic variation. Simulation of such a genetic constraint by using small values for h^2 results in predicted levels of divergence far beyond those observed over geological time. In agreement with other analyses, these results suggest that genetic constraints by themselves can account neither for stasis nor for the departures from stasis observed in the data set.

A model with a stationary optimum that undergoes Brownian motion faces the same problem as the neutral model. This model generally predicts too little evolution on short timescales and too much evolution on long timescales. The performance of a model with a stationary optimum that undergoes white noise motion is even worse than both of these models. In the white noise case, the model predicts too little evolution on all timescales over all realistic values of parameters.

The moving optimum model (Lynch and Lande 1995; Lande and Shannon 1996) formalizes Simpson's notion of phyletic evolution (see fig. 31 in Simpson 1944) by describing a situation in which a population (or higher taxonomic unit) is challenged with an adaptive peak that moves at a constant rate. This model predicts rates of evolution that are far too fast compared with those observed over geological time for most every parameter combination. We can thus refute a scenario of strict phyletic

gradualism in accounting for the observed patterns of divergence.

Effective population size is the Achilles' heel of peak shift models. The particular model that we explored (Lande 1985, 1986) could account for divergence on all timescales, using central values for inheritance and selection, only if N_e is in the range 200–750. Small population size is required for populations to escape from one adaptive peak and be drawn to another, distant peak. With $N_e > 200$, there are virtually no realistic sets of parameters that will allow this model to account for the divergence that is observed on both short and moderately long timescales. Lande (1985) addressed this limitation by proposing that the peak shift might be accomplished in one of n small, isolated populations—similar to Mayr's (1954) model of peripatric speciation. Here we face the quandary that such very small populations are unlikely to be represented in the fossil record. They are therefore unlikely to produce the divergence data that we aim to explain.

The displaced optimum model (Lande 1976) accounts for the divergence data over many parameter combinations and over all timescales. Amount of divergence is limited in this model by static and dynamic properties of the adaptive landscape. The phenotypic mean chases an optimum that has instantaneously moved to a new position and then resides near that new position for many generations. Contrary to the peak shift model, the displaced optimum model does not require evolution in small populations, nor does it demand that we invoke special metapopulation structures. Next, we discuss what the success of this model implies about the causes of long-term evolutionary stasis.

Persistent Configuration of the Adaptive Landscape

The explanation of stasis via selection requires not just stabilizing selection but also long-term constraints on the position of the adaptive peak. This requirement is apparent in our results, which resoundingly reject models that produce perpetual change (neutrality and moving optimum). Although the universality of environmental change makes persistence in peak position difficult to accept (Hansen and Houle 2004), many lineages experience stable biotic interactions for millions of generations. Boucot (1978, 1990) discusses many examples of terrestrial and marine communities that maintain stable composition over periods ranging from 1 to 300 million years. It is not unrealistic to assume that lineages situated in such communities experience long-term stability in the position of their adaptive peaks. Furthermore, stabilizing selection is not a simple external, environmental issue that is imposed purely from without. Stabilizing selection arises from the interaction between organisms and their environment—

both internal and external. Stabilizing selection can thus be thought of as resulting from external, normalizing forces (e.g., predation, competition) together with the internal, stabilizing constraints that abound at all levels of organization as a consequence of functional interactions.

Long-term persistence in the position and configuration of the adaptive landscape may be promoted by interactions among characters. Such interactions are a key ingredient in multivariate stabilizing selection (Berg 1959). The strength of the stabilizing effect of these interactions is captured in multivariate analyses that measure correlational selection (i.e., the off-diagonal elements in the γ - and ω -matrices; Lande and Arnold 1983). While there are still dismally few estimates of the strength of correlational selection (for examples, see Arnold 1988; Brodie 1992), there is considerable evidence for its frequent operation. For example, morphologists have long realized that character complexes such as the components of vertebrate dentition engender stability as a consequence of the strong functional interactions for which they are coselected. These complex phenotypic interactions help to build the genetic and phenotypic correlations that create phenotypic integration (Olson and Miller 1958; Lande 1980)—a manifestation of long-maintained stabilizing selection.

Constrained Optima and the Existence of Stable Adaptive Zones

The surprising success of the displaced optimum model over all timescales supports the notion of stable adaptive zones. In our implementations of this model, we displaced the optimum just once and asked, How closely does the phenotypic mean approach the new optimum over each time interval? For the range of displacements we employed (0.01 – 10σ), the mean rapidly evolved to the immediate vicinity of the new optimum. Surprisingly, such single displacements by a characteristic amount accounted for the divergence data over all timescales. Recurrent displacements of the optimum are likely in the natural world, so it is especially surprising that a single displacement fits the data so well. The success of this model may be attributed to our single displacements being surrogates for the net displacement of the optimum in evolving lineages. Thus, suppose the successful parameter value for displacement is 3σ . On short timescales, 3σ might be a surrogate for an actual, single displacement of this magnitude. On longer timescales, the actual optimum might move several or many times, but nevertheless if its total or net displacement is typically not more than 3σ , then the model will accurately predict the results of that process. Although we used a single displacement on longer timescales, the model can succeed because it correctly predicts phenotypic divergence in response to net displacement of the optimum. This view

of the model's success directs our attention to factors that could limit net displacement of the optimum to a value such as 3σ .

Simpson's (1944, 1953) concept of adaptive zones provides a useful framework for considering limits on the movement of the adaptive optimum. In Simpson's worldview, an adaptive zone is an ecological mode of life that imposes particular constraints on phenotype. Thus, the aerial insectivore adaptive zone necessitates flying as a condition of occupancy. This zone has supported independent invasions and radiations by birds and bats (Simpson 1944, p. 193). In his diagrams of adaptive zones, Simpson draws a band across a space in which morphology is plotted against time. Within this band, an evolving lineage tracks a moving optimum that sometimes bifurcates or otherwise diverges from the ancestor, but the optimum, and hence the phenotypic mean, remains within the band or adaptive zone. This idea is similar to Gingerich's "time form lattice" concept (Gingerich 2001; fig. 9)—the notion that reflecting boundaries exist on the amount of change possible, beyond which organisms will no longer be considered members of the same lineage. Returning to Simpson's imagery, the success of the displaced optimum model suggests that adaptive zones are typically less about 6 phenotypic standard deviations in width. What might account for this characteristic width?

All of the factors that promote stabilizing selection (e.g., interspecific competition, developmental constraints) may also contribute to the existence of narrow adaptive zones. In addition, the boundaries and hence the width of adaptive zones may be defined to a large extent by energy balances and intrinsic limits on design. These limitations arise from restrictions imposed by the strengths and intrinsic properties of materials, such as bone and cartilage. Such physical constraints are implicated in the trend for flying vertebrates to have much smaller body sizes than the largest of their flightless relatives.

A Data-Based Vision of the Adaptive Landscape

The adaptive landscape is arguably the most important integrative concept in evolutionary biology (Arnold et al. 2001). We estimated the curvature (H) and width of the adaptive landscape (a function of ω^2), as well as the distance of the phenotypic mean from the adaptive peak ($|\bar{z} - \theta|$), using a database of standardized linear (β) and quadratic (γ) selection gradients estimated from many contemporary populations (Kingsolver et al. 2001). These calculations suggest that phenotypic means are typically very close to the adaptive peak (46% are within 1 phenotypic standard deviation of the optimum, and 65% are within 2 standard deviations; figs. 8, 9; table 2). Such proximity reinforces the intuition of many naturalists that pop-

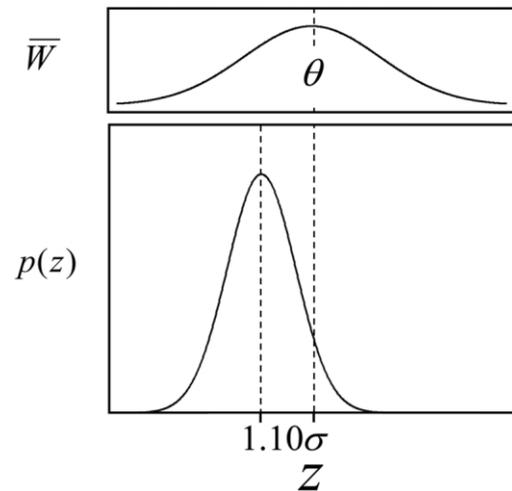


Figure 9: Vision of the adaptive landscape using estimates of parameters based on a recent summary of selection studies (table 2). In the top panel, a Gaussian adaptive landscape, \bar{W} as a function of \bar{z} , is shown in which $\omega^2 = 3$, a modal value for this selection parameter. The bottom panel shows a normally distributed trait, z , with unit variance and a mean situated 1.10 phenotypic standard deviations away from the optimum of the adaptive landscape, a median value for this displacement.

ulations are highly adapted to local conditions. Close proximity to the adaptive peak also implies that strong directional selection is unlikely to be sustained for long periods of time, in agreement with analysis of evolutionary rates in contemporary populations by Kinnison and Hendry (2001). The proximity of the mean to the adaptive peak may be even closer than our calculations indicate. Hereford et al. (2004) and Hersch and Phillips (2004) have suggested that, for a variety of reasons, the distribution of β reported by Kingsolver et al. (2001) is artificially shifted toward higher absolute values. Similarly, although the range of ω^2 is large, its frequency distribution is strongly leptokurtic with a modal value of 3.214 (fig. 7; table 2). Thus, the strength of stabilizing selection can often be strong in nature, stronger than is usually assumed in theoretical studies.

We need additional multivariate analyses of selection to make stronger inferences about the adaptive landscape. Because studies of nonlinear selection are overwhelmingly univariate (Kingsolver et al. 2001), we need to consider how this deficiency affects our vision of selection surfaces and adaptive landscapes (Phillips and Arnold 1989; Simms 1990; Blows and Brooks 2003). Univariate analyses are likely to overestimate the strength of stabilizing selection, because they do not account for indirect effects arising from stabilizing selection on correlated characters (Lande and Arnold 1983, eq. [15c]). If nonlinear selection on correlated characters is both stabilizing and disruptive, de-

pending on the character, the bias introduced by univariate analysis is more difficult to characterize. A recent survey suggests that this more complicated circumstance is likely. Blows and Brooks (2003) survey 19 data sets that included three or more characters and found that the selection surface was saddle shaped in a large majority of cases. Saddles have also been found in a number of other multivariate cases (Arnold 1988; Arnold and Bennett 1988; Simms 1990; Brodie 1992; Simms and Rausher 1993). This prevalence of saddles is consistent with the observation that γ appears to be symmetrically distributed about 0 (Kingsolver et al. 2001). Finally, Blows and Brooks (2003) find that the absolute value of nonlinear selection is stronger when it is measured along the eigenvectors of trait space rather than along the original trait axes. In this sense, nonlinear selection is probably underestimated in the majority of studies, which do not use canonical axes. In summary, although we know that univariate analyses can distort our vision of selection surfaces, we cannot say whether there is a positive or negative bias in the compilations of nonlinear selection coefficients reported in Kingsolver et al. (2001) and in this article (fig. 7). Our predominantly univariate view of the landscape will undoubtedly change with additional study of multivariate selection. From the multivariate studies conducted so far, it appears that selection surfaces, and perhaps adaptive landscapes, are often saddle shaped.

Implications

Comparative methods for inferring adaptation are often based on models for evolutionary process (e.g., Felsenstein 1985; Martins and Hansen 1996; Hansen 1997; Hansen and Orzack 2005). The present results suggest that the most commonly used process model, Brownian motion, is not the best predictor of evolutionary pattern. Felsenstein's (1985) independent contrasts method, for example, assumes Brownian motion-like evolution, typically modeled as random genetic drift or as Brownian motion of the optimum of a stabilizing selection function. We found that neither of these models produced a satisfactory fit to Gingerich's (2001) large database. The recurrent pattern of failure of these models is that too much divergence is predicted on moderate to long timescales. In particular, these models make the incorrect assumption that the variance among replicate lineages increases linearly with time. Felsenstein (1985) proposes that contrasts be inversely weighted by variances with this time-dependent feature. As a consequence, contrasts with long divergence times may be hugely devalued in assessing character correlations. Furthermore, contrasts are unlikely to be independent under other, non-Brownian motion models of stabilizing selection.

Phylogenetic comparative methods generally have focused on characters exhibiting evolutionary change while ignoring the possibility that ancestral characters retained for long periods are adaptations maintained by stabilizing selection (Hansen 1997; Butler and King 2004). Our results help to validate Hansen's (1997) model, which envisions organisms evolving in response to stabilizing selection imposed by stable adaptive optima and treats interspecific variation as arising from variation in adaptive optima. The important additional element that arises from our analysis is the necessity of invoking a limit or barrier to peak movement. Thus, our results suggest that uncertainty in the reconstruction of ancestral character states may be enormously inflated by using an inappropriate, neutral model of character evolution (Schluter et al. 1997) rather than one with a constrained optimum.

Inference of phylogenetic trees requires assumptions about evolutionary process (Felsenstein 2003). In phylogenetic analysis based on molecular data, incorporation of an explicit process model, with estimation of model parameters from the data, has long been the norm. No such consensus approach exists in the realm of phenotypic characters (Wiens 2000), perhaps because of a lack of consensus on a process model. Our results suggest that process models with constrained movement of an intermediate optimum may serve the role in inference of phylogeny from phenotypic data that Kimura's (1983) neutral model has served for molecular data. One such quantitative genetic model (the displaced optimum model) is capable of accounting for phenotypic divergence over a huge range of timescales but unrealistically assumes a single displacement of the adaptive peak. The next generation of process models for phenotypic traits should include stabilizing selection but explore the consequences of autocorrelation and various constraints on peak movement.

For inference of phylogeny, slowly evolving characters are often preferred over rapidly evolving characters because homoplasy is minimized. This choice is sometimes justified on the argument that rapidly evolving characters are adaptive, whereas invariant or slowly evolving characters are not. Our analysis suggests instead that invariance is a consequence of stabilizing selection and constraints on peak displacement. From this perspective, synapomorphies arise when an ancestor invades a new adaptive zone and a trait is thereafter maintained by stabilizing selection, with only modest displacements of the optimum. In other words, shared derived characters should attract our attention for the insights they may provide into evolutionary process, ecology, and history. They do more than define groups.

Conclusions

Our results suggest that properties of the adaptive landscape (topography, stabilizing selection, peak movement) account for the alternation between stasis and bursts of evolution that is commonly observed in the fossil record. In contrast, theories of punctuated equilibria (Gould 2002) emphasize divergence coincident with speciation and do not make predictions about the magnitude of divergence. Our results indicate that a model of an adaptive landscape with a single, displaced optimum can explain observed patterns of phenotypic divergence across timescales ranging from 1 to several million generations. When viewed in light of the evidence for rapid evolution in contemporary populations and the stasis that characterizes many fossil lineages, these results suggest that an intermediate optimum governs the tempo of evolution by virtue of constrained movement inside adaptive zones.

Acknowledgments

We thank P. Gingerich, A. Hendry, J. Kingsolver, and M. Kinnison for providing data; N. Eldredge for sharing an unpublished manuscript; B. Ajie, A. J. Boucot, R. Bürger, T. Hansen, A. Hendry, P. Hohenlohe, L. D. Houck, D. Houle, M. Kirkpatrick, R. Lande, J. Losos, and two anonymous reviewers for helpful discussion and comments; and A. Johnson and G. Johnson of Chatfield, Manitoba, for providing an inspirational work environment. This project was supported by National Science Foundation grants IR-CEB-0110666 to L. D. Houck, DEB-0447554 to S.J.A. and R. Bürger, and DBI-0208328 to S.E.

Literature Cited

- Arnold, S. J. 1988. Quantitative genetics and selection in natural populations: microevolution of vertebral numbers in the garter snake *Thamnophis elegans*. Pages 619–636 in B. S. Weir, E. J. Eisen, M. M. Goodman, and G. Namkoong, eds. Proceedings of the Second International Conference on Quantitative Genetics. Sinauer, Sunderland, MA.
- . 1992. Constraints on phenotypic evolution. *American Naturalist* 140(suppl.):S85–S107.
- Arnold, S. J., and A. F. Bennett. 1988. Behavioural variation in natural populations. V. Morphological correlates of locomotion in the garter snake (*Thamnophis radix*). *Biological Journal of the Linnean Society* 34:175–190.
- Arnold, S. J., M. E. Pfrender, and A. G. Jones. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica* 112/113:9–32.
- Bailey, N. T. J. 1964. The elements of stochastic processes with applications to the natural sciences. Wiley, New York.
- Barrowclough, G. F. 1980. Gene flow, effective population sizes, and genetic variance components in birds. *Evolution* 34:789–798.
- Barton, N. H., and B. Charlesworth. 1984. Genetic revolutions, founder effects, and speciation. *Annual Review of Ecology and Systematics* 15:133–164.
- Begon, M., C. B. Krimbas, and M. Loukas. 1980. The genetics of *Drosophila subobscura* populations. XV. Effective size of a natural population estimated by three independent methods. *Heredity* 45:335–350.
- Berg, R. L. 1959. The ecological significance of correlation pleiades. *Evolution* 14:171–180.
- Blows, M. W., and R. Brooks. 2003. Measuring nonlinear selection. *American Naturalist* 162:815–820.
- Blows, M. W., and A. A. Hoffmann. 2005. A reassessment of genetic limits to evolutionary change. *Ecology* 86:1371–1384.
- Bookstein, F. L. 1987. Random walk and the existence of evolutionary rates. *Paleobiology* 13:446–464.
- Boucot, A. J. 1978. Community evolution and rates of cladogenesis. *Evolutionary Biology* 11:545–655.
- . 1990. Conclusions. Pages 529–581 in A. J. Boucot, ed. *Evolutionary paleobiology of behavior and coevolution*. Elsevier, Amsterdam.
- Bradshaw, A. D. 1991. Genostasis and the limits to evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 333:289–305.
- Brock, J. P. 2000. The evolution of adaptive systems. Academic Press, San Diego, CA.
- Brodie, E. D., III. 1992. Correlational selection for color pattern and antipredator behavior in the garter snake *Thamnophis ordinoides*. *Evolution* 47:1284–1298.
- Bunce, M., M. Szulkin, H. R. L. Lerner, I. Barnes, B. Shapiro, A. Cooper, and R. N. Holdaway. 2005. Ancient DNA provides new insights into the evolutionary history of New Zealand's extinct giant eagle. *PLoS Biology* 3:44.
- Bürger, R. 2000. The mathematical theory of selection, recombination, and mutation. Wiley, Chichester.
- Bürger, R., and M. Lynch. 1995. Evolution and extinction in a changing environment: a quantitative-genetic analysis. *Evolution* 49:151–163.
- Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *American Naturalist* 164:683–695.
- Charlesworth, B., R. Lande, and M. Slatkin. 1982. A neo-Darwinian commentary on macroevolution. *Evolution* 36:474–498.
- Darwin, C. 1859. On the origin of species by means of natural selection. 6th ed. J. Murray, London.
- Eldredge, N., and S. J. Gould. 1972. Punctuated equilibria: an alternative to phyletic gradualism. Pages 82–115 in T. J. M. Schopf, ed. *Models in paleobiology*. Freeman, Cooper, San Francisco.
- Eldredge, N., J. N. Thompson, P. M. Brakefield, S. Gavrillets, D. Jablonski, J. B. C. Jackson, R. E. Lenski, B. S. Lieberman, M. A. McPeck, and W. Miller III. 2005. The dynamics of evolutionary stasis. *Paleobiology* 31:133–145.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *American Naturalist* 125:1–15.
- . 2003. *Inferring phylogenies*. Sinauer, Sunderland, MA.
- Frankham, R. 1995. Effective population size/adult population size ratios in wildlife: a review. *Genetical Research* 66:95–107.
- Gingerich, P. D. 1983. Rates of evolution: effects of time and temporal scaling. *Science* 222:159–161.
- . 2001. Rates of evolution on the time scale of the evolutionary process. *Genetica* 112/113:127–144.
- Gould, S. J. 2002. *The structure of evolutionary theory*. Belknap, Cambridge, MA.

- Haldane, J. B. S. 1949. Suggestions as to quantitative measurement of rates of evolution. *Evolution* 3:51–56.
- Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51:1341–1351.
- Hansen, T. F., and D. Houle. 2004. Evolvability, stabilizing selection, and the problem of stasis. Pages 130–150 in M. Pigliucci and K. Preston, eds. *Phenotypic integration*. Oxford University Press, Oxford.
- Hansen, T. F., and S. H. Orzack. 2005. Assessing current adaptation and phylogenetic inertia as explanations of trait evolution: the need for controlled comparisons. *Evolution* 59:2063–2072.
- Hendry, A. P., and M. T. Kinnison. 1999. The pace of modern life: measuring rates of contemporary microevolution. *Evolution* 53:1637–1653.
- Hereford, J., T. H. Hansen, and D. Houle. 2004. Comparing the strengths of directional selection: how strong is strong? *Evolution* 58:2133–2143.
- Hersch, E. I., and P. C. Phillips. 2004. Power and potential bias in field studies of natural selection. *Evolution* 58:479–485.
- Husband, B. C., and S. C. H. Barrett. 1992. Effective population size and genetic drift in tristylous *Eichhornia paniculata* (Pontederiaceae). *Evolution* 46:1875–1890.
- Jones, A. G., S. J. Arnold, and R. Bürger. 2003. Stability of the G-matrix in a population experiencing mutation, stabilizing selection, and genetic drift. *Evolution* 57:1747–1760.
- . 2004. Evolution and stability of the G-matrix on a landscape with a moving optimum. *Evolution* 58:1639–1654.
- Jorde, P. E., and N. Ryman. 1996. Demographic genetics of brown trout (*Salmo trutta*) and estimation of effective population size from temporal change of allele frequencies. *Genetics* 143:1369–1381.
- Karlin, S., and H. M. Taylor. 1981. *A second course in stochastic processes*. Academic Press, New York.
- Kimura, M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.
- Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. *American Naturalist* 157:245–261.
- Kinnison, M. T., and A. P. Hendry. 2001. The pace of modern life. II. From rates of contemporary microevolution to pattern and process. *Genetica* 112/113:145–164.
- Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30:314–334.
- . 1979. Quantitative genetic analysis of multivariate evolution, applied to brain : body size allometry. *Evolution* 33:402–416.
- . 1980. Genetic variation and phenotypic evolution during allopatric speciation. *American Naturalist* 116:463–479.
- . 1985. Expected time for random genetic drift of a population between stable phenotypic states. *Proceedings of the National Academy of Sciences of the USA* 82:7641–7645.
- . 1986. The dynamics of peak shifts and the pattern of morphological evolution. *Paleobiology* 12:343–354.
- Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–1226.
- Lande, R., and S. Shannon. 1996. The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution* 50:434–437.
- Lynch, M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. *American Naturalist* 136:727–741.
- Lynch, M., and W. G. Hill. 1986. Phenotypic evolution by neutral mutation. *Evolution* 40:915–935.
- Lynch, M., and R. Lande. 1995. Evolution and extinction in response to environmental change. Pages 234–250 in P. M. Kareiva, J. G. Kingsolver, and R. B. Huey, eds. *Biotic interactions and global change*. Sinauer, Sunderland, MA.
- Lynch, M., and B. Walsh. 1999. *Genetics and analysis of quantitative traits*. Sinauer, Sunderland, MA.
- Martins, E. P., and T. P. Hansen. 1996. A microevolutionary link between phylogenies and comparative data. Pages 273–288 in P. H. Harvey, A. J. Leigh Brown, J. Maynard Smith, and S. Nee, eds. *New uses for new phylogenies*. Oxford University Press, Oxford.
- Mayr, E. 1954. Change of genetic environment and evolution. Pages 157–180 in J. Huxley, A. C. Hardy, and E. B. Ford, eds. *Evolution as a process*. Macmillan, New York.
- . 2001. *What evolution is*. Weidenfeld & Nicholson, London.
- Mousseau, T. A., and D. A. Roff. 1987. Natural selection and the heritability of fitness components. *Heredity* 59:181–197.
- Olson, E. C., and R. L. Miller. 1958. *Morphological integration*. University of Chicago Press, Chicago.
- Partridge, L. 1978. Habitat selection. Pages 351–376 in J. R. Krebs and N. B. Davies, eds. *Behavioral ecology*. Blackwell Scientific, Oxford.
- Phillips, P. C., and S. J. Arnold. 1989. Visualizing multivariate selection. *Evolution* 43:1209–1222.
- Roopnarine, P. D. 2003. Analysis of rates of morphologic evolution. *Annual Review of Ecology and Systematics* 34:605–632.
- Schluter, D., T. D. Price, A. Ø. Mooers, and D. Ludwig. 1997. Likelihood of ancestor states in adaptive radiation. *Evolution* 51:1699–1711.
- Sheets, H. D., and C. E. Mitchell. 2001. Why the null matters: statistical tests, random walks and evolution. *Genetica* 112/113:105–125.
- Simms, E. L. 1990. Examining selection on the multivariate phenotype: plant resistance to herbivores. *Evolution* 44:1177–1188.
- Simms, E. L., and M. D. Rausher. 1993. Patterns of selection on phytophage resistance in *Ipomoea purpurea*. *Evolution* 47:970–976.
- Simpson, G. G. 1944. *Tempo and mode in evolution*. Columbia University Press, New York.
- . 1953. *The major features of evolution*. Columbia University Press, New York.
- Stanley, S. M., and X. Yang. 1987. Approximate evolutionary stasis for bivalve morphology over millions of years: a multivariate, multilinesage study. *Paleobiology* 13:113–139.
- Storz, J. F., H. R. Bhat, and T. H. Kunz. 2001. Genetic consequences of polygyny and social structure in an Indian fruit bat, *Cynopterus sphinx*. II. Variance in male mating success and effective population size. *Evolution* 55:1224–1232.
- Turner, T. F., J. P. Wares, and J. R. Gold. 2002. Genetic effective size is three orders of magnitude lower than adult census size in an abundant, estuarine dependent marine fish (*Sciaenops ocellatus*). *Genetics* 162:1329–1339.
- Wiens, J. J. 2000. *Phylogenetic analysis of morphological data*. Smithsonian Institution, Washington, DC.
- Wright, S. 1968. *Evolution and the genetics of populations*. Vol. 1. University of Chicago Press, Chicago.