Stabilizing Selection and the Comparative Analysis of Adaptation

Thomas F. Hansen

STABILIZING SELECTION AND THE COMPARATIVE ANALYSIS OF ADAPTATION

Thomas F. Hansen

Division of Zoology, Department of Biology, University of Oslo, P.O. Box 1050, Blindern, N-0316 Oslo, Norway
E-mail: thomasha@bio.uio.no

Abstract.—Comparative studies tend to differ from optimality and functionality studies in how they treat adaptation. While the comparative approach focuses on the origin and change of traits, optimality studies assume that adaptations are maintained at an optimum by stabilizing selection. This paper presents a model of adaptive evolution on a macroevolutionary time scale that includes the maintenance of traits at adaptive optima by stabilizing selection as the dominant evolutionary force. Interspecific variation is treated as variation in the position of adaptive optima. The model illustrates how phylogenetic constraints not only lead to correlations between phylogenetically related species, but also to imperfect adaptations. From this model, a statistical comparative method is derived that can be used to estimate the effect of a selective factor on adaptive optima in a way that would be consistent with an optimality study of adaptation to this factor. The method is illustrated with an analysis of dental evolution in fossil horses. The use of comparative methods to study evolutionary trends is also discussed.

Key words.—Adaptation, comparative method, macroevolution, optimality, phylogenetic constraint, stabilizing selection.

In evolutionary biology there are two major ways to test whether a trait is adapted to a particular function. The optimality approach investigates whether the trait is optimized with respect to the performance of the function, and the comparative approach looks across species to see whether the trait is predictably associated with demand for the function. Traditionally, the comparative approach has been the less rigorous of the two, but the last two decades have seen drastic improvements in methodology (reviewed in Harvey and Pagel 1991; Martins and Garland 1991; Miles and Dunham 1993; Losos and Miles 1994; Maddison 1994; Martins and Hansen 1996a). A new, modern comparative method has emerged. This method is characterized by the explicit inclusion of phylogenetic relationships that, among other things, are used to correct for similarities due to common ancestry that might otherwise obscure adaptive similarities due to convergent evolution. Attention also shifted from actual trait values to inferred historical changes in the trait and its environment (Ridley 1983). Association between changes in trait and environment is taken as evidence of adaptation in much the same way as the actual association between trait and environment was used before. This approach was motivated by cladistic methodology, which provided the means for reconstruction of historical changes (e.g., Brooks and McLennan 1991), but have also had strong influence on the comparative approach to adaptation in general.

The focus on evolutionary change in general, and the cla- distic methodology in particular, has unfortunately discouraged the study of adaptation in terms of maintenance by stabilizing selection. Cladists are adamant in allowing only derived traits as putative adaptations; characters maintained in their ancestral states need no further explanation (e.g., Coddington 1988, 1990; Baum and Larson 1991; Brooks and McLennan 1991). However, as pointed out by Frumhoff and Reeve (1994), the demonstration that a trait is ancestral does not rule out the possibility that it is maintained as an adaptation by stabilizing selection. The change and maintenance of character states are simply different evolutionary processes (e.g., Gould and Vrba 1982; Coddington 1988; Harvey and Pagel 1991; Williams 1992; Reeve and Sherman 1993). Both of these processes are worthy of study, and, as the study of adaptation in terms of function, design or optimality in extant populations is mainly concerned with the latter, there exists a need for comparative methodology focused around character maintenance. An aim of this paper is to provide an example of how this can be done on a basis that treats interspecific variation as variation in adaptive optima.

The study of adaptive maintenance of traits may also have been hampered by attempts to restrict the term adaptation to traits that have demonstrably changed as a consequence of selection for the adaptive function. Such history-laden definitions make maintenance by stabilizing selection, the very subject of an optimality study, irrelevant by definition (Reeve

1341

and Sherman 1993). Hence, I use the term adaptation in the broad sense of any trait maintained by selection for current function regardless of its historical origin (i.e., aptation in the terminology of Gould and Vrba 1982).

Many comparative studies are concerned with quantitative characters or characters that reflect underlying quantitative variation. For such characters, it may be stasis and not change that needs explanation (Williams 1992). Usually, quantitative characters show ample genetic variation (Houle 1992) and appreciable quantitative variation arises by mutation each generation (Lynch 1988). On this basis, evolutionary quantitative genetics has demonstrated that selection, and even genetic drift, has immense power to change characters on macroevolutionary time scales. Rates of phenotypic evolution are usually much smaller than expected from neutral drift-mutation models (Lynch 1990; Spicer 1993; Cheetham et al. 1993, 1994; Björklund 1994), and this despite the weakness of the tests employed (Savalli 1993). This suggests that character maintenance by stabilizing selection is a dominant mode of evolution. Optimality studies of adaptation are consistent with this view, as they treat the maintenance of a character at a local optimum as a core assumption (Mitchell and Valone 1990). The optimality assumption is used to test hypotheses about the constraints determining the position of the optimum (Mitchell and Valone 1990; Parker and Maynard Smith 1990; Sober 1993). A complementary comparative study of adaptation should compare adaptive optima across species assuming that each species is at an optimum. The influence of a selective factor on the adaptive optimum can then be investigated by identifying two or more selective regimes and comparing trait values in species influenced by different regimes.

In this paper, I first develop a model of evolutionary change that incorporates strong stabilizing selection around an adaptive optimum as the dominant evolutionary force. In this model phenotypic change is due to changes in the selective and genetic constraints that determines the position of the optimum. Thus, the model is compatible with both the assumptions made in an optimality study and what is known about macroevolutionary change in quantitative characters. In a previous paper, Hansen and Martins (1996) discussed how models of microevolutionary change can be turned into a phylogenetic pattern of statistical covariances between species. I use this approach to derive, from the evolutionary model, a statistical method for estimating the effect of a selective factor on the position of adaptive optima. Hence, if two or more selective regimes can be identified, the method can be used to assess the effect of changes in the selective factor. While the modeling is simple and certainly open for much improvement, it illustrates several neglected issues in the comparative analysis of adaptation.

ON THE TARGET OF ESTIMATION IN A COMPARATIVE ANALYSIS OF ADAPTATION

As in optimality studies, we shall make the fundamental assumption that the trait under investigation is in fact an adaptation in the sense that it is at any time kept close to a fitness optimum by selection. Thus, the question cannot be whether the trait is an adaptation in this absolute sense. Of course, this is a legitimate question, and the topic of studies that compare interspecific patterns with neutral models, but it is not the topic of this paper. The present concern is with a different kind of hypothesis, namely whether the trait shows adaptation to a specific function, or more precisely, whether direct selection caused by a specific functional demand influences the position of the trait optimum and thereby the trait itself.

Consider the hypothesis that the high-crowned (hypsodont) cheek teeth of some herbivorous mammals are an adaptation to withstand the high levels of tooth wear caused by a grazing diet (Simpson 1944; Fortelius 1985). This hypothesis may be tested by comparing crown heights across species with different amounts of grasses in their diet (Janis 1988). In doing such a study, we do not expect the dental adaptive optimum to be determined only by diet. Other background factors, such as metabolic needs, digestive capabilities, longevity, and the size and shape of the jaw will influence the position of the optimum. As many such factors may change slowly, they create constraints on the rate of adaptation to diet and, as they vary across species, they create interspecific variation that may confound the variation due to diet. A few known and measurable confounding variables may be included in the analysis, but, however careful the study, scores of unknown or unmeasurable factors will not be taken into account (Ridley 1988; Grafen 1989; Nee et al. 1996; Ridley and Grafen 1996). This leads to an important problem in the study of adaptation; how to correct for the confounding influence of such inevitable background factors. This difficulty is the backbone of recent critiques of the comparative method such as dePinna and Salles (1990), Lauder et al. (1993), Frumhoff and Reeve (1994), and Leroi et al. (1994). While the problem is undeniable, and the cited papers raise a number of valid points, it is not clear that comparative studies, as opposed to other ways of studying adaptation, are particularly vulnerable to confounding variables. It is even tempting to turn the situation on its head and view the comparative method as a device for reducing the influence of confounding variables. By studying a relationship over a number of different species, the hope is that effects of uncontrolled factors are randomized so that they average to zero (for similar arguments, see Doughty 1996; Nee et al. 1996).

Imagine that we start with an infinite number of populations identical to the common ancestor of the group under consideration and allow them to evolve independently but with the selective factor in a fixed state (e.g., they are all grazers). I shall refer to the average adaptive optimum eventually reached by these populations as the primary optimum with respect to this state of the selective factor. This primary optimum is a hypothetical entity different from the actual optimum realized in specific populations, but it is a better measure of adaptation, because it reflects the influence of the selective factor against a maximally randomized background.

A comparative analysis of primary optima must assume that the selective factor can produce a systematic effect that can be similarly interpreted across a variable background. It must also assume that the selective factor does not predict multiple adaptive peaks. Clearly, the primary optimum does not exclude confounding factors that functionally covary with the selective factor. Only the accidental effects of history

within the group are randomized and, unsurprisingly, the problem of inferring causation from correlation remains. With these qualifications, I shall show how estimation of primary optima is possible even in the presence of phylogenetic structure, but this requires information about phylogeny and the history of the selective regime.

A MODEL OF ADAPTATION ON A MACROEVOLUTIONARY TIME SCALE

The Evolutionary Approach to a Primary Optimum

Consider a functionally important, quantitative trait under Gaussian stabilizing selection. On a macroevolutionary time scale, the standard model from evolutionary quantitative genetics (Lande 1976, 1979) predicts an essentially instantaneous approach of the trait mean to its optimum. The optimum thus reached is a function of all selective demands on the trait, incorporating numerous functional trade-offs and conflicts. This actual optimum is likely different from the primary optimum with respect to any subset of the total selective regime. Let θ be the primary optimum of the trait with reference to the state of a particular selective factor of interest, such as diet, and assume, temporarily, that this remains constant. Evolutionary changes on a macroevolutionary time scale then mainly consist of changes in the position of the actual optimum. These changes are influenced by selection in the direction of the primary optimum induced by the selective factor, but also by numerous changes in the genetic and selective background that are unrelated to the selective factor. Following Grafen (1989), we may think of the background as a sum of many at least partially independent genetic and selective factors of individually small effects. Thus, the change in the actual optimum and therefore the trait mean, Y, is a sum of the effects of many small random changes in the background and a systematic effect caused by the selective factor. Hence, it seems appropriate to model this change, dY, on a short stretch of time, as a normally distributed random variable.

The expectation or deterministic part of this change represents the effect of direct selection due to the selective factor under consideration and is thus directed toward the primary optimum. In addition it is reasonable to expect the rate of evolution toward the primary optimum to increase with distance from it. This is partially because the strength of direct selection and benefits of releasing constraints may increase, and partially because unconstrained genetic variation that allows change toward the optimum is more likely to be available. This includes the higher likelihood of such variation being generated by mutation or environmental change as illustrated in Fisher's (1958, pp. 42–44) geometric model of adaptation. A simple mathematical representation of these considerations is to model the expected change as $E[dY] = -\alpha(Y - \theta)$. The parameter α controls the rate of adaptation to the primary optimum. We shall return to its interpretation in a worked example below.

The variance of dY reflects perturbations due to selective demands from unconsidered selective factors, perturbations due to genetic correlations with other traits, environmental fluctuations, and even genetic drift and random mutation. There are few reasons for these factors to vary systematically

with Y, so they are assumed to be constant, $Var[dY] = \sigma^2$. The parameter σ^2 measures the magnitude of the perturbations.

We now turn to the long-term behavior of this model. Let Y_i be the mean state in any species, i, and Y_a the mean state in an ancestor, a, of i. Let t_{ai} be the time separating i from a. On the assumption that subsequent perturbations in the background are independent on all but very short time scales, it follows that Y_i is normally distributed with expectation and variance conditional on Y_a given as

$$E[Y_i|Y_a] = [1 - Exp(-\alpha_{ai})]\theta + Exp[-\alpha t_{ai}]Y_a, \quad (1)$$
$$Var[Y_i|Y_a] = v[1 - Exp(-2\alpha t_{ai})] \quad (2)$$

where $v = \sigma^2/2\alpha$. Equation (1) shows that the descendant starts out similar to the ancestor and then becomes less and less similar with time at a rate determined by α. Finally, all influence of the ancestral state is lost and the expectation becomes θ, the primary optimum. Thus, equation (1) describes how the constraints of the ancestral state are lost and adaptation is gained. Equation (2) shows how the variance among species increases as they get separated from a common ancestor. Eventually, as ancestral constraints are lost, the interspecific variance reaches an equilibrium, v, between the perturbing and the adapting forces. If α is small, the equilibrium variance may become very large and the species may keep drifting apart (technically the process becomes nonstationary). If α is very large, species will adapt instantaneously to new conditions, and proportionally large perturbations, σ^2, are needed to explain interspecific variation.

The process described is an Ornstein-Uhlenbeck process (Gardiner 1985). The Ornstein-Uhlenbeck process has been suggested before as a basis for comparative studies (Felsenstein 1988; Garland et al. 1993; Martins 1994; Hansen and Martins 1996; Martins and Hansen 1996b), as it is well known as a model of evolution by genetic drift and stabilizing selection (Lande 1976). However, the interspecific variation and phylogenetic constraints seen in many quantitative traits can hardly be due to drift alone. Hence, I have suggested the above as a novel biological interpretation of the process where changes in selective, genetic, or environmental background factors are the source of interspecific variation.

Changes in the Selective Regime

We shall now allow the selective factor to vary in the phylogeny in the sense that the primary optimum of a species may change through its history. We begin by dividing the history of the species up into segments of time in which the optimum remains constant (continuous change can also be handled as outlined in the appendix). Let the state of the primary optimum in the kth segment be $\theta(k)$. In the appendix, it is shown that the expected phenotype of the species is a weighted sum of the positions the primary optimum has had throughout its history, and in place of equation (1) we get

$$E[Y_i|Y_a] = \Sigma_k[Exp(-\alpha t_e) - Exp(-\alpha t_b)]\theta(k) + Exp[-\alpha t_{ai}]Y_a, \quad (3)$$

where t_e is the time back to the end of the kth time segment and t_b the time back to the beginning of the kth segment. This equation shows that ancient values of θ are discounted

by a factor that decrease exponentially with distance from present. The rate at which the past is discounted is determined by α. A low rate of adaptation to new conditions makes ancient adaptations relatively more important.

Consider the special case where only two distinct selective regimes are present in the phylogeny. The only difference between these is that the primary optimum is θ_1 in one regime and θ_2 in the other. Hence, $\theta(k)$ takes the values θ_1 and θ_2 in alternating segments. By collecting together all segments with the same selective regime, equation (3) can be written as a weighted average of the two optima and the common ancestral state (the root, r, of the tree) as

$$E[Y_i|Y_r] = c_{i1}\theta_1 + c_{i2}\theta_2 + c_{i3}Y_r, \quad (4)$$

where c_{i1} is the sum of weights from segments where the optimum was θ_1, c_{i2} is the sum of weights from segments where the optimum was θ_2 and c_{i3}, the coefficient for the contribution of the root, is $Exp[-\alpha t_{ri}]$, where t_{ri} is the time separating i from the root. The model may be described as a regression on the (appropriately weighted) times the species has evolved under each selective regime. Note that the coefficients sum to one ($c_{i1} + c_{i2} + c_{i3} = 1$).

With the aim of estimating θ_1 and θ_2, we can now write a linear model. Let $Y = \{Y_1,..., Y_n\}'$ be a column vector of the mean phenotypes of the species on the tree. This vector can be written as a sum of the expectation given in equation(4) and a residual error

$$Y = C\theta + \epsilon. \quad (5)$$

The residual error, ϵ, is a $N(0, V)$ error vector with variance matrix V (see appendix). The elements of the variance matrix are given in the next section. In the usual notation of linear models, the mean is written as the product of a design matrix, C, and a parameter vector $\theta = \{\theta_1, \theta_2, Y_r\}'$. The design matrix is an n × 3 matrix with elements c_{ij}. This sets the stage for estimation of θ by generalized least squares or other techniques.

Phylogenetic Covariance

The next step is to describe the elements of the variance matrix. The ijth element of V is the covariance between the ith and the jth species. Based on Hansen and Martins (1996), the appendix shows that these are given by the formula

$$Cov[Y_i, Y_j] = Exp[-\alpha t_{ij}]Var[Y_a] = vExp[-\alpha t_{ij}][1 - Exp(-2\alpha t_{ra})] \quad (6)$$

where t_{ij} is the time separating the species i and j (i.e., phylogenetic distance) and t_{ra} is the time separating their most recent common ancestor, a, from the root, r, of the tree. This formula assumes that the two branches leading to the species i and j evolved independently after the split from their common ancestor. Thus, covariance due to convergent evolution and common environment is not included. Note that the phylogenetic covariance is independent of the optimum and of the actual state of the ancestor. We see that the phylogenetic covariance equals the variance of the common ancestor multiplied by an exponential decay with separation time. If the underlying process is stationary so that all ancestors have the same variance, v, then the correlation between any two spe-

cies is $Exp[-\alpha t_{ij}]$, which depends only on the rate of adaptation, α, and the time separating the species, t_{ij}. Note that α determines both the phylogenetic covariance in equation (6) and the discount of the past in equations (1) and (3).

STATISTICAL ANALYSIS OF THE MODEL

Assumptions and Requirements

A phylogeny with branch lengths is necessary to determine the design and variance matrices. The phylogeny is a major unquantified source of error, but this problem is shared with and can be dealt with as in most other comparative methods (Losos 1995; Martins 1996; Martins and Hansen 1996a). Thus, the estimators derived herein are conditional on a hypothesis about phylogeny. Furthermore, the selective regime must be overlaid on the phylogeny without error. This assumption is also shared with a large class of comparative methods based on reconstruction of character states on the phylogeny (Martins and Hansen 1996a). The selective regimes can be overlaid on the tree through direct historical (paleontological) information or through a reconstruction based on current states. The method will be most reliable in situations where the broad pattern of change is fairly well known, so that the majority of the uncertainty can be attributed to the evolution of the quantitative trait. This is reasonable if there are only one or a few changes in the selective regime or when defined taxa are to be compared. However, for evolutionary labile explanatory variables, reconstructions based on parsimony, for example, will have substantial random error and be biased toward less change. Random error is unavoidable, but calls for statistical analysis. In principle, uncertainty of explanatory variables can be included by treating them as random effects. Martins and Hansen (1997) discuss this point in more detail. In that paper we also indicate how ancestral states and their standard errors can be estimated for many models of evolution.

Estimation

If the design and variance matrices are known, then θ can be estimated by generalized least squares (GLS), which is a generalization of ordinary least squares that allows for correlated data (e.g., Grafen 1989). The GLS estimator of θ is

$$\hat{\theta} = (C'V^{-1}C)^{-1}C'V^{-1}Y, \quad (7a)$$

$$Var[\hat{\theta}] \approx (C'V^{-1}C)^{-1}, \quad (7b)$$

where C' is transpose of C. The GLS estimator is the best linear unbiased estimator and robust in the senses of not requiring normality and remaining unbiased even when the variance matrix is misspecified (Searle et al. 1992). The diagonal elements of the matrix $Var[\hat{\theta}]$ are the variances of the individual parameter estimates.

As usual in ANOVAs, there may be linear dependencies in the design matrix that preclude identification of all parameters. In such cases the model can be rewritten to estimate differences between parameters (for details, see, e.g., Dobson 1990). Interpretation based on the difference between selective regimes is in any case more reliable.

The dispersion parameter, $v = \sigma^2/2\alpha$, can be estimated based on the (generalized) residual sum of squares:

$$RSS = (Y - \hat{Y})'(V/v)^{-1}(Y - \hat{Y}), \qquad (8)$$

where $\hat{Y} = C\hat{\theta}$ is the vector of predictions from the model. An unbiased minimum variance estimator of v is then

$$\hat{v} = \frac{RSS}{n - p}, \qquad (9)$$

where n is the number of species and p the number of independent parameters (i.e., the rank of C). Under the assumption of normality this is also a restricted maximum-likelihood estimator (Searle et al. 1992).

The estimators above are conditional on the parameter α. Estimation of α is not straightforward, as it enters nonlinearly into both the variance and the design matrix. A simple way to deal with this is to perform the analysis on a range of different a priori reasonable values of α. Results are then given conditionally on α. One might, for example, conclude that a certain selective factor is important given that α is in a certain range. An estimate of α can subsequently be obtained by comparing likelihoods from this procedure. An assessment of the uncertainty in this estimate can be given as a support region (Edwards 1972). It would also be possible to fit all parameters jointly by maximum likelihood, but I recommend fitting θ and v conditionally on α, as this allows conventional (linear model) interpretation of these parameters and may be less sensitive to misspecification of the model.

Multistate, Multiple, and Continuous Selective Factors

The model can easily incorporate any number and type of explanatory variables. Let x be any qualitative or quantitative explanatory variable, as in a usual linear model. We may then write the optimum as $\theta = \Sigma_j x_j \beta_j$, where the βs are parameters that measure the influence of the explanatory variable (i.e., θ is modeled exactly as the mean in a standard linear model). By fitting this into equations (3), the appropriate weights of each β-parameter can be obtained. The influence of β_j on species i can thus be seen to be

$$c_{ij} = \sum_k x_{jk}[Exp(-\alpha t_e) - Exp(-\alpha t_b)] \qquad (10)$$

where x_{jk} is the value of x_j in segment k. By using these values of c in the design matrix a vector of β-parameters can be estimated as discussed above. This means that the selective factor can be allowed to have several states and that multiple selective factors and their interaction can be assessed or controlled for.

Model Checking

Conditional on α, model checking and hypothesis testing can be conducted exactly as in a standard linear model by use of generalized sums of squares, as in equation(8), in place of ordinary sums of squares. If the residuals are normally distributed, it can be shown that the RSS has a $\chi^2(n - p)$ distribution and, if SSF and SSR are the RSSs of a full and a reduced model, respectively, then it can be shown that

$$F = \frac{(SSR - SSF)/\Delta p}{SSF/(n - p)}, \qquad (11)$$

has a $F(\Delta p, n - p)$ distribution under the reduced model. Here, Δp is the difference in degrees of freedom between the full and reduced model and p the number of independent parameters in the full model. Hence, standard F-tests can be used to decide whether to include extra variables.

Just as ordinary sums of squares measure the amount of variation in a dataset, generalized sums of squares measure interspecific variation when species are phylogenetically correlated. Hence, SST, the RSS of the model that only includes a grand mean, can be used as an overall measure of interspecific variation. The amount of interspecific variation explained by a model can be defined as percent of reduction in residual sum of squares relative to SST, as $R^2 = (SST - RSS)/SST$.

The residuals, ϵ, are predicted to be independent of the fitted mean and normally distributed with variance matrix V. This can be checked by plotting the residuals against the mean and by examining a set of standardized residuals for departures from normality. A vector of independent residuals with unit variance can be obtained by multiplying ϵ with any square root, $V^{-1/2}$, of V^{-1} (a Cholesky decomposition of V^{-1} could be used). If gross departures from these assumptions are observed, this is an indication that the model is not appropriate. In such cases modifications of the explanatory variables or a standard transformation of the data to a more appropriate scale may solve the problem.

A critical assumption of the model is that the parameters v and α do not vary with the selective regime or the predicted value of the phenotype. Heteroscedacity or excessive clustering in the residuals may indicate that these assumptions are violated. On extensive data one might consider fitting a model where these parameters are allowed to be different in different selective regimes or vary with the predicted mean. Martins and Hansen (1997) discuss how to include the predicted mean in the variance matrix.

AN ILLUSTRATION: HYPSODONTY AS AN ADAPTATION TO GRAZING IN FOSSIL HORSES

The horse family (Equidae) has an excellent fossil record and has provided several classical examples of adaptive evolution (Simpson 1944, 1951, 1953; MacFadden 1992). One of these is the evolution of high-crowned (hypsodont) teeth in association with "the great transformation" from a browsing to a grazing diet (Simpson 1951). Grasses contain silica, which combined with the increased intake of dust and grit, makes a grazer more prone to tooth wear (Simpson 1944; Fortelius 1985; Janis and Fortelius 1988). It is thus reasonable to expect compensatory adaptations in tooth morphology. Increased crown height is one such candidate adaptation. Comparative investigations of the hypothesis that hypsodonty is an adaptation to grazing, or more precisely ground feeding, have been undertaken and support the hypothesis both among horses and more generally across ungulates (Simpson 1944; Fortelius 1985; Janis 1988). However, it has never been investigated by the use of a statistical comparative method that takes phylogeny into account. Here, I reanalyze this classical case as an illustration of the above method.

Incidentally, Simpson's (1944, 1953) view of macroevolution in general and horse evolution in particular has many

Myr

ExtantEquus

Cormohipparion goorisi
Nannippus minor
Neohipparion coloradense
Hipparion tehonense
Merychippus insignis
M. primus

Equus scotti
4
2.5
3.5 Onohippidium galushai
10
Dinohippus leidyanus
M. matthewi
8.5 Protohippus simus
H. affinis
4
4
Megahippus mckennai
5
Archaeohippus blackbergi
4
Hypohippus avus
2
3.5
0.5
1
0.5
0.5
2
Anchitherium clarenci
Parahippus cognatus
P. tyleri
Parahippus leonensis
Miohippus gemarose
3
2
5.5
7
1.5
9
1.5
Miohippus quartus
Miohippus agnectans
4
4

Mesohippus bairdii
5
3
1 Mesohippus barbouri

Epihippus

Myr

3.0

3.2
496kg
1.9
(321kg) 1.8
60kg
2.6 311kg
230kg
2.2
2.1
266kg
0.8
0.6
(105kg)
162kg
(403kg)
1.2
1.85
1.0
1.0
0.6
195kg
0.7
102kg
136kg
161kg
122kg)
209kg
(92kg)
(95kg)(88kg)0.7
0.4
0.3
77kg
44kg
0.5
0.8
0.6
87kg
132kg
111kg
(54kg)
(80kg)
(69kg)
0.4
0.6
54kg
(54kg)

0.5
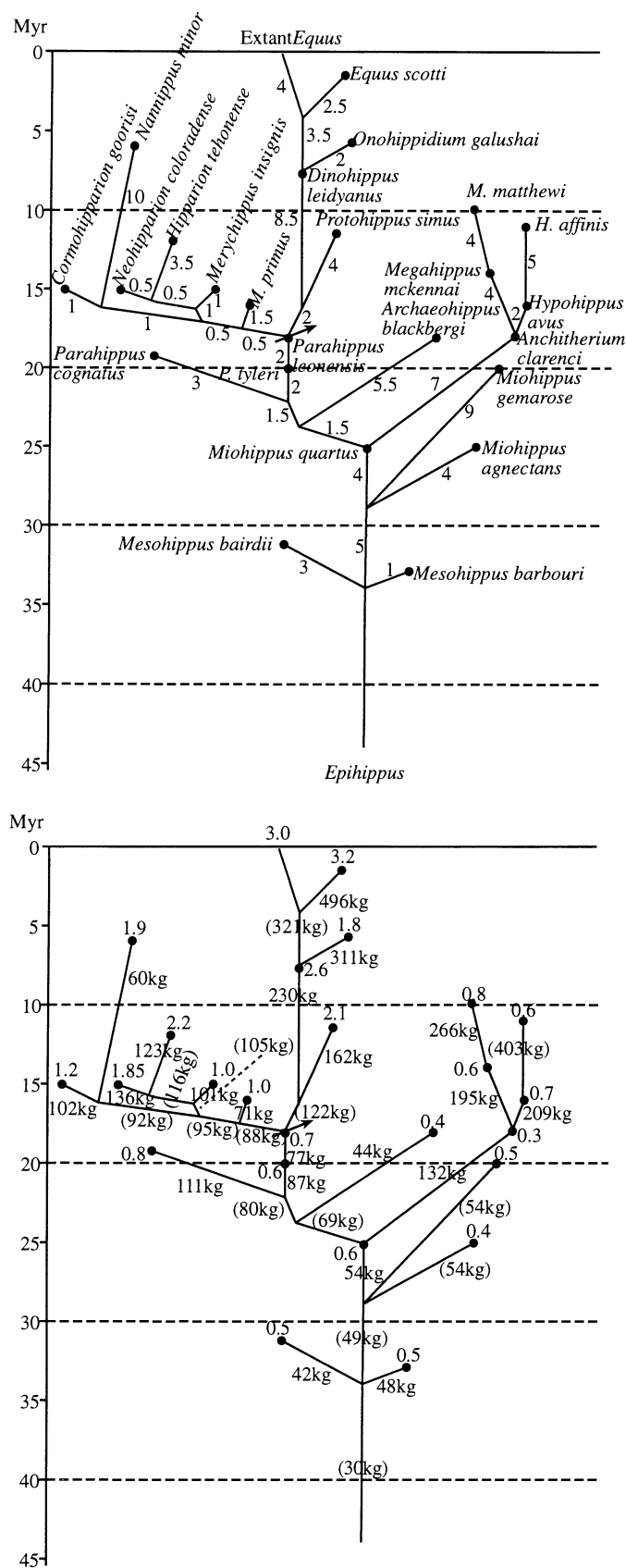(49kg)
0.5
42kg
48kg

(30kg)

FIG. 1. (Top) A conjectural phylogenetic history of horses starting from the genus *Epihippus*. The 24 horses used in this study are

similarities to the model presented in this paper. Simpson emphasized that species are at any time at (actual) adaptive optima, and that these could slowly and erratically change to improve the "primary adaptation" (evolve toward the primary optimum) as constraints from "secondary adaptations" (the background) are released (see e.g., Simpson 1944, pp. 86–93).

Data on crown heights are available for many fossil horses. As data, I use an index of hypsodonty, the ratio of maximum unworn crown height to maximum crown length of the first upper molar (M^1), originally from MacFadden (1992) and given in Figure 1. Included are 24 species descending from the genus *Mesohippus*. I compiled a phylogeny with branch lengths for these species as given in Figure 1. Species older than *Mesohippus* were excluded not to bias the sample in direction of ancestral browsing and as *Mesohippus* and descendants seem to have reached a stage of dental evolution characterized by homogeneity of the cheek teeth (Simpson 1951) making the M^1 a reliable indicator of the general dentition.

The great transformation seems to have started at about 18 M.Y.B.P. (Simpson 1951; MacFadden 1992; MacFadden and Cerling 1994). It is associated with an explosive radiation of species (MacFadden and Hulbert 1988) and various evolutionary changes that can be interpreted as adaptations to life on the dry, open plains that spread at the time. It might be an oversimplification to represent the shift as a precise, discrete event. Grasslands probably did not emerge instantaneously. Note though, that we do not model the shift in terms of a species suddenly adopting a full-fledged grazing habit. Dietary preferences are likely influenced by tooth morphology as well as other adaptations. The shift is best seen in terms of the emergence of a new primary optimum generated by the grassland niche. The actual optimum for a species at this time probably reflected a mixed diet, but the option of utilizing the newly emerged grassland created possibilities for new beneficial changes to happen. Or rather, this is the hypothesis we want to test by estimating the change in the primary optimum.

Formulating the Model for Evolution of Hypsodonty

To formulate the model, take the species *Protohippus simus* in Figure 1 as an example. Let Y_i be its observed index of

←

marked with a bullet. Numbers are branch lengths in million yr. The shift in selective regime from browsing to grazing is marked with an arrow. The phylogeny is compiled from information given by MacFadden (1992) in such a way that ages of species are consistent with the age given for the hypsodonty data; ages of species and genera are consistent with known ranges of occurrence; and the topology of the tree is as consistent as possible with information given in MacFadden (1992). In addition, phylogenetic distances given in MacFadden (1985, 1986, 1988) are used whenever compatible with the above. (Bottom) Data used in this study. The index of hypsodonty given for each horse is read off from figure 11.6 in MacFadden (1992). The estimated body size for each branch is based on MacFadden (1986). When available, a size estimate for the endpoint of the branch is used. When there is no estimate for the endpoint (numbers in parenthesis), a guess is made based on a weighted average of surrounding species or on data from related species not used in this study.

TABLE 1. Primary adaptive optima for hypsodonty as a function of diet for a range of α-values. The optimum is θ_1 for grazers and θ_2 for browsers. An α of 0.01 reflects very slow (constrained) evolution, while an α of 100 is effectively equivalent to instantaneous adaptation. The maximum-likelihood (ML) estimate of α is 0.247 and the two-unit support region is (0.107, 0.530). The difference between the optima is statistically significant at $P < 0.01$ for all α. The R^2 is a measure of the amount of interspecific variation explained by diet. Note that R^2 is weighted by phylogenetic covariance and should therefore not be used to compare the fit of models with different values of α.

α	$\hat{\theta}_1 - \hat{\theta}_2 \pm 2SE$	$\hat{\theta}_2 \pm 2SE$	Variation explained (R^2)	Support (log likelihood)
0.01	13.45 ± 8.14	0.73 ± 2.95	33%	-20.8
0.1	2.45 ± 0.92	0.60 ± 0.53	56%	-15.1
0.2	1.89 ± 0.53	0.58 ± 0.32	69%	-13.0
0.247 (ML)	1.78 ± 0.47	0.57 ± 0.28	72%	-12.8
0.3	1.70 ± 0.44	0.57 ± 0.26	74%	-13.0
0.4	1.60 ± 0.40	0.57 ± 0.24	75%	-13.7
0.5	1.54 ± 0.39	0.57 ± 0.24	74%	-14.6
1.0	1.38 ± 0.38	0.57 ± 0.24	70%	-17.4
100.0	1.31 ± 0.38	0.57 ± 0.25	68%	-18.6

hypsodonty (2.1 in this case). Through its history this species has evolved as a browser from the root until 18 M.Y.B.P. From then it evolved as a grazer until 12 M.Y.B.P., where the sample was taken. Hence, if we index grazer as 1 and browser as 2, c_{i1} is computed from the interval that starts 6 million yr before the species and c_{i2} is computed from the interval that starts at the root and ends 6 million yr before the sample was taken. Where to place the root is somewhat arbitrary. For simplicity, I place the root infinitely far back in the past. A full model that includes and estimates the root gives essentially the same results in this case. For illustration, let the value of α be 0.25 (in units of million yr^{-1}).

$$Y_i = c_{i1}\theta_1 + c_{i2}\theta_2 + \epsilon_i$$

$$= [\text{Exp}(-\alpha 0) - \text{Exp}(-\alpha 6)]\theta_1$$

$$+ [\text{Exp}(-\alpha 6) - \text{Exp}(-\alpha\infty)]\theta_2 + \epsilon_i$$

$$= [1 - \text{Exp}(-1.5)]\theta_1 + [\text{Exp}(-1.5) - 0]\theta_2 + \epsilon_i$$

$$= 0.78\theta_1 + 0.22\theta_2 + \epsilon_i, \tag{12}$$

where θ_1 is the optimum with respect to grazing and θ_2 the

optimum with respect to browsing. Thus, *P. simus* is judged to be 78% adapted to grazing and 22% constrained by the ancestral selective regime. Alternatively, we could say that *P. simus* is still 22% adapted as a browser.

In comparative studies there are often several explanatory variables or confounding variables to control for. To illustrate how this can be done I shall include body size as a covariate in the model. In general, it is a good idea to control for size due to its wide ranging effects on the morphological, physiological, and ecological background factors that determine the position of an actual optimum. In this case, body size also has relevance as an explanatory variable, as it has been suggested that hypsodonty is an adaptation to the increased tooth wear that might follow from the increased food intake or increased life span of a larger animal (Simpson 1944; Radinsky 1984). However, it has also been argued on functional reasons that relative tooth wear should be independent of both food intake and life span (Fortelius 1985; Janis and Fortelius 1988). I used inferred body weights of fossil horses to assign each branch in the phylogeny a size measure (Fig. 1b). If β_1 is the regression of hypsodonty on size and β_2 the regression on the interaction between diet and size, the full model becomes

$$Y_i = c_{i1}\theta_1 + c_{i2}\theta_2 + z_{ij}\beta_1 + z_{i2}\beta_2 + \epsilon_i, \tag{13}$$

where z_{i1} is the weighted average of body sizes throughout the history of the species computed as in equation (10) above. For *P. simus*, z_{i1} becomes 134 when α is 0.25. The interaction coefficient z_{i2} is computed similarly as the weighted average of the interactions between size and diet. The interaction on a single branch is computed as size in the grazing regime and as zero in the browsing regime. For *P. simus*, $z_{i2} = 120$ at $\alpha = 0.25$.

Results

Results of the analysis, not including body size, are presented for a range of different α-values in Table 1. For all values of a there is a strong effect of diet on the primary optimum both in terms of absolute difference and in terms of variation explained. On the maximum-likelihood estimate of α ($\hat{\alpha} = 0.247$, obtained by a grid search), diet explains 72% of the interspecific variation in hypsodonty. On this value of α, the estimated optimum with respect to grazing is

TABLE 2. Results including size in the model. The parameter β_1 is the regression of the browser optimum on size (in units 100 kg^{-1}), while $\beta_1 + \beta_2$ is the regression of the grazer optimum on size. Here, θ_1 and θ_2 are the intercepts of the optima at zero weight. The ML estimate of α is 0.553 and the two-unit support region is (0.241, ∞). The inclusion of size is statistically significant at $P < 0.05$ for values of α above 0.316.

α	$\hat{\theta}_1 - \hat{\theta}_2 \pm 2SE$	$\hat{\theta}_2 \pm 2SE$	$\hat{\beta}_1 \pm 2SE$	$\hat{\beta}_2 \pm 2SE$	Variation explained	Support
0.01	18.55 ± 15.22	0.47 ± 3.84	0.51 ± 4.57	-3.16 ± 8.34	35%	-20.4
0.1	2.31 ± 1.66	0.49 ± 0.78	0.11 ± 0.55	0.03 ± 0.87	57%	-14.9
0.247	1.32 ± 0.77	0.49 ± 0.39	0.08 ± 0.26	0.22 ± 0.40	77%	-10.4
0.4	1.05 ± 0.58	0.49 ± 0.31	0.06 ± 0.20	0.27 ± 0.30	83%	-8.8
0.5	0.96 ± 0.53	0.49 ± 0.29	0.06 ± 0.19	0.29 ± 0.27	84%	-8.6
0.553 (ML)	0.93 ± 0.52	0.50 ± 0.29	0.06 ± 0.18	0.30 ± 0.26	85%	-8.55
0.6	0.90 ± 0.51	0.50 ± 0.28	0.06 ± 0.18	0.30 ± 0.26	85%	-8.6
0.7	0.85 ± 0.49	0.50 ± 0.28	0.06 ± 0.17	0.31 ± 0.25	85%	-8.7
1.0	0.77 ± 0.47	0.50 ± 0.28	0.06 ± 0.17	0.32 ± 0.24	85%	-9.0
100.0	0.68 ± 0.46	0.50 ± 0.28	0.05 ± 0.17	0.33 ± 0.24	85%	-9.4

$\theta_1 = 2.35$ while that for browsing is $\theta_2 = 0.57$; this amounts to a profound difference in dental morphology. Thus, the hypothesis that hypsodonty is an adaptation to grazing is strongly supported.

The interpretation of the estimated α-values can be aided by relating α to the time it takes to move from one optimum to another. The time it takes for the expected phenotype, starting in an ancestral state and evolving in a new selective regime, to have moved halfway to the primary optimum is $t_{1/2} = \ln(2)/\alpha$. This half time is a measure of the time it takes before adaptation to the new selective regime is expected to be more influential than the constraints from the ancestral state. For $\alpha = 0.25$, the half-time is about 2.8 million yr and for the α-values in the two-unit support range, the half-time ranges from 1.3 to 6.5 million yr. These values mean that phylogenetic correlations and constraints are not negligible, as some species are within this range of each other or the ancestral browsing regime. If α is as small as 0.01, the half-time is 69 million yr and even extant *Equus*, which have evolved as grazers for 18 million yr, would be 84% constrained by the ancestral browsing condition. However, if α is about 0.25, as in our case, extant *Equus* are predicted to be about 99% adapted as grazers.

Body size does not have a very large effect (Table 2). The amount of variation explained is increased by a little over 10%. The best estimate is that the grazing optimum would increase from 1.61 to 2.98 as weight increases from 50 to 500 kg. For a browser it increases from 0.53 to 0.80. While these figures are not functionally negligible, size (or longevity) cannot be ascribed the major role in the evolution of hypsodonty. However, we may continue to entertain the idea that size has a moderate influence, at least for grazers. Perhaps the increased stress associated with grazing makes a species more vulnerable to additional demands on the dentition. There is however little basis for rejecting the alternative that it was the evolution of large size that was made possible by hypsodonty (also suggested by Simpson 1944, 1953), or simply that hypsodonty and large size are independent adaptations to life on open plains.

Phylogenetic correlations and constraints seem less important when size is taken into account ($\hat{\alpha} = 0.553$, $t_{1/2} = 1.25$ million yr). Even a model postulating instantaneous adaptation ($\alpha \to \infty$, $t_{1/2} \to 0$) has a reasonable fit to the data. Thus, while diet (or something closely associated with it) is clearly indicated as the major selective factor acting on hypsodonty, we may identify two alternative interpretations of the residual variation depending on whether we choose to explain the rather slow evolution of hypsodonty as a consequence of unspecified phylogenetic constraints or as an adjustment to gradually increasing size among horses living in a grassland habitat.

EVOLUTIONARY TRENDS

The models with $\alpha = 0.01$ in Tables 1 and 2 illustrate a situation where the rate of adaptation is very slow. In these models, the estimated grazing optimum lies far outside the range of observed data and its interpretation as an optimum is not biologically very meaningful. If such models show the best fit, evolution is better interpreted as a slow directional trend toward increased hypsodonty. We may formalize this as a limiting case of the primary optimum model by taking the limits $\alpha \to 0$ and $\alpha\theta \to \tau$, where τ is a constant. These limits mean that the rate of approach to the primary optimum goes to zero relative to the distance from it, so that it is not reachable. The parameter τ represents the trend in the sense of average change per time. Formally, the model reduces to a Brownian motion with a trend.

In this model, two selective regimes may impose different trends, τ_1 and τ_2 on the evolutionary process. Applying the limits to equation (3), it can be shown that the coefficient for the contribution of a segment equals the length of the segment. Hence, the coefficients c_{i1} and c_{i2} equal the total times the species has evolved in selective regimes 1 and 2, respectively. The coefficient for the root, c_{i3}, equals one. Thus, as in the optimum model, the species vector, **Y**, follows a linear model of the form $\mathbf{Y} = \mathbf{C}\tau + \epsilon$, where $\tau = \{\tau_1, \tau_2, Y_r\}'$, **C** is an $n \times 3$ design matrix with elements c_{ij} and ϵ is a $N(\mathbf{0}, \mathbf{V})$ residual vector. By taking the limit $\alpha \to 0$ in equation (6), the phylogenetic covariance becomes $\mathrm{Cov}[Y_i, Y_j] = \sigma^2 t_{ra}$, where t_{ra} is the time separating the most recent common ancestor from the root (as in Felsenstein 1985). The parameter σ^2 measures the strength of the perturbing forces as it did before. The statistical analysis of this "primary trend" model can be conducted exactly as discussed for the primary optimum model by using the appropriate design and variance matrices. Note though, that when all species are extant, linear dependencies in **C** will preclude estimation of some parameters. However, a contrast such as $\tau_1 - \tau_2$ can always be estimated.

DISCUSSION

This paper has presented a comparative methodology consistent with optimality studies of adaptation. For this to be possible, adaptation to a selective factor must be seen in terms of the effect this factor has on the adaptive optimum. The concept of a primary adaptive optimum was introduced to represent the effects of the selective factor against a maximally randomized background. The modeling illustrates how primary optima can be estimated in a relatively simple statistical framework provided two or more selective regimes can be identified on a phylogeny.

Existing comparative methods are not designed for studying adaptation in this sense. A main deficiency of all continuous character methods is that they do not take the selective history of the species into account. According to the model presented in this paper, the selective history has major influence on the estimates of adaptation whenever species are phylogenetically correlated. This arises from the fact that both phylogenetic correlations and the influence of past selective regimes ultimately derive from similarities between species and their ancestors and therefore disappear at the same underlying rate (α in the model). Although this is strictly a result derived from a specific model, it is likely to have general validity, as species living in different selective regimes can hardly be perfectly adapted to local conditions at the same time as they are correlated with each other. The implication of this observation is that whenever a comparative study needs to take phylogenetic correlations into ac-

count, it also needs to consider the effect of past selective regimes.

Modern comparative studies of adaptation have emphasized the origin and change of character states rather than their maintenance by stabilizing selection. Consider a hypothetical cladistic analysis of the hypsodonty data above. A such study might have employed the usual division of fossil horses into brachydont and hypsodont taxa. Hypsodont taxa having an index of hypsodonty larger than one and brachydont taxa having an index less than one. A parsimony reconstruction would indicate about two changes from brachydonty to hypsodonty on the phylogeny in Figure 1, both occurring in the grazing part of the phylogeny. This provides some support for a connection between grazing and hypsodonty, but anyone of the standard statistical methods that are used in such cases would have had difficulties rejecting a null hypothesis of random association. The low number of changes do not provide much statistical power. This may be contrasted with the extremely strong support found in the present study. This support ultimately derives from treating hypsodonty as what it is, a quantitative character, and from utilizing fully the abundant evidence for maintenance of hypsodonty among grazers and brachydonty among browsers.

While the focus on evolutionary change may have originated in cladistic methodology, it goes beyond this. For example, independent contrasts methods (Felsenstein 1985) are designed for estimating the correlation between evolutionary changes and must be substantially modified to study adaptation in the present sense. Several authors have also been influenced by Ridley's (1983) argument that a single adaptive radiation should provide only a single data point for the analysis. From this, it is argued that unresolved polytomies should be reduced to a single data point (Grafen 1989, 1992; Pagel 1992), or even that invariant parts of the phylogeny should be collapsed into a single node (Ridley and Grafen 1996). Such approaches may have some justification in the study of evolutionary change, but they throw out information about the maintenance of character states (see also Purvis and Garland 1993).

Many comparative methods have taken a phenomenological or statistical approach to the effects of phylogeny and are not necessarily tied to any particular evolutionary model. However, to apply these methods, it is necessary to choose specific ways of including phylogeny and history. This amounts to making specific microevolutionary assumptions and such methods are best used in connection with explicit evolutionary models. Some statistical approaches, particularly Grafen's (1989, 1992) "standard" regression, can provide a framework for the estimation of primary optima by including the appropriate design and variance matrices. For a categorical response variable, Pagel's (1994) method could be used to assess the influence of a selective history.

Phenomenological methods are sometimes seen as attempting to estimate a phylogenetic effect and then remove or correct for this when estimating adaptation (e.g., Stearns 1983; Cheverud et al. 1985; Grafen 1989, 1992; Gittleman and Kot 1990; Legendre et al. 1994). In light of the above model, this is problematical, as the phylogenetic and the adaptive effects are not fully separable (see also Lynch 1991). Although such methods may be interpreted as estimating the effect of present-day selective regimes, and are to an extent rescued by the positive correlations that usually exists between the history of the selective regime and its present state, they may yield suboptimal estimates of adaptation, as the historical component is ignored.

The necessity of knowing both the phylogeny and the history of the selective regime may seem as strict requirements for a comparative method. Frumhoff and Reeve (1994) have described many difficulties in accurately reconstructing a selective regime from observations of extant species (but see Maddison 1995). To this we may note that even the use of crude information is better than ignoring it altogether. There can be no doubt that the phylogeny and the placement of the great transformation in Figure 1a contain inaccuracies, but likewise, there can be no doubt that this representation is an order of magnitude more accurate than a nonphylogenetic labeling of all horses as either full-fledged grazers or full-fledged browsers. Even the admittedly very crude representation of the evolution of body size in Figure 1b can hardly be worse than the assumption that all species have remained at their final size throughout relevant history.

A comparative study must assume that the association between traits and environment can be taken as evidence for adaptation (Ridley 1983; Harvey and Pagel 1991). Recently, Leroi et al. (1994) have argued that interspecific correlation between trait and environment cannot be taken as evidence for adaptation, because a number of other mechanisms such as genetic drift, indirect selection, pleiotropy, genotype-by-environment interaction, and environmental variation can sometimes produce patterns indistinguishable from natural selection. DePinna and Salles (1990), Lauder et al. (1993), and Frumhoff and Reeve (1994) have made similar points. Although this critique is useful in directing attention toward interesting alternative hypotheses, it should not be seen as a discouragement for the comparative approach. Indeed, for any hypothesis it is possible to construct ad hoc alternatives that explain the data equally well. This does not mean that the data cannot be taken as evidence in favor of the adaptive hypothesis. It merely informs us that an hypothesis cannot be verified. The comparative method is able to produce evidence for hypotheses about adaptation, because it is always possible that the adaptive hypothesis can be falsified by the approach in the sense that the data may fail to support the predicted effect (Coddington 1990).

A comparative analysis should not be performed in isolation. It should be informed by an explicit hypothesis as to the functional basis of the adaptation that provides more detailed predictions of pattern and therefore makes the hypothesis more easily falsifiable. In this way it becomes much harder to find alternative nonadaptive hypotheses that predict similar patterns. Of course, one should not allow alternative hypotheses to be constructed in a purely ad hoc fashion; they must also be motivated by information external to the comparative data. If the analysis fails to distinguish between well-motivated hypotheses, adaptive or not, this is not a reason to turn away. Rather, it is a challenge to refine the hypotheses to make contrasting predictions that can be tested against data.

ACKNOWLEDGMENTS

LITERATURE CITED

BAUM, D. A., AND A. LARSON. 1991. Adaptation reviewed: a phylogenetic methodology for studying character macroevolution. Syst. Zool. 40:1–18.

BJORKLUND, M. 1994. Processes generating macroevolutionary patterns of morphological variation in birds: a simulation study. J. Evol. Biol. 7:727–742.

BROOKS, D. R., AND D. H. MCLENNAN. 1991. Phylogeny, ecology, and behavior: a research program in comparative biology. Univ. of Chicago Press, Chicago.

CHEETHAM, A. H., J. B. JACKSON, AND L.-A. C. HAYEK. 1993. Quantitative genetics of Bryozoan phenotypic evolution. I. Rate tests for random change versus selection in differentiation of living species. Evolution 47:1526–1538.

———. 1994. Quantitative genetics of Bryozoan phenotypic evolution. II. analysis of selection and random change in fossil species using reconstructed genetic parameters. Evolution 48:360–375.

CHEVERUD, J. M., M. M. DOW, AND W. LEUTENEGGER. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. Evolution 39:1335–1351.

CODDINGTON, J. A. 1988. Cladistic tests of adaptational hypotheses. Cladistics 4:3–22.

———. 1990. Bridges between evolutionary pattern and process. Cladistics 6:379–386.

DEPINNA, M. C. C., AND L. O. SALLES. 1990. Cladistic tests of adaptational hypotheses: a reply to Coddington. Cladistics 6:373–377.

DOBSON, A. J. 1990. An introduction to generalized linear models. Chapman & Hall, London.

DOUGHTY, P. 1996. Statistical analysis of natural experiments in evolutionary biology: comments on recent criticisms of the use of comparative methods to study adaptation. Am. Nat. 148:943–956.

EDWARDS, A. W. F. 1972. Likelihood. exp. ed. Johns Hopkins Univ. Press, Baltimore, MD.

FELSENSTEIN, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

———. 1988. Phylogenies and quantitative characters. Annu. Rev. Ecol. Syst. 19:445–471.

FISHER, R. A. 1958. The genetical theory of natural selection. 2d ed. Dover, New York.

FORTELIUS, M. 1985. Ungulate cheek teeth: developmental, functional and evolutionary interrelations. Acta Zool. Fenn. 180:1–76.

FRUMHOFF, P. C., AND H. K. REEVE. 1994. Using phylogenies to test hypotheses of adaptation: a critique of some current proposals. Evolution 48:172–180.

GARDINER, C. W. 1985. Handbook of stochastic methods. 2d ed. Springer-Verlag, Berlin, Germany.

GARLAND, T., JR., A. W. DICKERMAN, C. M. JANIS, AND J. A. JONES. 1993. Phylogenetic analysis of covariance by computer simulation. Syst. Biol. 42:265–292.

GITTLEMAN, J. L., AND M. KOT. 1990. Adaptation: statistics and a null model for estimating phylogenetic effects. Syst. Zool. 39:227–241.

GOULD, S. J., AND E. S. VRBA. 1982. Exaptation—a missing term in the science of form. Paleobiology 8:4–15.

GRAFEN, A. 1989. The phylogenetic regression. Phil. Trans. R. Soc. Lond. B 326:119–157.

———. 1992. The uniqueness of the phylogenetic regression. J. Theor. Biol. 156:405–423.

HANSEN, T. F., AND E. P. MARTINS. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50:1404–1417.

HARVEY, P. H., AND M. D. PAGEL. 1991. The comparative method in evolutionary biology. Oxford Univ. Press, Oxford.

HOULE, D. 1992. Comparing evolvability and variability of quantitative Traits. Genetics 130:195–204.

JANIS, C. M. 1988. An estimation of tooth volume and hypsodonty indices in ungulate mammals, and the correlation of these factors with dietary preference. Pp. 367–387 in D. E. Russell, J.-P. Santero, and D. Sigogneau-Russell, eds. Teeth revisited: proceedings of the VIIth international symposium on dental morphology, Paris 1986. Memoirs de Musee national d'historie naturelle, no. 53.

JANIS, C. M. AND M. FORTELIUS. 1988. On the means whereby mammals achieve increased functional durability of their dentitions, with special reference to limiting factors. Biol. Rev. 63:197–230.

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.

LAUDER, G. V., A. M. LEROI, AND M. R. ROSE. 1993. Adaptations and history. Trends Ecol. Evol. 8:294–297.

LEGENDRE, P., F.-J. LAPOINTE, AND P. CASGRAIN. 1994. Modeling brain evolution from behavior: a permutational regression approach. Evolution 48:1487–1499.

LEROI, A. M., M. R. ROSE, AND G. V. LAUDER. 1994. What does the comparative method reveal about adaptation? Am. Nat. 143:381–402.

LOSOS, J. B. 1995. An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. Syst. Biol. 43:117–123.

LOSOS, J. B., AND D. B. MILES. 1994. Adaptation, constraint, and the comparative method: phylogenetic issues and methods. Pp. 60–98 in P. C. Wainwright and S. M. Reilly, eds. Ecological morphology: integrative organismal biology. Univ. of Chicago Press, Chicago.

LYNCH, M. 1988. The rate of polygenic mutation. Genet. Res. Cambridge 51:137–148.

———. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. Am. Nat. 136:727–741.

———. 1991. Methods for the analysis of comparative data in evolutionary biology. Evolution 45:1065–1080.

MACFADDEN, B. J. 1985. Patterns of phylogeny and rates of evolution in fossil horses: hipparions from the Miocene and Pliocene of North America. Paleobiology 11:245–257.

———. 1986. Fossil horses from "Eohippus" (Hyracotherium) to Equus: scaling, Cope's law, and the evolution of body size. Paleobiology 12:355–369.

———. 1988. Fossil horses from "Eohippus" (Hyracotherium) to Equus, 2: rates of dental evolution revisited. Biol. J. Linn. Soc. 35:37–48.

———. 1992. Fossil horses: systematics, paleobiology and evolution of the family Equidae. Cambridge Univ. Press, New York.

MACFADDEN, B. J., AND T. E. CERLING. 1994. Fossil horses, carbon isotopes and global change. Trends Ecol. Evol. 9:481–486.

MACFADDEN, B. J., AND R. C. J. HULBERT. 1988. Explosive speciation at the base of the adaptive radiation of Miocene grazing horses. Nature 336:466–468.

MADDISON, D. R. 1994. Phylogenetic methods for inferring the evolutionary history and processes of change in discretely valued characters. Annu. Rev. Entomol. 39:267–292.

MADDISON, W. P. 1995. Calculating the probability distribution of ancestral states reconstructed by parsimony on phylogenetic trees. Syst. Biol. 44:457–473.

MARTINS, E. P. 1994. Estimating rates of character change from comparative data. Am. Nat. 144:193–209.

———. 1996. Conducting phylogenetic comparative analyses when the phylogeny is not known. Evolution 50:12–22.

MARTINS, E. P., AND T. GARLAND JR. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. Evolution 45:534–557.

MARTINS, E. P., AND T. F. HANSEN. 1996a. Phylogenetic comparative methods: the statistical analysis of interspecific data. Pp. 22–75 in E. P. Martins, ed. Phylogenies and the comparative method in animal behavior. Oxford Univ, Press, Oxford.

———. 1996b. A microevolutionary link between phylogenies and comparative data. Pp. 273–288 in P. Harvey, J. Maynard Smith, and A. Leigh-Brown, eds. New uses for new phylogenies. Oxford Univ. Press, Oxford.

———. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149:646–667.

MILES, D. B., AND A. E. DUNHAM. 1993. Historical perspectives in ecology and evolutionary biology: the use of phylogenetic comparative analysis. Annu. Rev. Ecol. Syst. 24:587–619.

MITCHELL, W. A. AND T. J. VALONE. 1990. The optimization research program: studying adaptations by their function. Q. Rev. Biol. 65:43–52.

NEE, S., A. F. READ, AND P. H. HARVEY. 1996. Why phylogenies are necessary for comparative analysis. Pp. 399–411 in E. P. Martins, ed. Phylogenies and the comparative method in animal behavior. Oxford Univ. Press, Oxford.

PAGEL, M. D. 1992. A method for the analysis of comparative data. J. Theor. Biol. 156:431–442.

———. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc. R. Soc. Lond. B Biol. Sci. 255:37–45.

PARKER, G. A., AND J. MAYNARD SMITH. 1990. Optimality theory in evolutionary biology. Nature 348:27–33.

PURVIS, A., AND T. GARLAND JR. 1993. Polytomies in comparative analyses of continuous characters. Syst. Biol. 42:569–576.

RADINSKY, L. 1984. Ontogeny and phylogeny in horse skull evolution. Evolution 38:1–15.

REEVE, H. K., AND P. W. SHERMAN. 1993. Adaptation and the goals of evolutionary research. Q. Rev. Biol. 68:1–32.

RIDLEY, M. 1983. The explanation of organic diversity: the comparative method and adaptations for mating. Clarendon Press, Oxford.

———. 1988. Why not to use species in comparative tests. J. Theor. biol. 136:361–364.

RIDLEY, M., AND A. GRAFEN. 1996. How to study discrete comparative methods. Pp. 76–103 in E. P. Martins, ed. Phylogenies and the comparative method in animal behavior. Oxford Univ. Press, New York.

SAVALLI, U. M. 1993. An application of the neutral model to the evolution of tail length in the genus Euplectes (Aves, Ploceidae). Evolution 47:696–699.

SEARLE, S. R., G. CASELLA, AND C. E. MCCULLOCH. 1992. Variance components. Wiley, New York.

SIMPSON, G. G. 1944. Tempo and mode in evolution. Columbia Univ. Press, New York.

———. 1951. Horses: the story of the horse family in the modern world and through sixty million years of history. Oxford Univ. Press, New York.

———. 1953. The major features of evolution. Columbia Univ. Press, New York.

SOBER, E. 1993. Philosophy of biology. Westvies Press, Boulder, CO.

SPICER, G. R. 1993. Morphological evolution of the Drosophila virilis species group as assessed by rate tests for natural selection on quantitative characters. Evolution 47:1240–1254

STEARNS, S. C. 1983. The influence of size and phylogeny on patterns of covariation among life-history traits in mammals. Oikos 41:73–187.

WILLIAMS, G. C. 1992. Natural selection: domains, levels, and challenges. Oxford Univ. Press, New York.

Corresponding Editor: J. B. Walsh

APPENDIX
Analysis of the Adaptation Model

The model outlined in the main text can be represented as the stochastic differential equation

$$dY_i = -\alpha[Y - \theta(t)]\,dt + \sigma dW, \quad Y(0) = Y_a, \qquad (A1)$$

where dW are uncorrelated $N(0, dt)$ random variables, $\{\theta(t)\}$ is a realization of the primary optima through time, and the state of an ancestor, Y_a, is used as initial value. Conditional on this realization of the optima, the expectation of a species can be computed by solving the differential equation $dE[Y_i] = -\alpha(E[Y_i] - \theta(t))dt$, with initial condition $E[Y_i(0)] = Y_a$ (Gardiner 1985). The solution is

$$E[Y_i \mid Y_a] = \alpha \int_0^{t_{ai}} \theta(s)\mathrm{Exp}[-\alpha(t_{ai} - s)]\,ds + \mathrm{Exp}[-\alpha t_{ai}]Y_a, \quad (A2)$$

where t_{ai} is the time separating the species from the ancestor. Equation (3) in the main text is obtained by integrating equation (A2) in the case where $\theta(t)$ shifts between discrete values.

The second moment of Y_i is, conditional on $\{\theta(t)\}$, given as the solution to the differential equation $dE[Y^2] = \{-2\alpha[E(Y^2i) - \theta(t)E(Y_i)] + \sigma^2\}dt$, with $E[Y^2_i(0)] = Y^2_a$ (Gardiner 1985). By using the solution for $E[Y_i]$ given in equation (A2) we obtain

$$\mathrm{Var}[Y_i \mid Y_a] = \frac{\sigma^2}{2\alpha}[1 - \mathrm{Exp}(-2\alpha t_{ai})]. \qquad (A3)$$

Note that this is independent of $\{\theta(t)\}$.

Assuming that species evolve independently after they split, the covariances between species can be computed from the formula $\mathrm{Cov}[Y_i, Y_j] = \mathrm{Cov}[E[Y_i Y_a], E[Y_j Y_a]]$ (Hansen and Martins 1996). Plugging in from equation (A2) we obtain

$$\mathrm{Cov}[Y_i, Y_j] = \mathrm{Exp}[-\alpha t_{ij}]\mathrm{Var}[Y_a], \qquad (A4)$$

where t_{ij} is the time separating the species and $\mathrm{Var}[Y_a]$ is the variance of their most recent common ancestor, which can be obtained from equation (A3).

Hansen and Martins (1996) showed that a set of species evolving according to a linear stochastic differential equation as (A1) on a phylogenetic tree have a joint multivariate normal distribution. Hence, the moments computed in equations A2–A4 constitute a complete specification of their distribution.