Chapter 14
Use and Misuse of Comparative Methods in the Study of Adaptation

Thomas F. Hansen

Abstract The comparative method can be used to test hypotheses of adaptation by comparing groups of species that meet different adaptive challenges. This requires attention to phylogenetic correlations and to historical lags in achieving adaptation. The modern phylogenetic comparative method has provided some partial solutions to these problems, but the field has also suffered from a systemic lack of demand for biological justifications of its statistical procedures. Consequently, assumptions have been made for statistical convenience and are often inconsistent with the relevant biology. I argue that common comparative tests of adaptation, including Brownian-motion based phylogenetic linear models and inferred-changes methods based on reconstructing ancestral states, violate essential characteristics of adaptation as a biological process. I discuss the requirements for biologically consistent comparative analysis of adaptation, and I review work toward this goal.

14.1 Introduction

Ever since Darwin the comparative method has been a major tool for studying adaptation on macroevolutionary timescales. By comparing species living in different niches or environments, we can look for systematic differences in biological traits and relate these to functional needs in the environment. Consider deer antlers. The males of most species in the deer family (Cervidae) sport antlers that are used in sexual displays. These are cast after the mating season and regrown each year. This pattern suggests an influence of sexual selection. We can construct a comparative test for the involvement of sexual selection by comparing deer

T. F. Hansen (✉)
Department of Biology, CEES, University of Oslo, Blindern,
P.O. Box 1066, 0316 Oslo, Norway
e-mail: Thomas.Hansen@bio.uio.no

species likely to experience different strengths of sexual selection. This can be based on ecological data on the severity of competition among males for females. For example, one could compare the size of antlers in species with monogynous versus polygynous mating systems. In a now classical paper, Clutton-Brock et al. (1980) did this by comparing the size of antlers across species with different-sized breeding groups, and they indeed found larger antlers in species with more competitive mating systems.

There are, however, many difficulties with such analyses. Perhaps the most fundamental of these relates to the fact that we are working with observational data and not with controlled experiments. This means that there are usually many possible interpretations of any one pattern, and many unknown or at least uncontrolled factors that can influence the result. Hence, it is crucial that comparative tests of adaptation are not done in isolation from other biological information. Studies have to start from specific hypotheses with a priori biological motivation, and the truism that hypotheses can only be falsified and not confirmed is even more to the point in observational than it is in experimental studies. In the case of antlers, we have a priori reasons to evoke sexual selection, but there are many different sexual-selection hypotheses to choose among, and even if some form of sexual selection is operating, it is almost certainly not the only relevant selective force. The growth of antlers requires considerable resources, and different diets may put different costs on antler growth (Geist 1998). For example, Geist argued that one factor that allowed the gigantic antlers of the Irish elk was a mineral-rich diet of willow (see also Gould 1998). Antlers may also be a hindrance of movement and the costs of this may depend on the habitat (open or closed), the predation pressure, the shape of the antler, and the size of the animal. Antlers are also integrated with the rest of the organism. Strong static and ontogenetic allometries exist within species, and the positive evolutionary (among-species) allometry (Fig. 14.1) that culminates in the relatively huge antlers in large-bodied species such as the Irish elk could be a non-adaptive side effect of body-size evolution (Gould 1973, 1974, 1977). Sexual selection may also act on alternative traits, and because of the positive allometry, large antlers in deer with competitive mating systems could be a non-adaptive side effect of selection on sexual size dimorphism. The mating system correlates with body size, and Clutton-Brock et al. (1980) even suggested that the positive evolutionary allometry of antler size may be a result of adaptive evolution of larger male bodies in more polygynous mating systems.

Adaptation is not instantaneous, and the ancestry of the species must be considered when evaluating its antler traits. For example, although the Irish elk had relatively huge antlers that may have deviated positively from the evolutionary allometry, so do the antlers of its closest living relatives, the fallow deers (Fig. 14.1). Lister et al. (2005) and Hughes et al. (2006) obtained ancient DNA from the Irish elk and showed that it indeed was related to fallow deers in the genus *Dama* (Fig. 14.2). Hence, the antlers of the Irish elk may reflect ancestral constraints and cannot automatically be assumed to be fully adapted to the "current" environment of the species.
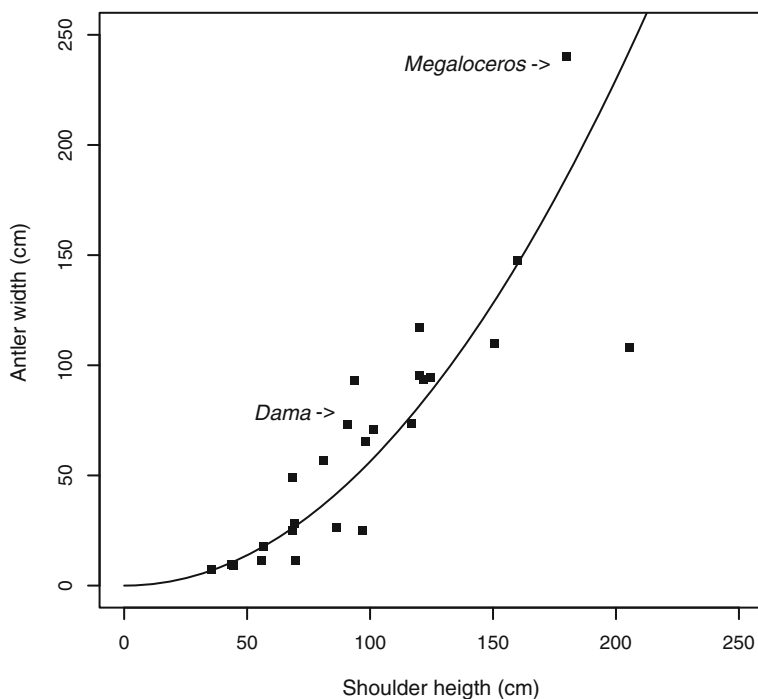
Fig. 14.1 Allometry of antler size against shoulder height in deer based on data from Clutton-Brock et al. (1980) with the Irish elk added (data from Gould 1974). Allometric curve estimated from non-phylogenetic least-squares regression on log-scale. The Irish elk (*Megaloceros*) and its closest living relative, the fallow deer (*Dama*), both deviate positively from the allometric relationship

Such ancestral constraints generate phylogenetic covariance (=phylogenetic correlation) between species and may thus violate the independence assumptions of standard statistical methods. This problem came in focus in the 1970s and contributed to a negative view of species comparison as non-rigorous methods that could only suggest, but not test, hypotheses. The emergence of so-called phylogenetic comparative methods that dealt with this statistical problem was therefore an important development in evolutionary biology. The foundational paper of Felsenstein (1985) presenting the method of independent contrasts showed how statistically correct analyses could be conducted based on a phylogeny with branch lengths and an assumed model of evolution. Together with rapid growth in the quality and availability of molecular phylogenetic information, this led to a renaissance of large-scale comparative studies and in the last decade, we have seen proliferation of ever more sophisticated methods for comparative analysis (see Martins 2000; Garland et al. 2005; Cooper et al. 2010; Freckleton et al. 2011; Nunn 2011; Stone et al. 2011; O'Meara 2012; Pennell and Harmon 2013; and this volume for some twenty-first century reviews).
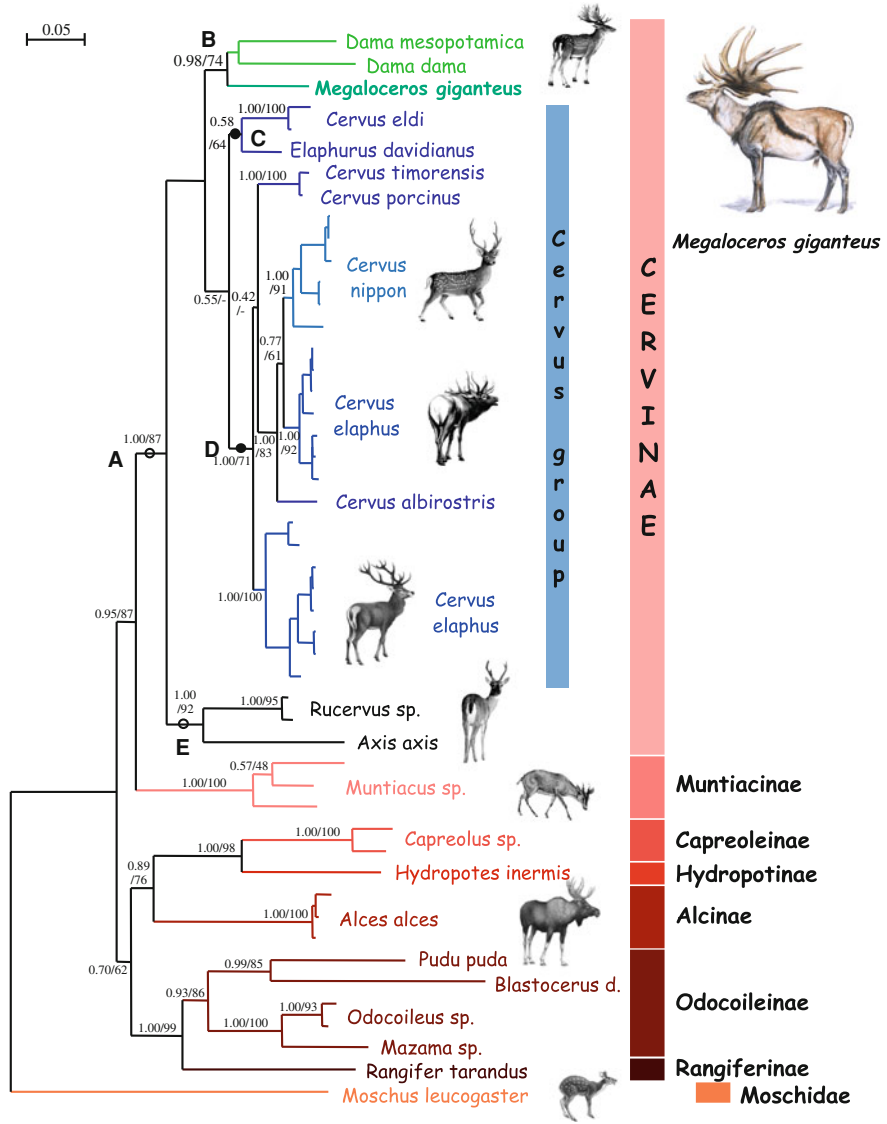
Fig. 14.2 Molecular phylogeny for the deer family including the Irish elk based on ancient DNA. Reprinted from Hughes et al. (2006) with permission from Elsevier

Developers of phylogenetic comparative methods have been concerned with solving statistical problems and have sometimes lost sight of the biological contexts in which the methods are to be used. Nowhere is this more evident than in the study of adaptation. For a long time, phylogenetic comparative methods for quantitative traits were built almost exclusively around a single model of evolutionary change, the Brownian-motion model, which introduces serious interpretational problems

when it is used to represent adaptive evolution. It is plainly inconsistent with adaptation toward an optimal state in given niche (Hansen and Orzack 2005). Starting with Hansen (1997) and Butler and King (2004), alternative methods that were more appropriate for testing adaptation were proposed, but have only recently been in regular use. Over the last few years, these methods have seen rapid development. Here, I review some of these developments and discuss the logic of and problems with the comparative study of adaptation.

14.2 The Concept of Adaptation

To better understand how the comparative method can be informative about adaptation, it is useful to clarify how I think about the concept and to outline some of the different notions of adaptation that have played a role in the comparative methods literature.

While it is possible to talk loosely about adaptation of a whole organism to a niche, as in saying that the polar bear is adapted to the arctic, quantitative studies must focus on specific biological traits. Adaptations are also for something specific. A trait is an adaptation for some task, X, if it helps perform X. The polar bear's white fur is an adaptation for camouflage against the arctic snow if it does help make the bear less visible and this improves the bear's fitness. Note that we may speak of adaptation in this sense even if the relevant task disappears. The white fur is an adaptation for camouflage against the snow even if the snow disappears and the species come to exist in a snow-free environment, say the zoo.

While all (genetic) adaptation is due to selection, not all selection leads to adaptation. A trait may be under selection for a number of reasons, but only selection causally generated by the success in performing task X will build adaptation for task X. This observation is related to Sober's (1984) distinction between selection *for* a trait and selection *of* a trait. A trait Y may be selected because it is correlated with some other trait under selection. This is called indirect selection, and while it is a potent mechanism for evolution, it is no more a mechanism for generating adaptations in Y than non-selective factors such as genetic drift or gene flow. Furthermore, a trait may be selected for more than one reason, and selection of trait Y for task Z will not generate adaptation for task X except by circumstance.

It seems reasonable to expect most traits to be influenced by several sources of both direct and indirect selection. The antlers of a deer may be under direct selection for impressing females, but this must be seen against a background of indirect selection acting on body size and mineral metabolism, and it will have to compete and interact with direct selection originating in male–male conflict and predation.

For this reason, I think adaptation is best understood in the context of a balance of forces. In fields such as behavioral ecology and life-history theory, investigations often start with the assumption that a trait is optimized by selection, and

adaptation for task *X* is studied by asking how variations in the need for *X* may shift the position of the optimum (Michell and Valone 1990; Reeve and Sherman 2001). This assumes that there are constrains from other selective forces. By saying that the shape of the antlers of a deer is adapted to attract females, we mean that female preferences have been able to shift the optimal antler shape away from the value it would have had if the females did not care. Note that this makes adaptation into a question of degree. The antlers may not be, and are indeed not expected to be, optimal for attracting females. It is enough that the female preference has some influence on the fitness function of the antlers.

In Hansen (1997), I proposed to understand the comparative study of adaptation in a similar way. As is done in within-species optimality studies of adaptation, we can start by assuming that the traits are optimal and then use the comparative method to compare the position of the optima in different environments. A comparative study of deer with monogynous versus polygynous mating systems can reveal whether antlers tend to evolve toward different states in the two situations. If they do, this is evidence that the antler optima are influenced by mating system or some variables correlated with mating system.

This view of the comparative study of adaptation is consistent with the most common ways of studying adaptation within species, but it differs dramatically from how adaptation is usually conceived in phylogenetic or historical studies. For example, there is a cladistic tradition for doing comparative studies of adaptation based on identifying trait changes and associating these with states or changes in the environment (e.g., Ridley 1983; Coddington 1988; Baum and Larson 1991; Brooks and McLennan 1991; Maddison 1994; Larson and Losos 1996). In this approach, only apomorphisms are candidate adaptations. This view is also reflected in Gould and Vrba's (1982) distinction between adaptation and exaptation, where the term adaptation is restricted to traits that have demonstrably originated due to the adaptive function (also favored by Sober 1984). In contrast, the broad non-historical concepts of adaptation favored by many, and in practice used by all, neontologists (Reeve and Sherman 1993) differ by focusing on whether traits are maintained by selection for the "adaptive" (including exaptive) function regardless of how the trait originated.

In comparative studies, these different notions of adaptation may lead to dramatically different approaches and conclusions. While the historical approach is limited to identifiable changes (or even apomorphisms), the non-historical approach will also utilize trait maintenance, absence of change, as evidence of adaptation (e.g., Williams 1992; Frumhoff and Reeve 1994; Westoby et al. 1995; Hansen 1997; Price 1997; Reeve and Sherman 2001). To exaggerate, a cladist may hold that the relatively huge antlers of the Irish elk cannot be adaptations to anything in the elk's environment, because the trait is ancestral (i.e., shared with the fallow deers). The counter argument is that the large-sized antlers must be adaptations for something because they have been maintained for long periods of time in balance with obvious selection pressures to reduce them. In Hansen (1997), I argued that we have comparative evidence for hypsodonty, high-crowned cheek teeth, being an adaptation to grazing in horses, because the trait has been

maintained in all grazing horses and it has not evolved in browsing horses. A cladist may instead say we have little evidence as a parsimony analysis reveals only a single shift to hypsodonty within this group and this is not sufficient statistical evidence to draw any conclusions.

Ironically, Gould did not heed his own rallying cry of "stasis is data" when discussing adaptation.

14.3   The Logical Structure of the Comparative Study of Adaptation

Sober (2008) points out that comparative studies involve a shift in explanandum from why a trait has a certain value to why there is a difference between groups in this value. That is, we shift from trying to explain why the antlers of Irish elk have an average span of 3 m to explain why polygynous and monogynous deer have different antler sizes. I want to argue, however, that this shift in explanandum does not need to entail a shift in explanans. In both cases, the explanation can be based on the assumption that the trait is at or near an optimum and the explanatory goal is to identify the factors that determine the positions or differences of optima.

Hansen and Bartoszek (2012) formalized this as follows. Let Y be the state of the trait we are studying, say antler size, and let us assume that this trait is at or close to a fitness optimum determined as a function of a number of factors, X, as $Y = f(X_1, \ldots, X_m)$. Assuming we know the function, f, and the exact state of all relevant factors, we could make a perfect prediction of the position of the optimum. In practice, however, we can only know, or at least measure, one or a few relevant factors, and we do not know the exact functional relationship between these and the optimal state. It is still possible to test hypotheses of adaptation. Let us say we want to test whether antler size is influenced by sexual selection, and we have observations of the strength of sexual selection from several species. This could come from direct measures of variance in mating success or from data on mating systems and some theory linking mating system to strength of sexual selection (e.g., arguing that sexual selection on males is stronger in polygynous mating systems). Let X_1 be our measure of strength of sexual selection, and rewrite the model as

$$Y = b_0 + b_1X_1 + r(X_1, \ldots, X_m), \tag{14.1}$$

where $r(X_1, \ldots, X_m) = f(X_1, \ldots, X_m) - (b_0 + b_1X_1)$ are biologically determined residual deviations from the model. Obviously, these residuals will be different in different species due to different values of the X-variables. It is still possible to test the influence of our focal variable X_1 by estimating the coefficient b_1. This will give the linear influence of X_1 on the optimum. For example, if X_1 is taken to be an indicator of polygynous versus monogynous mating systems, b_1 will be the average difference in optimal antler size for the two mating systems. I want again to underscore that this does not assume that there is one unique optimum for each

mating system. Each species has its own optimum, $f(X_1, \ldots, X_m)$, determined by many factors. The comparative method works by identifying systematic effects of some focal variables above the background noise due to changes in other unmeasured factors and then testing whether these effects are consistent with theoretical predictions. If not, the theory should be revised.

Many issues with comparative methods revolve around the model residuals, which may violate any of the standard linear model assumptions. Because the residuals depend on biological factors that may be shared among related species, they cannot be assumed to be independent. The residuals may also depend on the focal factor or on variables that are related with the focal factor. Given that some factors may also be discrete or of major influence, the residuals are not necessarily normally distributed. While these problems are serious, they can all be diagnosed and at least partially dealt with. The focus of the modern phylogenetic comparative method has been to solve the non-independence problem. We will now look at how this is done.

14.4 Phylogenetic Comparative Methods

Most comparative analyses are standard regression or ANOVA types of analyses with species trait means as the dependent variable. In this setting, the phylogenetic comparative method is a standard linear model modified to account for phylogenetic correlations in the residuals. Formally, the model can be written as

$$\mathbf{y} = \mathbf{D}\beta + \mathbf{r} \tag{14.2}$$

$$\mathbf{r} \sim N(0, \mathbf{V}) \tag{14.2a}$$

where $\mathbf{y}$ is a vector of species observations, $\mathbf{D}$ is a design matrix with predictor variables, $\boldsymbol{\beta}$ is a vector of parameters to be estimated, and $\mathbf{r}$ is a vector of residuals assumed to be normally distributed with mean zero and, a not necessarily diagonal, variance matrix, $\mathbf{V}$. If the $\mathbf{D}$ and $\mathbf{V}$ are specified, generalized least squares (GLS) estimates can be used to obtain unbiased minimum-variance estimates of the parameter vector $\boldsymbol{\beta}$, and if $\mathbf{D}$ and $\mathbf{V}$ depend on additional parameters with unknown values, as they typically do in adaptation models, then so-called estimated GLS, where the additional parameter values are estimated by maximum likelihood, can be used (e.g., Martins and Hansen 1997; Chaps. 5 and 6).

While this is straightforward in principle and can also be extended to generalized linear models such as logistic regression (Martins and Hansen 1997; Hadfield and Nakagawa 2010; Ives and Garland 2010; Chap. 9), the crux is how to model the $\mathbf{D}$ and $\mathbf{V}$ matrices. The standard approach, which forms the basis of nearly all phylogenetic comparative methods, is to leave the design matrix fixed and model the covariances in $\mathbf{V}$ as proportional to shared branch lengths on a phylogeny. This is justified if the response and predictor variables jointly follow a multivariate Brownian-motion process (Felsenstein 1985). In this case, regressions

of variables on each other are linear and residuals are normally distributed with covariances proportional to shared branch lengths. This is the basis for the method of independent contrasts, which may be regarded as an algorithm for implementing GLS when residual covariances are proportional to shared branch lengths.

It is essential to realize, however, that other evolutionary processes lead to different patterns. Hansen and Martins (1996) give an overview of the phylogenetic correlation patterns expected under different evolutionary processes, and many do not provide a simple proportionality of covariance to branch lengths (see Chap. 15). Furthermore, it is also typical that model residuals have patterns of phylogenetic signal that are different from the patterns in the data themselves (Labra et al. 2009).

14.5 Problems with the Standard Phylogenetic Comparative Method

Confusing phylogenetic signal in the data (e.g., the traits) with phylogenetic signal in model residuals is perhaps the most common error in the application of modern phylogenetic comparative methods (Fig. 14.3). The core assumptions of independence, normality, and homoscedasticity that are made in the standard regression and ANOVA models apply to the residuals from the model and not to the response or predictor variables per se. Still, it is entirely common to see tests for phylogenetic signal being conducted on the data and used to justify the use, or not, of phylogenetic corrections. This is a fundamental error and has no doubt lead to many misapplications of phylogenetic corrections (see Hansen and Orzack 2005; Labra et al. 2009; Revell 2010; Hansen and Bartoszek 2012 for discussions of the problem).

Figure 14.3a illustrates a common pattern in cross-species regression analyses. This is essentially the standard textbook illustration of the need for phylogenetic corrections, but this pattern is in fact totally consistent with the assumptions of standard non-phylogenetic regression. Even if there is a strong phylogenetic signal in both the response (y-axis) and the predictor variable (x-axis), there is no indication of a phylogenetic pattern in the residual deviations. This situation arises when adaptation is rapid. Then, we do not expect a phylogenetic signal in model residuals, but if related species tend to occur in similar environments (i.e., having similar values of their predictor variables), then we still expect a phylogenetic signal in the response variable. Correcting for phylogeny in this situation is throwing the baby out with the bathwater.

My impression from the relatively few studies that have reported phylogenetic signal in model residuals is that this situation is more common than situations with strong phylogenetic signal in model residuals as illustrated in Fig. 14.3b. If so, then the application of phylogenetic comparative methods has done more harm than good in the study of adaptation. Standard non-phylogenetic methods would usually have been a better choice than methods based on independent contrasts and the like.
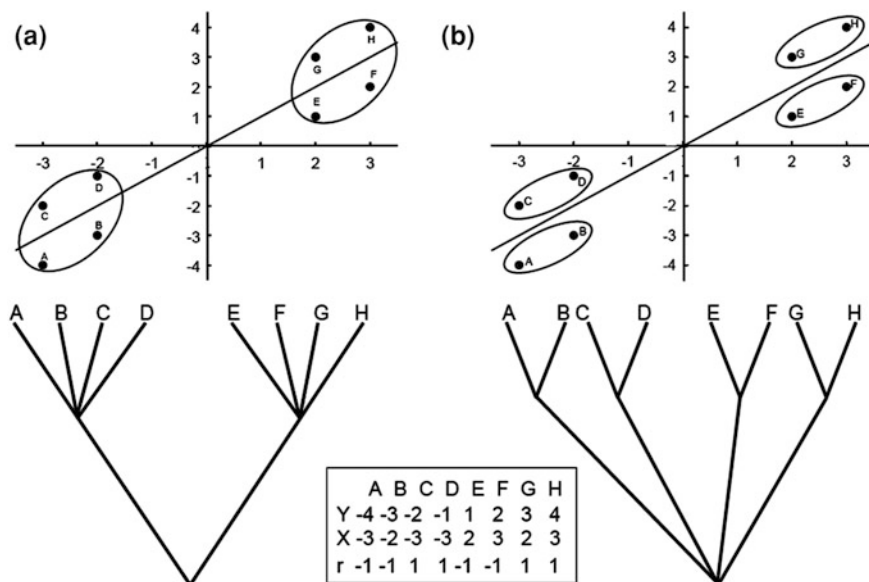
Fig. 14.3 Sources of phylogenetic effects. The trait Y on the y-axes is adapting to an environment X on the x-axes, but species $A–H$ deviate from the regression line. In **a**, the species are divided into two clades as shown in the phylogeny, and even if both Y and X are associated with the phylogeny (all species in a clade have the same sign of Y and X values), the residual deviations, r, from the line are not associated with the phylogeny (species within a clade are deviating in different directions). In **b**, the species within a clade are deviating in similar directions, and a phylogenetic GLS would be appropriate. In **a**, the phylogenetic effect in the trait is generated by the association of the trait with a phylogenetically-structured environment, while in **b**, it is also influenced by phylogenetically correlated maladaptations. Developed in collaboration with A. Labra and modified from Labra et al. (2009)

Furthermore, this problem is but one instantiation of a suite of problems deriving from a statistically convenient, but biologically unjustified, separation of "non-phylogenetic" adaptation from "phylogenetic" residual deviations. For example, methods such as phylogenetic autocorrelation (e.g., Cheverud et al. 1985) and phylogenetic eigenvector regression (e.g., Diniz-Filho et al. 1998) are based on explaining trait variation with one or more descriptors of phylogenetic distance (e.g., phylogenetic eigenvectors) and analyzing trait–trait or trait–environment relations in the reminder. This removes phylogenetically structured adaptation from consideration. While phylogenetic (and spatial) eigenvector regressions are useful descriptive tools, they are not well suited for estimating adaptation (see Freckleton et al. 2011; Diniz-Filho et al. 2012 for debate).

Figure 14.4 illustrates how any adaptive process that is sufficiently slow to generate a phylogenetic signal in model residuals will also generate systematic deviations from the optimal state that would manifest as a lag or "bias" in the mean structure of the model (Hansen et al. 2008; Hansen and Bartoszek 2012;
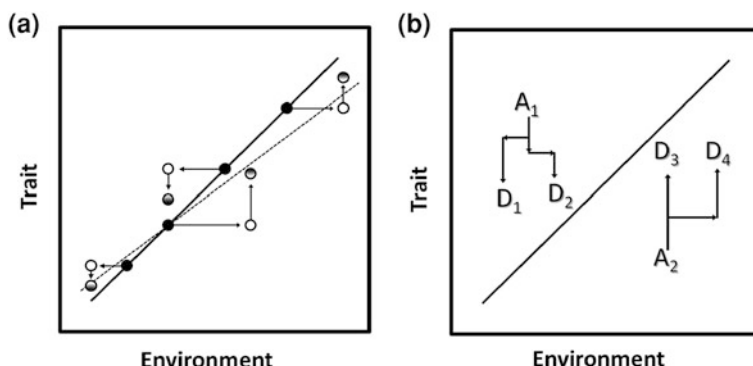
Fig. 14.4 Maladaptation, optimality, and phylogenetic effects. **a** How maladaptation flattens the evolutionary regression. A set of ancestral species (*black dots*) sit on a line describing the (primary) optimal relation between the trait and the focal environment. Secondary environmental changes may shift the species off the line in random directions (*horizontal arrows* to *open dots*), leading to a more shallow "evolutionary regression" (*dotted line*). The species then undergo adaptive evolution (*vertical arrows*), and the evolutionary regression will approach the optimal regression as the ancestral maladaptation is reduced. **b** How maladaptation leads to residual correlation. Two maladapted ancestral species, "*A*", speciate while evolving toward their (primary) optimal states. The descendant species, "*D*", deviate from optimality in a correlated manner. This correlation disappears as they reach their optima. Inspired by Sober (2008), and developed in collaboration with A. Labra

Bartoszek et al. 2012). This shows that the past history of the predictor variables should enter into the design matrix, **D**, and the weighting of the past should match with the degree of phylogenetic signal in the residual variance matrix, **V**. If not, estimates of the optimal relation to the predictor variables will be biased. For example, estimated regression slopes may be more shallow than the optimal relation between variables as illustrated in Fig. 14.4a. This means that the standard practice of modeling residual variance and mean structure separately is inconsistent with evolution toward optimal states. Species simply cannot provide unbiased information about optimal trait–environment relations at the same time as they retain ancestral residual correlations with other species (Fig. 14.4b).

Until the mid 2000s, almost all phylogenetic comparative analyses of continuous traits were based explicitly or implicitly on the assumption that model residuals evolve as an undirected Brownian motion. The Brownian motion has the property that the expected state of a descendant must equal the state of its ancestor. Hence, there is no mechanism for a systematic decrease of ancestral discrepancies between trait and environment. Instead, the degree of maladaptation will increase gradually through time due to the undirected random changes. Hansen and Orzack (2005) pointed out that this leads to what they called the problem of inherited maladaptation. If the tip species live in different environments, then the environment must have changed somewhere on the phylogeny, and then the change in the predictor variable implies that the species must jump to a new state where it has the same residual deviation as it had before. It inherits the maladaptation of its
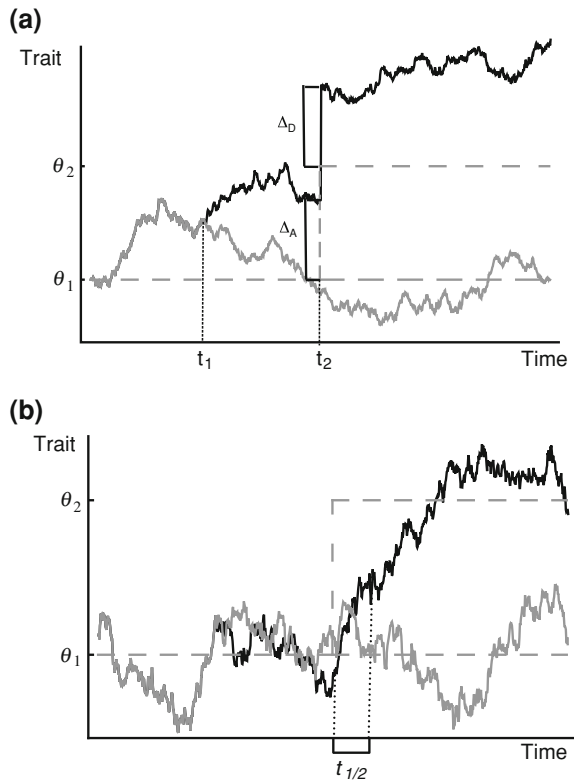
(a)

(b)

Fig. 14.5 Stochastic-process models of adaptation along a (two-species) phylogeny. **a** Evolution follows a Brownian motion. The ancestral species splits into two (*black* and *gray*) at time t_1, and at time t_2, one of the sister species (the *black*) experience an environmental change as its optimum (*dashed gray lines*) changes from θ_1 to θ_2. In a standard phylogenetic comparative analysis, this would be modeled by adding the difference $(\theta_2 - \theta_1)$ to the species (as the model is the sum of the environmentally predicted value and residuals evolving as Brownian motion). Note how this implies that ancestral maladaptation (Δ_A) must be transferred to maladaptation in the new environment (Δ_D) even if the ancestral species happened to be close to the new optimum. Note also how the Brownian motions tend to drift further and further away from the optimum. Maladaptation is increasing (on average) even in a constant environment. **b** Same as **a** except that the species now evolve according to an Ornstein–Uhlenbeck process around their optima. Here, the species tend to be pulled toward their optima if they diverge too far, and note how the two species lose their ancestral correlation much faster than with the Brownian motion. When the environment changes, the (*black*) species is gradually pulled toward its new optimum. The phylogenetic half-life $(t_{1/2} = \ln2/\alpha)$ is the time it takes to get halfway there (on average)

ancestor even when this implies it must jump across or away from its "optimal" state (Fig. 14.5a). This shows that, while the standard model can be used to estimate the statistical influence of an environmental variable on a trait, it is not a biologically coherent model of adaptation to fixed optima, and its estimates should not be taken as estimates of optimal states.

14.6 Modeling Adaptation

A way to approach the above problems is to base the comparative method on the Ornstein–Uhlenbeck process in place of the Brownian motion (Fig. 14.5; and see Chap. 15). The Ornstein–Uhlenbeck process models trait change as a sum of a white noise and a deterministic pull toward a particular state (the "optimum"). The model is

$$dy = -\alpha(y - \theta)dt + \sigma\, dW, \tag{14.3}$$

where dy is the trait change in a time interval dt, θ is the "optimum," and the dW is independent normally distributed random variables with mean zero and variance proportional to dt (i.e., a white noise). The parameters α and σ describe, respectively, the strength of the pull toward the optimum and the standard deviation of the stochastic changes. If α is zero, the pull to the optimum disappears and the Brownian motion is regained.

In Hansen (1997), I proposed a solution to the above problems based on assuming that the predictor variables acted on the optimum, θ, and thus only indirectly on the trait. Hence, the focus of estimation was shifted from the direct relationship between trait and environment and on to the relationship between the environment and the optimal state. In this way, species are allowed to be influenced by past environments, to lag behind their current optimal state, and to deviate in manners consistent with the deviation of their relatives. In the simplest cases, the model predicts that past environments have an influence proportional to an exponential function of the elapsed time since the environment occurred multiplied by α and that the covariance between related species is proportional to the exponential of the time separating them multiplied by α. The exact equations for this and various extensions of the model can be found in Hansen (1997) and Hansen et al. (2008).

This setup aligns the comparative method with within-species optimality studies of adaptation. In both cases, adaptation is studied as the influence of variables on an assumed optimum. As I argued above, the main reason why species trait means are different is because they evolve around different local optima determined by many factors. Optimality studies are concerned with testing the effects of one or a few variables on the optimum, and residual deviations are mainly due to the fact that we cannot know all relevant variables. This means that residual deviations are not due to maladaptation in the general sense, but reflect "maladaptation" relative to a focal selective agent under study. To make this distinction explicit, I introduced the concept of a *primary optimum*, defined as the average optimum reached by a number of species evolving in the same niche for sufficient time to allow ancestral constraints to disappear (Hansen 1997). This term was inspired by Simpson's (1944) concept of a primary adaptive zone. The idea is that the niche or adaptive zone is described by one or a few factors, say the grazing adaptive zone being defined by using grasses as major food source. The hypothesis that a trait, say hypsodont teeth, is an adaptation to this niche would then predict

that the primary optimum for tooth crown height of grazers is different from the primary optimum of other dietary niches. The local optimum of particular species in this niche may deviate from the primary optimum due to differences in secondary variables that do not enter the niche definition, say body size, skull shape, or the availability of alternative food sources, but if the trait is significantly influenced by the primary niche, then we predict systematic differences between the primary optima.

More specifically, our approach starts as a linear model of the primary optimum, determined by one or a few predictor variables that will differ across species or parts of the phylogeny. It then uses an Ornstein–Uhlenbeck model for evolution around this, not necessarily constant, primary optimum. The simplest setup is to map different niches onto a phylogeny and then estimate the value of the primary optimum in each of these niches. If θ_j is the value of the primary optimum in niche j, Hansen (1997) showed that the predicted trait value for species i is

$$\hat{y}_i = c_{0i}y_a + c_{1i}\theta_1 + c_{2i}\theta_2 + \cdots + c_{ki}\theta_k, \tag{14.4}$$

where y_a is the ancestral state at the root of the phylogeny, and the coefficients c_{ji} represent the influence of environmental state j on species i. Each period in the species' past history that was associated with a niche contributes a term $e^{-\alpha t_e} - e^{-\alpha t_b}$ to the coefficient of that niche with t_e and t_b being the times back to the end and the beginning of the period. Hence, each coefficient c_{ij} is the sum of such terms for each period in which the species i was associated with niche j. The coefficient c_{oi} equals $e^{-\alpha t_r}$, where t_r is the time back to the root of the phylogeny. Hence, all coefficients are between zero and one, and they sum to one. The coefficient corresponding to a particular environmental state will be large when the species has spent a lot of its history associated with this state, but more recent associations are weighted more heavily than more ancient ones. The larger the rate of adaptation, α, the more the weighting is shifted toward recent environments, and when α approaches infinity, then only the current environment is weighted and the model converges on a standard non-phylogenetic linear model.

There are many variations and extensions of this basic model. Butler and King (2004) developed a method for evaluating and comparing different niche arrangements with information-theoretical criteria. See also Ingram and Mahler (2013) for further extensions (Chap. 18). This is useful because the Achilles heel of the method is the need for reconstructing the niches on a phylogeny. The test of adaptation relies on the reconstruction of the historical past, and as I will discuss below, ancestral-state reconstruction is often unreliable and prone to systematic biases in the absence of true historical information. Hansen et al. (2008) presented another approach to deal with this problem based on assuming a model of evolution for the predictor variables and then using only tip-species observations as data. Their model relied on Brownian motion for the predictors and is thus only applicable to continuous predictor variables, although is possible to combine this with discrete reconstructed niches in ANCOVA types of models for the primary optimum (Escudero et al. 2012; Voje and Hansen 2013). Extensions to randomly

evolving discrete predictors (e.g., discrete niches) based on simple Markov chain models are possible, but have not yet been developed. Bartoszek et al. (2012) extended the method to multivariate response variables that can both influence each other's primary optima and be stochastically correlated. A special case of the Bartoszek et al. (2012) model extends Hansen et al. (2008) by allowing predictor variables to follow an Ornstein–Uhlenbeck model. Beaulieu et al. (2012) extended the method in a different direction by allowing not only the optima but also the α-parameter and σ-parameter to differ among reconstructed niches. This allows testing of niche- and clade-dependent rates of adaptation and evolution in a quite general and flexible manner (Chap. 15). See also Lajeunesse (2009) for implementation of the method into a meta-analytic framework.

It is important to understand that little is gained by just using the Ornstein–Uhlenbeck process to model the variances and covariances in the residual **V** matrix as done in many studies and implemented in some software. This simply transforms the form of the relationship between covariance and phylogeny and does not capture the effects of species tracking optima throughout the phylogeny. It does not adequately address any of the above-mentioned problems, which derive from the artificial separation of adaptation and residual deviation.

14.7 Using the Method to Study Adaptation of Deer Antler Size

For illustration, I present some analyses of a data set on deer antler size from Plard et al. (2011) and also analyzed by Bartoszek et al. (2012). The basic data are measures of antler length in 32 species from the deer family (Cervidae), and to test the hypothesis that antler size is influenced by sexual selection, we use a classification from Clutton-Brock et al. (1980) of the mating system into large, medium, and small breeding groups. The idea is that there will be stronger sexual selection in the large breeding groups. The small breeding group is close to monogynous systems. In Fig. 14.6a, we see a parsimony mapping of the breeding-group systems on a phylogeny. I used the program Slouch (Hansen et al. 2008) to run an analysis with log antler length as response variable and breeding-group-size niche as a predictor variable with three states. The estimated primary optima for antler size corrected for body mass are shown in Fig. 14.6b. They indicate a strong effect in the predicted direction, and the results are consistent with sexual selection being important. The phylogenetically corrected R^2 is 41 % after body size is corrected for and 87 % for the whole model.

These estimates are conditional on the best estimates of the α-parameter and σ-parameter of the Ornstein–Uhlenbeck model. Figure 14.6c shows the likelihood surface for these parameters that to aid interpretation have been transformed to two new parameters: $t_{1/2} = \ln2/\alpha$ and $v_y = \sigma^2/2\alpha$. The phylogenetic half-life, $t_{1/2}$, is the average time it takes to evolve half the distance from the ancestral state towards an
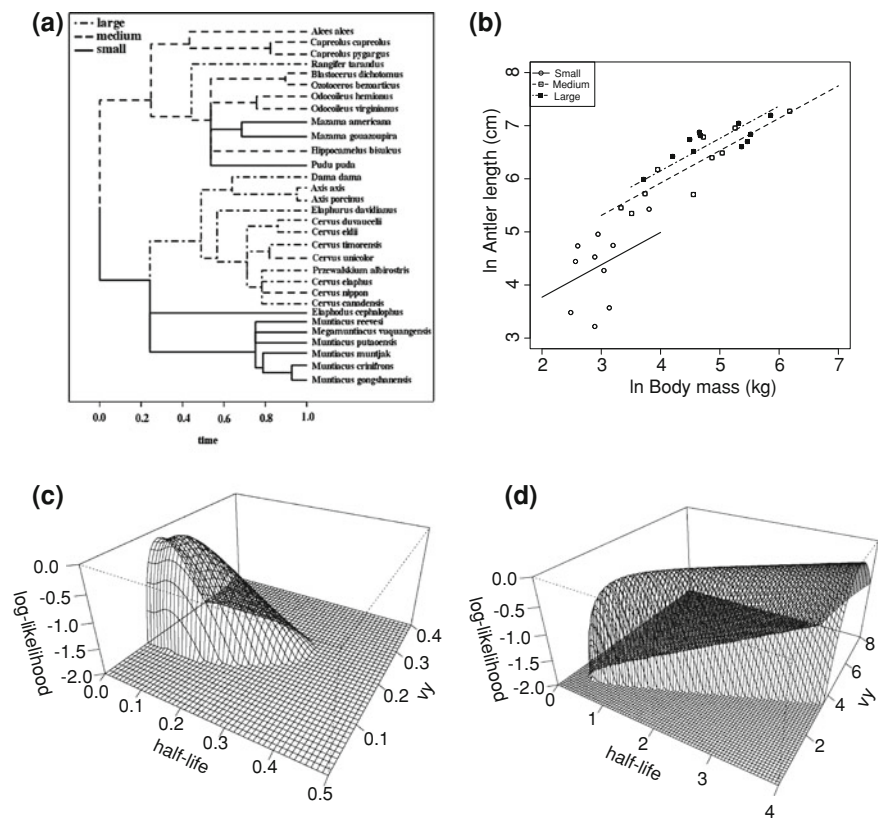
Fig. 14.6 Analyzing adaptation of antler size to breeding-group size using Slouch. **a** Mapping of *small*, *medium*, and *large* breeding groups as "niches" on the phylogeny (reprinted from Bartoszek et al. 2012 with permission from Elsevier). **b** Estimated primary optima for log antler length on the three breeding group sizes given as regressions on log body mass. The slope on log body mass is 0.61 ± 0.18 ($\pm$standard error), and the intercepts for the *small*, *medium*, and *large* breeding groups are, respectively, 2.55 ± 0.56, 3.48 ± 0.83, and 3.71 ± 0.90. The model explains $R^2 = 87$ % of the variance. **c** Log-likelihood surface for phylogenetic half-life $t_{1/2}$ and stationary variance, v_y for this model. The best estimates are $t_{1/2} = 0.07$th and $v_y = 0.16$. **d** Log-likelihood surface for the same parameters from a model with only an intercept. The best estimates here are $t_{1/2} = 2.4$th and $v_y = 3.6$, but the surface has a ridge that extends outward to infinity indicating that the phylogenetic effect is indistinguishable from the pattern expected from Brownian motion. Note different scales from **c**. Flat areas in the likelihood surfaces correspond to parameter values for which the log-likelihood is more than two units worse than the best estimate. Data for analysis originally from Plard et al. (2011)

optimum in a new niche as illustrated in Fig. 14.5b. The stationary variance, v_y, is a measure of how much the trait tends to deviate from the primary optimum when evolution has come to a stochastic equilibrium. The likelihood function peaks at a half-life of $t_{1/2} = 7$ % of time from tip to root of the phylogenetic tree. Hughes et al. (2006) estimated that the most recent common ancestor of the Cervidae lived

between 16 and 23 million years ago. Taking this into account, the half-life is between $t_{1/2} = 1.1$ and $t_{1/2} = 1.6$ million years. While this corresponds to a considerable lag in adaptation from a microevolutionary perspective, it is practically undistinguishable from instant adaptation on the timescales of this phylogeny. Instantaneous adaptation ($t_{1/2} = 0$) has only marginally lower likelihood. We can exclude long half-lives, however. The log-likelihood has dropped about two units when the half-life has reached 30 % of the distance from root to tip, which may be described as a moderate phylogenetic effect.

Now consider Fig. 14.6d, which shows the log-likelihood surface for $t_{1/2}$ and v_y for a model with only an intercept (i.e., no predictor). In this case, the phylogenetic half-life measures the phylogenetic signal in the trait (log antler size) as opposed to in the model residuals. The best estimate of the half-life in this case is 2.4 times the time back to the most recent common ancestor of the Cervidae, or between 38 and 55 million years. Half-lives that are much longer than the time back to the common ancestor are essentially indistinguishable from the pattern of a Brownian motion, which has a half-life of infinity. Consequently, the likelihood surface in Fig. 14.6d has a long ridge that extends out to infinity with only a microscopic drop in likelihood. In contrast, zero half-life is here soundly rejected (about 8 log-likelihood units worse). Hence, if we look at antler size in isolation, we see a pattern that resembles Brownian motion, and any conventional check for phylogenetic signal would be consistent with the use of a standard phylogenetic method as the independent contrasts. Looking at the phylogenetic signal in model residuals in Fig. 14.6c, however, shows that this would be a serious mistake. The weak phylogenetic signal in the residuals means that a non-phylogenetic analysis would be close to optimal. The reason why this happens is that antler size tracks predictor variables with strong phylogenetic effects (as seen in Fig. 14.6a) and thus inherits this phylogenetic signal even if there is little phylogenetic signal in the residual deviations. This is precisely the situation that was illustrated in Fig. 14.3a.

Although consistent with a strong effect of sexual selection on antler size, the fitted model is crude. The characterization of the breeding group niches is qualitative, the reconstruction of niches on the phylogeny is almost certainly seriously inaccurate, the phylogeny itself is inaccurate (cmp. Figure 14.2 with Fig. 14.6a), and the antler- and body-size measurements have unknown measurement error. It is indeed remarkable that the model explains so much of the variance. Bartoszek et al. (2012) give a more detailed analysis of the same data. There, antler size and sexual size dimorphisms are analyzed together as coevolving variables based on a multivariate Ornstein–Uhlenbeck process around niche-dependent optima. This analysis shows that antler size and sexual size dimorphisms are not coevolving in the sense of influencing each others' primary optima, but they show correlated evolution in their stochastic deviances that must be due to correlations in some secondary variables, and their primary optima depend on breeding group size in the predicted manner.

14.8 Interpreting the Parameters in the Ornstein–Uhlenbeck Model

Lande (1976) showed that the Ornstein–Uhlenbeck process could be derived from a simple model of quadratic stabilizing selection and genetic drift, and this is sometimes used to justify and interpret the parameters in the comparative method. Under this interpretation, the α-parameter becomes equal to the product of additive genetic variance and the curvature of the stabilizing selection function and the σ^2 parameter becomes equal to the additive variance divided by the effective population size. This is on a generational timescale, however, and on the million-year timescales that usually separate distinct species, the rates of evolution that can be predicted from estimates of evolvability and strengths of selection are so high as to be essentially instantaneous (Hansen 2012).

The conclusion from this is that evolution on a fixed fitness surface is usually so rapid that we do not expect any ancestral effects or any similarity of related species beyond what is due to similar positions of their adaptive optima. We know, however, that related species are similar, and this requires an explanation. A clue to explanation may be to ask why a primary optimum may not be reached instantaneously.

A possibility suggested by Hansen (1997) was based on Simpson's (1944) idea that primary adaptations may be slow due to the need to establish a number of secondary adaptations before it can reach its fullest expression. This idea is perhaps most clearly expressed by Kemp (2006, 2007), who referred to it as the correlated-progression hypothesis. Here, we imagine the focal biological trait as embedded in a complex network of coadapted interactions with other traits. Any large change in a focal trait is thus unlikely to be beneficial due to internal selective constraints even if there is external selection to change it. A small change may, however, be beneficial on balance, and this will set up selection on other trait to adjust to this small change, which will again make possible a further change in the focal trait, and so on. Thus, external directional selection on a trait (or suite of traits) may result in a slow correlated progression. On this view of evolution, a small α-parameter may result from strong internal selective constraints. This point is discussed in more detail in Labra et al. (2009) and Hansen (2012).

The definition of the primary optimum as an average optimum over repeated reruns of evolution in a given niche assumes that evolution in the niche tends to evolve around this optimum. It does not capture cases with two or more distinct centers of attraction, as with distinct alternative strategies or disruptive selection where more extreme alternatives are favored. We may, however, extend the concept to cover niche-dependent trends of evolution. A primary trend may be defined as the average directional rate of evolution over repeated reruns of evolution in the given niche. In the Ornstein–Uhlenbeck model, this manifests itself in cases where a long phylogenetic half-life is combined with primary optima that are far outside the range of among-species variation. In this case, the average directional change per time is equal to $\alpha\theta$, and the trends can thus be estimated by

multiplying the estimated primary optima with the estimated α-parameter. In these situations, the model behaves like a Brownian motion with niche-dependent trends (see Hansen 1997).

14.9 Software and Applications

These adaptation methods have now been used in many studies and are at least partially implemented in several software packages. The popularity of the methods is to a large degree due to Butler and King's (2004) R package Ouch, which can estimate optima in fixed niches mapped onto a phylogeny and assess the fit of different niche arrangements. Later, several other packages with complementary functionality have been developed. Slouch, originally developed by Hansen et al. (2008) to fit the model with randomly evolving predictor variables, has also been extended to handle fixed niches mapped on the phylogeny combined with random effects. It differs from Ouch by using different parameterizations of the model ($t_{1/2}$ and v_y in place of α and σ^2) and by using a grid search in place of a numerical optimization algorithm. The latter is an advantage in that it makes the uncertainty in parameter estimation more apparent and guards against convergence on suboptimal peaks in the likelihood, but a disadvantage with large numbers of species, as it can be slow and cumbersome to use. Slouch can also be used to correct for known observation variance in both response and predictor variables. Beaulieu et al's (2012) Ouwie allows not only the primary optima but also the α-parameter and σ-parameter to depend on niches, and is hence useful for estimating environment-dependent rates of evolution and adaptation (in this respect, it extends Brownie, O'Meara et al. 2006). MvSlouch (Bartoszek et al. 2012) allows several traits to evolve around a multivariate primary optimum determined by fixed niches or other randomly evolving variables. A restricted multivariate model can also be fitted by Ouch. The multivariate models are very parameter rich, however, and their application is technically difficult and requires much data and well-specified hypotheses to constrain the number of parameters. The program Surface (Ingram and Mahler 2013) includes routines for identifying and comparing models of convergent evolution based on alternative niche reconstructions on the phylogeny (Chap. 18). Versions of the model can also be fit by the programs Ape (Paradis et al. 2004), Brownie (O'Meara et al. 2006), Compare (Martins 2004), and Geiger (Harmon et al. 2008).

I am aware of more than 70 published applications of these methods, which will be reviewed elsewhere (Hansen, Pienaar, Voje, Bartoszek in preparation). The main impression is that most of these studies are able to find evidence of adaptation in that they can convincingly reject alternative non-adaptive models and provide parameter estimates in accordance with a priori expectations from the adaptive hypothesis. Importantly, adaptation is also sometimes rejected. For example, Labra et al. (2010) found no signs that the size of the parietal (third) eye was adapted to any climatic or thermophysiological variables in a lizard genus (*Liolaemus*), thereby providing evidence against long-standing hypotheses of

thermobiological functionality for this organ. It is commonly found that adaptation is fast enough to produce none to mild residual phylogenetic correlations, but this is not always the case, and it remains necessary to consider phylogeny. The standard phylogenetic comparative model would, however, only have been appropriate in a small minority of cases and would then also need to be reinterpreted in terms of estimating trends. A recurring problem with many of these studies is the failure to report parameter estimates, or reporting them without units. This limits the possibility for interpreting and generalizing from this body of work.

14.10 Testing for Adaptation in Qualitative (Categorical) Traits

The study of adaptation has always been one of the major uses for comparative methods. Most usage is simply looking for predicted associations between traits and environments without controlling for phylogeny, or based on phylogenetic corrections that are not justified with appropriate process models. Either of these approaches may be informative in a qualitative sense, but falls short of delivering logically consistent quantitative estimates of precisely defined parameters. For continuous quantitative characters, the explicit adaptation models are now sufficiently well developed to fill the need.

The situation is less clear for qualitative traits. Although there are stochastic-process models appropriate for categorical traits, there has been no discussion of how to parameterize adaptation along the lines I discussed above. Simple Markov-chain models can be used to estimated the intensities (probabilities per time) of change between states of a variable as a function of the state of another variable (e.g., Pagel 1994), and this allows tests of adaptation in similar ways as above (e.g., Hansen and Orzack 2005). In particular, maintenance of a trait A in the presence of environment B can be used as evidence of adaptation of trait A to environment B. Armbruster (2002) used this method to test whether bract color (green/white vs. pink/purple) in *Dalechampia* blossoms depended on type of bee pollinator (euglossine vs. megachilid). He found no support for this in a study of 37 species. He did, however, find support for an influence of vegetative colors on the evolution of bract colors, concluding that bract colors may be changing due to indirect selection stemming from a pleiotropic relation to stem pigments.

In analogy with the adaptation model for continuous traits, one could define adaptation as a systematic niche-dependent effect on the rates of transition between states, or equivalently, on the equilibrium probabilities of the states. This could be done directly in a generalized linear model setup or indirectly by allowing the categorical trait to be a threshold function of an underlying continuous variable that itself follows the adaptation model, but such models remain to be developed. Felsenstein (2012) has shown how a threshold model on traits evolving as Brownian motion can be fit to discrete comparative data, and it should be possible to extend this to Ornstein–Uhlenbeck models (see also Chap. 16).

14.11 Ancestral-State Reconstruction and Inferred-Changes Methods

Perhaps the most common comparative test of adaptation is to reconstruct character and environmental states on a phylogeny and then use these as data in various statistical analyses. In the cladistic tradition, this is done by inferring trait changes with parsimony methods and then using these inferred changes as data in subsequent analyses (Maddison 1994). Ancestral states and changes can also be inferred by assuming specific evolutionary process models such as Brownian motion or Markov chains (Martins and Hansen 1997; Schluter et al. 1997) and then either the reconstructed states at nodal values or the inferred changes along branches can be used as data in subsequent analyses.

Unfortunately, inferred changes on a phylogeny are poorly suited for statistical analysis (Frumhoff and Reeve 1994). In the cladistic tradition, inferred changes are treated as independent evolutionary events, and it is assumed that using them as data solves the phylogenetic correlation problem that arises when using species data. Ridley (1983, p. 18) makes this argument explicit when he writes "To recognize independent evolution is to distinguish primitive from derived character states. Derived characters are independently evolved characters." From this, he proposed a research strategy for testing adaptation based on using inferred changes from primitive to derived character states as independent data in statistical analyzes. In my opinion, his argument is mistaken in two interesting ways.

The first and most obvious problem is that the inferred changes to be used as data are not the actual changes that have happened in the history of life, but estimates thereof. Even if the true changes could be regarded as independent evidence, the estimates cannot. Inferred changes based on parsimony (or any other method) are not independent of each other. Whether a change is inferred to have happened on a particular branch in the phylogeny depends on what changes are inferred on other branches. Using inferred changes as data in non-phylogenetic statistical analysis will seriously violate the assumptions of independent sampling and is likely to give misleading information about evolution.

If the ancestral states are inferred from a statistical model such as a Markov chain or a Brownian motion, then it is possible to derive the joint statistical distribution of the inferred changes including their variances and covariances (Martins and Hansen 1997). This makes it possible to use them correctly in statistical analyses that take account of covariance and heteroscedacity. Doing this, however, reveals a fundamental circularity of the inferred-changes approach. First, a model is assumed and parameters may be estimated to infer the changes, and then, the inferred changes are used to make inferences about evolution. It is clear that those inferences are constrained to reflect the assumptions of the model used to infer the changes. For example, a set of ancestral states reconstructed from extant species data by assuming a Brownian-motion process will have the same average as the extant species. Hence, one will observe that there is no temporal trend in the (inferred) data, but this is just recovering a property of the assumed

model and does not constitute evidence against evolutionary trends. More subtle errors of this sort are hard to avoid. In any case, the likelihood principle of statistics tells us that given an evolutionary process model, then all the information about its parameters is contained in the probability of the observed data given the model (i.e., in the likelihood function). From this, it follows that reconstruction of ancestral states can never provide any information that cannot be obtained by a standard likelihood analysis of the observed tip species. The best that can be hoped from an inferred-changes method is a roundabout way to construct a likelihood function, and I am not aware of a single application that has achieved even that.

In addition to the statistical problem of using inferred changes as if they were true changes, there is also an interesting biological problem with Ridley's reasoning. It concerns his assumption that the true changes from primitive to derived character states can be treated as independent observations of an evolutionary process. The problem with this can be seen from our model (14.1) of how a trait is determined by a number of observed and unobserved factors. The model was: $Y = b_0 + b_1 X_1 + r(X_1, \ldots, X_m)$. If a change on a branch happens because the focal variable, X_1, changes with an amount ΔX_1, then we may predict a change $\Delta Y = b_1 \Delta X_1$, which would be independent of changes on other branches provided, and this is the problem, provided that there are no correlations between branches in the changes of secondary variables that make the residuals. If, due to some inherited common biology, X_2 has a tendency to increase in one part of the phylogeny, then the residuals from a regression of ΔY on ΔX_1 will tend to be similar in this part of the phylogeny. Put in other words, evolutionary changes in different parts of a phylogeny are expected to be correlated for exactly the same reason that species trait values in different parts of the phylogeny are expected to be correlated. The underlying reason is that both states and changes are dependent on shared third variables inherited from a common ancestor. This is also the reason why pairwise contrasts on a phylogeny cannot be taken to be independent except under specified precise assumptions about the underlying evolutionary process, such as in the derivation of independent contrasts from a multivariate Brownian motion. For example, it can be shown that independent contrasts are not independent if the species data were generated by an Ornstein–Uhlenbeck process.

In molecular evolution, the reconstruction of ancestral genes and proteins is becoming increasingly popular. This approach has an advantage over the reconstruction of higher level physiological, morphological, behavioral, and life-history traits in that molecular reconstructions are based on models of evolution that are better understood and delineated. Still, the fundamental problems are the same, and reconstructed genes and proteins should not be used to test laws of adaptation. Also here, the principled approach would be to estimate parameters of defined models based on extant species data without the vivid but statistically unnecessary pass through a reconstructed ancestor.

14.12 A Measurement-Theoretical Perspective on Comparative Methods

Measurement is the assignment of numbers to attributes of reality. Measurements are valid when relations among the numbers reflect relevant relations among the attributes of reality. Only then can inferences made from the numbers lead to meaningful inferences about nature. Measurement theory is the study of the mapping from reality to numbers. Houle et al. (2011) pointed out that a measurement-theoretical perspective has been lacking from biology and argued that a large number of common problems, ranging from relatively obvious errors such as ignoring units or using p values as measures of effect to more subtle problems involving violations of scale type or making statistical manipulations that violate the theoretical context of the study, can be diagnosed as violation of measurement-theoretical principles.

Comparative studies are extremely vulnerable to such problems because there is a need to simultaneously handle difficult statistical problems, error prone data of heterogeneous origin, and quantification of poorly understood theories of evolution on long timescales. The drive to solve statistical problems can easily come into conflict with studying meaningful biological relations, and without a firm grasp of the meaning of measurement, it is easy to loose the biological baby with the statistical bathwater.

An undercurrent of this essay has been that the statistical laundering that took place with the development of phylogenetic comparative methods has also weakened the connections to biology, and particularly so in the treatment of adaptation. I have argued that the assumption of evolution as Brownian motion has fundamental incompatibilities with adaptation in the sense of evolution toward niche-dependent optima. The Brownian-motion assumption arose from the need to model interspecies covariances as shared branch lengths so as to be able to use GLS techniques without having to estimate extra nuisance parameters in the variance matrix. This implicitly assumes that the effects of adaptation could be separated from the residual covariances and that the estimated parameters in regression and ANOVAs could be interpreted in the same way as they would in non-phylogenetic analyses. As shown above, the meaning of these parameters change. For example, the regression slope becomes influenced both by the optimal relation and by the expected evolutionary lag (Fig. 14.4).

Evolutionary regression studies are often conducted in a descriptive statistical manner without quantitative links to the theory that motivated the study. A qualitative prediction of, say, a positive relationship between two variables may motivate the study, but the context of this prediction is often not used to inform or constrain the statistical procedure. When statistical manipulations are decoupled from theoretical context, this will often lead to measurement-theoretical problems. Using interspecies trait covariances in place of residual covariances needs justification, and even if such justification can be derived in some cases, as when all variables follow a joint multivariate Brownian motion, it is typically incorrect, as

discussed above. I suspect this mistake was allowed to proliferate to such a degree because there was no demand to formally justify statistical methods and procedures from biological assumptions. I will now briefly discuss two more examples where statistical arguments void of biological justification have lead to serious errors in comparative studies.

One case concerns the estimation of allometric relationships, which is important both as a goal in itself by testing hypotheses about scaling relationships and functional adaptations, and due to its widespread use in controlling for body size. While essentially all theory about allometric scaling refers to a power relation between variables (i.e., $Y = aX^b$, where Y is a trait and X is usually a measure of body size, as in Fig. 14.1), there appeared in the 1970s a verbal broadening of the term allometry to describe all kinds of nonlinear trait relationships. This lead to a curious situation where predictions derived from or about allometry in the narrow sense of a power relationship were tested with statistical methods incompatible with the power relationship. For example, the allometric exponent b is traditionally estimated from a log–log regression ($\log[Y] = \text{Log}[a] + b\text{Log}[X]$), but in the context of broad-sense allometry, the exponent has no meaning and the theoretical necessity of the log-transformation disappears, so that researchers started to do "allometric" linear regression on the arithmetic scale (see Houle et al. 2011; Voje et al. 2014 for discussion and documentation). This was sometimes accompanied with arguments that the log-transformation was not necessary for statistical reasons (e.g., it was not necessary to stabilize the variance or to achieve normality). When theoretical context was thus forgotten, a theoretically necessary manipulation became confused with a statistical manipulation and was then often dropped for statistical reasons. One result of this was a substantial body of work and an emerging consensus that static and ontogenetic "allometries" are evolvable and flexible traits and not important constraints on evolution. Voje et al. (2014; Voje and Hansen 2013) have shown, however, that none of this work apply to allometry in the traditional narrow sense and that there is in fact considerable evidence that static allometries are highly constrained and may therefore act as evolutionary constraints on trait adaptation even on macroevolutionary timescales as hypothesized by Huxley (1932), Rensch (1959), Gould (2002), and others.

Another common violation of measurement-theoretical principles in cross-species regression studies involves the use of nonparametric curve-fitting techniques such as reduced major-axis regression. Such techniques are based on geometrical arguments on how to draw a line through points and not justified in terms of estimating parameters in a defined model. In particular, the common practice of using the reduced major-axis slope to estimate allometric exponents has obscured our empirical knowledge of allometric scaling relationships (Voje et al. 2014). This slope is computed as the ratio of the standard deviations of the variables involved and does not even include their covariance. Except when points deviate little from a straight line, it typically gives a seriously misleading estimate of a true underlying regression slope (Kelly and Price 2004; Hansen and Bartoszek 2012). Using the reduced major-axis slope to estimate an allometric exponent or a causal effect of a predictor variable on a trait are measurement-theoretical mistakes

because there are no connections between the statistical procedure and the entities to be estimated.

Reduced major-axis and other nonparametric regression techniques are sometimes said to improve on standard regression when there is "error" in both response and predictor variables. This is only true under highly specific and unlikely circumstances and seems to derive from an implicit assumption that all residual deviations are due to measurement error in the variables and not to biological deviations from the model (Hansen and Bartoszek 2012). In reality, most residual deviations from evolutionary regression models are due to biological deviations from the linear relationship as described in Eq. (14.1). Indeed, if this was not the case, there would be no reason to worry about phylogenetic correlations! In such cases, the "corrections" of the major axis and more general structural equation models are completely off the mark. In particular, reduced major-axis regression should never be used in comparative analyses, and every result drawn from this method should be reconsidered.

Yet, this is not to deny that measurement error is a major concern in comparative analyses. In a typical comparative analysis, data have to be gathered from many individuals within each of many species in a consistent manner. Often sample sizes for individual species are small and vary with orders of magnitude between species, so that some are much more reliable than others (Garamszegi and Møller 2010; 2011). Many comparative studies are also based on compiled data sets of dubious quality. For example, the data on antler-body-size regression shown in Fig. 14.1 and used in the classical analyses by Gould (1974) and Clutton-Brock et al (1980) are reported without any indication of uncertainty and appear to be of poor quality. As just one example, I traced the estimate of the body size of the fallow deer given as a 91.0-cm shoulder height to the following statement in Ward's (1903) *Record of big game*: "Height at shoulder about 3 feet" (Ward 1903, p. 64). Somehow this qualitative statement transmuted into a fixed quantitative measurement with three significant digits on the cm scale ready to be used in the comparative studies. This may seem outrageous, but I fear it is quite typical. Smith and Jungers (1997) documented even worse problems in a standard data set of primate body size used in many comparative studies.

The most important thing to do about this is to pay more attention to data quality and also not to accept measurements without units and indications of uncertainty (as I admittedly did in my example). If estimates of measurement error variation are available, for example, as standard errors of species means, methods have been developed to include these into phylogenetic comparative studies (e.g., Lynch 1991; Martins and Hansen 1997; Ives et al. 2007; Felsenstein 2008; Lajeunesse 2009; Hadfield and Nakagawa 2010; Hansen and Bartoszek 2012; Revell and Reynolds 2012; reviewed in Chap. 7). Also the bias in regression slopes induced by measurement error in predictor variables that sometimes motivates the use of nonparametric regression can easily be corrected with simple extensions of the standard phylogenetic linear model (Hansen and Bartoszek 2012).

14.13 Conclusion

The emergence of phylogenetic comparative methods may appear a textbook case of scientific progress. The growing recognition of statistical problems and inconsistencies led to increasing dissatisfaction with comparative biology, but this was resolved by the development of new methods of measurement and the incorporation of new types of data in the form of molecular phylogenies. Phylogenetic comparative methods then became the core of a new research paradigm that was more quantitative, statistical, and rigorous than what went before. Here, I have argued that this paradigm shift was not without its problems. In relation to the study of adaptation, the solutions to the statistical problems generated conceptual and interpretational inconsistencies that need to be resolved. I believe these inconsistencies at least now have been identified and that solutions are emerging at an accelerating pace. There is still more work to be done before we are able to deal with all the different types of data and theories that we would like to include in our analyses and tests of adaptation with comparative methods, but I hope this chapter can help clarify the principles for how this can be done.

References

Armbruster WS (2002) Can indirect selection and genetic context contribute to trait diversification? A transition-probability study of bolssom-colour evolution in two genera. J Evol Biol 15:468–486

Bartoszek K, Pienaar J, Mostad P, Andersson S, Hansen TF (2012) A comparative method for studying multivariate adaptation. J Theor Biol 314:204–215

Baum DA, Larson A (1991) Adaptation reviewed: a phylogenetic methodology for studying character macroevolution. Syst Zool 40:1–18

Beaulieu JM, Jhwueng D-C, Boettiger C, O'Meara BC (2012) Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. Evolution 66:2369–2383

Brooks DR, McLennan DH (1991) Phylogeny, ecology and behavior: a research program in comparative biology. University of Chicago Press, Chicago

Butler MA, King AA (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am Nat 164:683–695

Cheverud JM, Dow MM, Leutenegger W (1985) The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weights among primates. Evolution 39:1335–1351

Clutton-Brock TH, Albon SD, Harvey PH (1980) Antlers, body size and breeding group size in the Cervida. Nature 285:565–567

Coddington JA (1988) Cladistic tests of adaptational hypotheses. Cladistics 4:3–22

Cooper N, Jetz W, Freckleton RP (2010) Phylogenetic comparative approaches for studying niche conservatism. J Evol Biol 23:2529–2539

Diniz-Filho JAF, Ramos de Sant'ana CE, Bini LM (1998) An eigenvector method for estimating phylogenetic inertia. Evolution 52:1247–1262

Diniz-Filho JAF, Rangel TF, Santos T, Bini LM (2012) Exploring patterns of interspecific variation in quantitative traits using sequential phylogenetic eigenvector regressions. Evolution 66:1079–1090

Escudero M, Hipp A, Hansen TF, Voje KL, Luceño M (2012) Selection and inertia in the evolution of holocentric chromosomes in sedges (*Carex,* Cyperaceae). New Phytol 195:237–247

Felsenstein J (1985) Phylogenies and the comparative method. Am Nat 125:1–15

Felsenstein J (2008) Comparative methods with sampling error and within-species variation: contrasts revisited and revised. Am Nat 171:713–725

Felsenstein J (2012) A comparative method for both discrete and continuous characters using the threshold model. Am Nat 179:145–176

Freckleton RP, Cooper N, Jetz W (2011) Comparative methods as a statistical fix: the dangers of ignoring an evolutionary model. Am Nat 178:E10–E17

Frumhoff PC, Reeve HK (1994) Using phylogenies to test hypotheses of adaptation: a critique of some current proposals. Evolution 48:172–180

Garland T Jr, Bennett AF, Rezende EL (2005) Phylogenetic approaches in comparative physiology. J Exper Biol 208:3015–3035

Garamszegi LZ, Møller AP (2010) Effects of sample size and intraspecific variation in phylogenetic comparative studies: a meta-analytic review. Biol Rev 85:797–805

Garamszegi LZ, Møller AP (2011) Nonrandom variation in within-species sample size and missing data in Phylogenetic comparative studies. Syst Biol 60:876–880

Geist V (1998) Deer of the world: their evolution, behavior, and ecology. Swan Hill Press, Shrewsbury

Gould SJ (1973) Positive allometry of antlers in the "Irish elk", *Megaloceros giganteus*. Nature 244:375–376

Gould SJ (1974) The evolutionary significance of "bizarre" structures: antler size and skull structure in the "Irish Elk," *Megaloceros giganteus*. Evolution 28:191–220

Gould SJ (1977) Ontogeny and Phylogeny. Belknap, Cambridge

Gould SJ (1998) A lesson from the old masters. In: Gould SJ (ed) Leonardo's mountain of clams and the diet of worms. Harmon books, pp 179-196

Gould SJ (2002) The structure of evolutionary theory. Harvard University Press, Cambridge

Gould SJ, Vrba ES (1982) Exaptation—a missing term in the science of form. Paleobiology 8:4–15

Hadfield JD, Nakagawa S (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. J Evol Biol 23:494–508

Hansen TF (1997) Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351

Hansen TF (2012) Adaptive landscapes and macroevolutionary dynamics. In: Svensson EI, Calsbeek R (eds) The adaptive landscape in evolutionary biology. Oxford University Press, Oxford, pp 205–226

Hansen TF, Bartoszek K (2012) Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. Syst Biol 61:413–425

Hansen TF, Martins EP (1996) Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50:1404–1417

Hansen TF, Orzack SH (2005) Assessing current adaptation and phylogenetic inertia as explanations of trait evolution: the need for controlled comparisons. Evolution 59:2063–2072

Hansen TF, Pienaar J, Orzack SH (2008) A comparative method for studying adaptation to a randomly evolving environment. Evolution 62:1965–1977

Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W (2008) Geiger: investigating evolutionary radiations. Bioinformatics 24:129–131

Houle D, Pélabon C, Wagner GP, Hansen TF (2011) Measurement and meaning in biology. Quart Rev Biol 86:3–34

Hughes S, Hayden TJ, Douady CJ, Tougard C, Germonpre M, Stuart A, Lbova L, Carden RF, Hanni C, Say L (2006) Molecular phylogeny of the extinct giant deer, *Megaloceros giganteus*. Mol Phylogenet Evol 40:285–291

Huxley JS (1932) Problems of relative growth. Lincoln Mac Veagh-The Dial Press, New York

Ingram T, Mahler DL (2013) SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike information criterion. Methods Ecol Evol 4:416–425

Ives AR, Garland T Jr (2010) Phylogenetic logistic regression for binary dependent variables. Syst Biol 59:9–26

Ives AR, Midford PE, Garland T Jr (2007) Within-species variation and measurement error in phylogenetic comparative methods. Syst Biol 56:252–270

Kelly C, Price TD (2004) Comparative methods based on species mean values. Math Biosci 187:135–154

Kemp TS (2006) The origin of mammalian endothermy: a paradigm for the evolution of complex biological structure. Zool J Linn Soc 147:473–488

Kemp TS (2007) The origin of higher taxa: macroevolutionary processes, and the case of the mammals. Acta Zoologica 88:3–22

Labra A, Pienaar J, Hansen TF (2009) Evolution of thermal physiology in *Lioleamus* lizards: adaptation, phylogenetic inertia and niche tracking. Am Nat 174:204–220

Labra A, Voje KL, Seligmann H, Hansen TF (2010) Evolution of the third eye: a phylogenetic comparative study of parietal-eye size as an ecophysiological adaptation in *Liolaemus* lizards. Biological J Linn Soc 101:870–883

Lajeunesse MJ (2009) Meta-analysis and the comparative phylogenetic method. Am Nat 174:369–381

Lande R (1976) Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314–334

Larson A, Losos JB (1996) Phylogenetic systematics of adaptation. In: Rose MR, Lauder GW (eds) Adaptation. Academic press, San Diego, pp 187–220

Lister AM et al (2005) The phylogenetic position of the 'giant deer' *Megaloceros giganteus*. Nature 438:850–853

Lynch M (1991) Methods for the analysis of comparative data in evolutionary biology. Evolution 45:1065–1080

Maddison DR (1994) Phylogenetic methods for inferring the evolutionary history and processes of change in discretely valued characters. Ann Rev Entomol 39:267–292

Martins EP (2000) Adaptation and the comparative method. Trends Ecol Evol 15:296–299

Martins EP (2004) Compare, version 4.6b. Computer programs for the statistical analysis of comparative data. Distributed by the author at http://compare.bio.indiana.edu/. Technical report, Department of Biology, Indiana University, Bloomington, IN

Martins EP, Hansen TF (1997) Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am Nat 149:646–667

Mitchell WA, Valone TJ (1990) The optimization research program: studying adaptations by their function. Quart Rev Biol 65:43–52

Nunn CL (2011) The comparative approach in evolutionary anthropology and biology. The University of Chicago Press, Chicago

O'Meara BC (2012) Evolutionary inferences from phylogenies: a review of methods. Annu Rev Ecol Evol Syst 43:267–285

O'Meara BC, Ane C, Sanderson MJ, Wainwright PC (2006) Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922–933

Pagel MD (1994) Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc R Soc B 255:37–45

Paradis E, Claude J, Strimmer K (2004) Ape: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290

Pennell MW, Harmon LJ (2013) An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. Ann New York Acad Sci 1289:90–105

Plard F, Bonenfant C, Gaillard J-M (2011) Revisiting the allometry of antlers among deer species: male–male sexual competition as a driver. Oikos 120:601–606

Price T (1997) Correlated evolution and independent contrasts. Philos Trans R Soc B 355:1599–1606

Reeve HK, Sherman PW (1993) Adaptation and the goals of evolutionary research. Quart Rev Biol 68:1–32

Reeve HK, Sherman PW (2001) Optimality and phylogeny: a critique of current thought. In: Orzack SH, Sober E (eds) Adaptationism and optimality. Cambridge University Press, Cambridge, pp 64–113

Rensch B (1959) Evolution above the species level. Wiley, New York

Revell LJ (2010) Phylogenetic signal and linear regression on species data. Methods Ecol Evol 1:319–329

Revell LJ, Reynolds G (2012) A new bayesian method for fitting evolutionary models to comparative data with intraspecific variation. Evolution 66:2697–2707

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

Schluter D, Price T, Mooers AØ, Ludwig D (1997) Likelihood of ancestor states in adaptive radiation. Evolution 51:1699–1711

Simpson GG (1944) Tempo and mode in evolution. Columbia University Press, New York

Smith RJ, Jungers WL (1997) Body mass in comparative primatology. J Human Evol 32:523–559

Sober E (1984) The nature of selection: evolutionary theory in philosophical focus. Bradford books, Cambridge

Sober E (2008) Evidence and evolution: the logic behind the science. Cambridge University press, Cambridge

Stone GN, Nee S, Felsenstein J (2011) Controlling for non-independence in comparative analysis of patterns across populations within species. Phil Trans R Soc B 366:1410–1424

Voje KL, Hansen TF (2013) Evolution of static allometries: slow rate of adaptive change in allometric slopes of eye span in stalk-eyed flies. Evolution 67:453–467

Voje KL, Hansen TF, Egset CK, Bolstad GH, Pélabon C (2014) Allometric constraints and the evolution of allometry. Evolution 68: 866–885

Ward R (1903) Records of big game: with the distribution, characteristic, dimensions, weights, and horn and tusk measurements of the different species, 4th edn. Rowland Ward, Limited, London

Westoby M, Leishman MR, Lord JM (1995) On misunderstanding the 'phylogenetic correction'. J Ecol 83:531–534

Williams GC (1992) Natural selection: domains, levels, and challenges. Oxford University Press, Oxford