2001, Volume 10, Issue 3

Pages 77 -116

# Evolutionary Anthropology

## ISSUES, NEWS, AND REVIEWS

## IN THIS ISSUE



Wiley-Liss

A Division of John Wiley & Sons, Inc.

ARTICLES

# Comparative Methods for Studying Primate Adaptation and Allometry

CHARLES L. NUNN AND ROBERT A. BARTON

A well-known comparative biologist was once asked by a field biologist whether the latter's detailed and painstaking field study of orangutan behavior, carried out over many years, qualified as an example of the comparative method. "No," replied the comparative biologist, "that's an anecdote." The reply is somewhat harsh, as useful comparisons can be conducted both within and across species. The reply does emphasize, however, that analysis of patterns across species is fundamental to the study of adaptive evolution, particularly when variation needed to test hypotheses is present only at this interspecific level.[1–5] Examples in primatology include the occurrence of female sexual swellings in species with habitually multimale, rather than single-male, breeding systems,[6,7] the relationship between polygynandrous mating and relatively large testes size,[3,8] and the association between brain size and social group size.[9] Thus, in many cases, interspecific variation is required to test adaptive hypotheses.[10]

In this article, we review developments in comparative methodology that have emerged in the last fifteen years, with particular emphasis on the importance of incorporating phylogeny into comparative studies. We define "comparative study" as meaning analysis of trait variation across rather than within species, with the aim of testing hypotheses or generating new ones.[11] We provide information on how to implement phylogenetic comparative methods, including available computer programs (Box 1).

Charles Nunn is a Research Associate in the Department of Biology at the University of Virginia. His current research focuses on the macroevolutionary correlates of disease risk in mammals and methods for analyzing host-parasite trait evolution (http://faculty.virginia.edu/charlienunn). Email: charlie.nunn@virginia.edu.
Robert Barton is a Reader in Anthropology, and convenor of the Evolutionary Anthropology Research Group at the University of Durham, UK. His recent research uses comparative methods to study brain evolution. Email: R.A.Barton@durham.ac.uk

Evolutionary Anthropology 10:81–98 (2001)

For reasons of space, this review focuses mainly on the most commonly used methods, namely the method of independent contrasts[12] and the concentrated changes test.[13] Readers interested in other methods may wish to consult longer or more focused reviews.[10,14,15]

## WHY INCORPORATE PHYLOGENY IN COMPARATIVE STUDIES?

In recent years, biologists have identified many possible methods for the analysis of comparative data.[5,10,15–19] There has been much debate about the validity of these different methods in biological anthropology[20,21] and other fields such as ecology.[22–24] Although this debate has provided important insights, a consensus has emerged that comparisons must incorporate information about the phylogenetic relationships among the species under consideration.

Why is phylogeny necessary? There are three interrelated answers to this question. First, phylogeny allows the identification of independent data points. The concept of independence is as important for comparisons across species as it is for experimental and observational studies within species.[25,26] Closely related species tend to share traits because of their recent common ancestry,[10,27] but we can be more confident that a trait is an adaptation if it has evolved repeatedly, rather than once, in association with some other trait or environmental attribute.[28] Thus, it may be incorrect to consider a trait shared by multiple extant species as independent, and counted as multiple degrees of freedom, if it is shared among species through common descent rather than independent origin.[10]

Second, the usual aim of a comparative study is to document correlated trait evolution.[10] Hypothesizing that a dependent variable *Y*, such as brain size, is adaptively linked to an independent variable *X*, such as social group size, implies that the two variables have evolved together. Phylogenetic information allows us to test this hypothesis directly.

Finally, incorporation of phylogeny reduces the effects of unmeasured confounding variables. Such variables are particularly problematic when shared through common descent. Felsenstein[12] provided a hypothetical example to illustrate how ignoring phylogeny can lead to spurious results (his Figs. 5–7). We provide a real example in primates involving the relationship between body mass and group size,[2,29] which we also use to illustrate the methods discussed below. Conducting the analysis with species data points, there is a strong, positive relationship between group size and body mass in primates, with larger-bodied species living in larger

### Box 1. Implementing Phylogenetic Comparative Methods

Many computer programs have been developed to implement phylogenetic comparative methods. In searching for the latest programs, a good place to start is Joe Felsenstein's web page (http://evolution.genetics.washington.edu/phylip/software.html). The CAIC computer program,[53] which runs on a Macintosh, calculates contrasts for continuous traits and also allows analysis of mixed data using the BRUNCH algorithm (see Box 3). The latest version of CAIC (v. 2.6.2) is available from http://www.bio.ic.ac.uk/evolve/software/caic/index.html. Options for non-Macintosh platforms include PDAP[44] (information available at: http://www.wisc.edu/zoology/faculty/fac/Garland/PDAP.html), which runs on a PC, and COMPARE (http://darkwing.uoregon.edu/~compare4/), which is web-based and therefore runs on multiple platforms. Finally, PHYLIP also runs on a wide range of machines, and it provides tree-building methods (http://evolution.genetics.washington.edu/phylip.html).

Readers may find a variety of other methods useful for particular questions.[10] Phylogenetic autocorrelation methods can be used to assess whether a trait is correlated with phylogeny and can generate "phylogeny-free" values for comparative studies.[92,108] This approach can be implemented in the computer package COMPARE or by using the program Phylogenetic Autocorrelation (ftp://ftp.math.utk.edu/pub/luh/PA.hqx). Grafen's[54] Phylogenetic Regression program may be useful for some comparative studies (http://users.ox.ac.uk/~grafen/phylo/). Moreover, similar approaches have been developed.[15,18,67] Pagel's program Discrete[66] can be used to test for correlations among discrete variables. Finally, new methods of reconstructing ancestral character states with confidence limits, for example, using maximum likelihood, can be implemented using several programs, among them COMPARE, PDAP, or Schluter's[81] program ANCML, available from http://www.zoology.ubc.ca/~schluter/ancml.html).

All of these methods require as input a phylogenetic hypothesis. At present, many consider the composite estimate of phylogeny given by Purvis[26,55] to be the best option for primate comparative studies. One important advantage of this primate "supertree"[109] is that branch length estimates are provided, a requirement of many methods. For 203 species of primates in the Corbet and Hill[110] primate taxomony, the Purvis phylogeny has 160 resolved nodes. More general information on supertrees is available.[109,111] Web sites for recent phylogenies are TreeBase (http://www.herbaria.harvard.edu/treebase/) and the Tree of Life (http://phylogeny.arizona.edu/tree/phylogeny.html). Sequence information for primates and other species can be obtained from GenBank (http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html). The datasets used in this paper, along with links to many of the web sites provided here, can be obtained from the first author's web site: http://faculty.virginia.edu/charlienunn.

groups ($b = 0.326$, $F_{1,105} = 29.4$, $P < 0.0001$; data from Nunn and Barton[30] and Smith and Jungers[31]). However, this relationship is confounded by activity period and substrate use, as shown in Figure 1. Thus, within activity-substrate categories, no relationship is found between female body mass and group size (nocturnal species: $F_{1,10} = 0.10$, $P = 0.76$; diurnal-arboreal species: $F_{1,70} = 0.30$, $P = 0.58$; diurnal-terrestrial species: $F_{1,21} = 0.32$, $P = 0.58$; data on activity period and substrate use updated from Nunn and van Schaik[29]).

A plausible adaptive hypothesis for the pattern in Figure 1 is that diurnal primates reduce predation risk through gregariousness and therefore live in larger groups than do nocturnal species, which avoid predation through crypsis. Terrestrial species experience increased predation risk and are therefore larger in body size and live in even larger groups.[2,29] Hence, the true causal variables are likely to be activity period, substrate use, and both of these variables' interactions with predation risk. All of these variables tend to be shared through common descent (for example, all baboon species are terrestrial, have larger bodies, and probably experience greater risk of predation, yet these traits have not been acquired independently among these species). In fact, as we show below, there is no significant relationship between body mass and group size once phylogeny is taken into account. Because so many variables interact in behavioral, ecological, and morphological studies, confounding variables are a common problem in comparative biology.

Computer simulation studies have documented the magnitude of statistical errors that result when phylogenetic information is ignored.[32–37] The effects are staggering. A recent simulation study, for example, showed that Type I error rates (the probability of rejecting a true null hypothesis) can be as high as 44% when phylogeny is ignored, compared to an expected error rate of 5%.[37] In other words, a naïve comparative analysis, as in Figure 1, can be up to nine times more likely to detect a significant pattern when, in fact, no relationship exists. It is even theoretically possible for a nonphylogenetic analysis to indicate a statistical relationship that is in the opposite direction from the true evolutionary relationship.[38] Moreover,

**Computer simulation studies have documented the magnitude of statistical errors that result when phylogenetic information is ignored. The effects are staggering.**

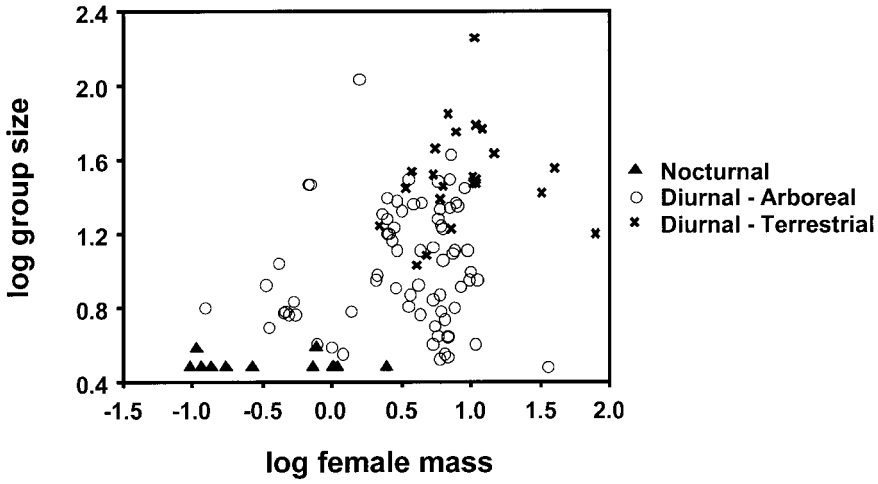



Figure 1. Association between body mass and group size in primates. Each data point represents a different species and is identified according to activity pattern and substrate use. A single three-value discrete variable is possible in this case, as all nocturnal primate species are arboreal.[29] We used female body mass rather than averaging sometimes markedly different values for males and females in dimorphic species.

the statistical power to detect associations is reduced when phylogeny is ignored.[32,33,37] This conclusion runs counter to the commonly expressed opinion that nonindependence is simply a "degrees of freedom" problem, one in which phylogenetic comparative methods are thought to reduce the number of data points for analysis and therefore result in lower statistical power. Although some phylogenetic comparative methods do appropriately result in fewer degrees of freedom, the most commonly used comparative method, independent contrasts, has the same degrees of freedom as a species analysis when the phylogeny is fully resolved.[52]

The computer simulations also illustrate an important assumption of nonphylogenetic studies. When using species values with no control for their historical relationships, the investigator assumes that the data points in the study are statistically independent. This is equivalent to assuming that the species are all equally related to one another and therefore linked by a "star phylogeny," with all branches emanating from a single ancestor.[12] This assumption is clearly invalid. Thus, it is not surprising that imperfect phylogenetic information often provides better statistical performance than ignoring phylogeny entirely.[33]

## PHYLOGENETIC COMPARATIVE METHODS: AN OVERVIEW

Comparative analysis requires information on phylogeny and a means to reconstruct evolutionary change.[10,12,14,39,40] Our confidence in a comparative test is strengthened when the phylogeny is well supported or when alternative models of character evolution, using several plausible phylogenies, give similar results.[10,41] In many cases, it is important that the phylogeny be reconstructed based on traits that are independent of the question at hand, although the effect of this should be considered on a case-by-case basis.[42] Many phylogenetic approaches to comparative testing also make assumptions regarding branch lengths, or the amount of time corresponding to each branch on the phylogeny.[12]

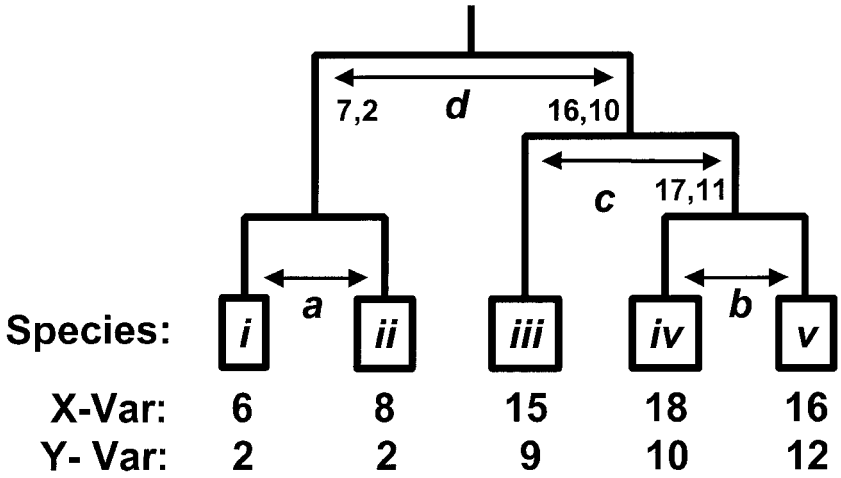
Although it may sound obvious, the first step in a comparative study is to state the hypothesis and its predictions. There is a good reason for emphasizing this step in comparative biology. Explicit formulation of predictions distinguishes one use of comparative methods, namely hypothesis testing, from another equally valid use of comparative methods to generate hypotheses.[7,10,11] The danger is that in exploratory analyses aimed at generating hypotheses, several comparative patterns are likely turn up significant and then inadvertently "become" *a priori* hypotheses, written up at a later stage as if the investigator had explicitly tested a hypothesis. Instead, if the study is exploratory, then the hypotheses that are formulated should be stated as such. These hypotheses can then be tested independently, either in comparative study of another appropriate group of organisms or by testing independent comparative, observational, and experimental predictions in the clade from which the hypothesis was generated.[43]

After formulating the hypothesis and collecting relevant data, the next step is to decide on the type of method to use. This is a critical step: it structures the overview of methods that follow because some phylogenetic comparative methods are appropriate for discrete data, others are appropriate for continuous data, and yet others are appropriate for a combination of discrete and continuous data. Continuous data can take any quantitative value, subject to measurement precision, and include variables such as body mass, group size, and longevity. Regression, correlation, and principal components analysis are statistical tests commonly used to examine relationships among continuous variables. Discrete data are those that have integer values. In phylogenetic comparative studies, these values are typically dichotomous, coded as 0 or 1, although more than two character states are possible. For example, most primate species can be classified as nocturnal (= 0) or diurnal (= 1). However, cathemerality could be added as an intermediate value to give three total character states, with adjustment of character states such that nocturnality = 0, cathemerality = 1 and diurnality = 2. Discrete data commonly are analyzed using tests of independence such as the Chi-square test, although different statistical tests are usually required when taking phylogeny into account.

A final situation concerns a mix of continuous and discrete data types. These data would normally be analyzed using analysis of variance (ANOVA) or, if there is more than one continuous variable, analysis of covariance (ANCOVA). Although some of these specific tests can be used in com-

## Box 2. Phylogenetically Independent Contrasts

Independent contrasts are calculated as differences between pairs of species, between a species value and an ancestral node, or between higher nodes. Differences are calculated beginning at the tips of the tree. In this example, contrast *a* is calculated as the difference between species *i* and *ii*, while contrast *b* is calculated as the difference between species *iv* and *v*. Contrasts then are calculated using higher nodes (that is, the "mixed" contrast *c* and internal contrast *d*), with values reconstructed as weighted averages of the descendent taxa (indicated on the phylogeny). Contrasts are thus calculated up the tree, maximizing the number of possible contrasts while using each branch no more than once.




With this phylogeny and data set, the unstandardized contrasts are:

| | contrast a | contrast b | contrast c | contrast d |
|---|---|---|---|---|
| X-variable | 2 | 2 | 2 | 9 |
| Y-variable | 0 | −2 | 2 | 8 |

The direction of subtraction for calculating contrasts is arbitrary. For plotting contrasts, however, some authors have recommended forcing the independent variable to be positive.[52] Thus, we forced the *X*-variable contrasts to be positive in this example, with the direction of subtraction retained for calculating the *Y*-variable contrasts. For example, contrast *d* is calculated as $16 - 7 = 9$ for *X*. Maintaining the direction of subtraction gives $10 - 2 = 8$ for *Y*. With five species, there are four contrasts on this fully resolved phylogeny. Methods for standardizing these contrasts (that is, incorporating branch lengths) are explained in the text and elsewhere.[10,12,52]

These contrasts are independent of one another and can be examined using standard statistical packages. Because the expected value of a contrast is 0, regressions and correlations must be forced through the origin.[10,52] Before forcing the intercept to 0, however, we recommend that this be tested statistically, as intercepts that differ significantly from 0 may indicate violations of the method or a nonlinear underlying relationship.[10] The statistical and evolutionary assumptions of contrasts should also be tested using established procedures.[10,52,53] Other assumptions of the method include correct topology, correct branch lengths, a Brownian motion model of evolution, and negligible intraspecific variation.[14]

parative tests, for example by generating null distributions using computer simulation,[44] alternative approaches based on independent contrasts are available.

This discussion brings up the issue of data collection and scoring of characters because discrete data often, but not always, represent an underlying continuous distribution of trait values in a discontinuous way.[44,45] Comparative tests using continuous measures are often better able to detect cross-species patterns because they provide more fine-grained variation to detect patterns.[44,46] Moreover, continuous variables are more likely to meet parametric statistical assumptions than are discrete categorizations. Whenever possible, then, continuous data should be used to test comparative predictions. In what follows, we consider methods that are appropriate for the different data types, concluding with a discussion of multivariate methods.

### All Continuous Data

When all variables are continuous, the preferred method is independent contrasts (Box 2). Contrasts are differences in trait values between species or higher nodes. As such, they represent independent evolutionary change since two species last shared a common ancestor and thus deal with the nonindependence of species values. Independent contrasts can be used explicitly to test hypotheses predicting correlated evolutionary change. Figure 2 provides an example using contrasts in basal metabolic rate and contrasts in body mass. This plot can therefore be interpreted as evolutionary increases in body mass are correlated with evolutionary increases in basal metabolic rate.

At first glance, treating differences in trait values (Box 2) as a measure of evolutionary change can be conceptu-

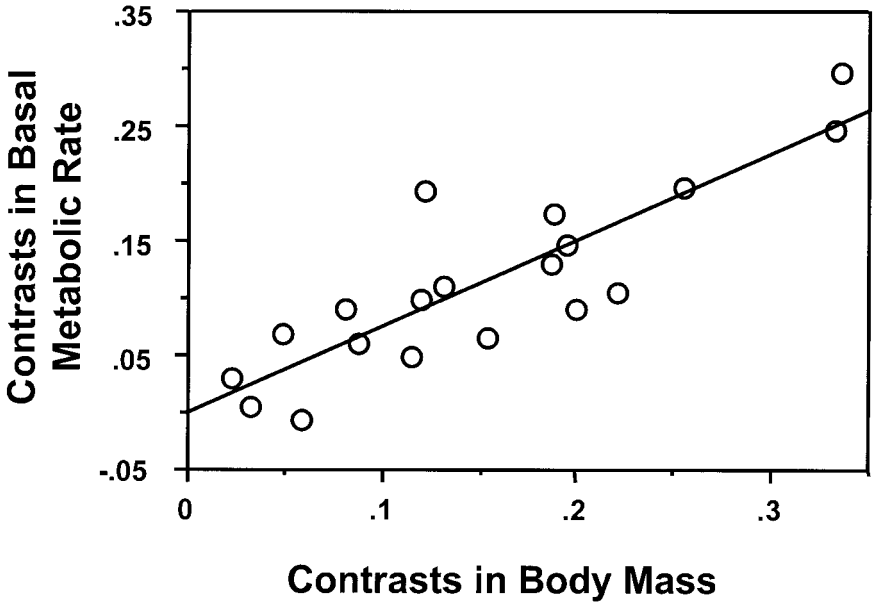



Figure 2. Association between basal metabolic rate and body mass in primates. Contrasts in basal metabolic rate are positively correlated with contrasts in mass ($r^2 = 0.92$, $t = 14.2$, $P < 0.0001$). The slope of the line and the deviations from it (residuals) can be interpreted in the conventional way. For example, Kleiber's Law[112] states that basal metabolic rate increases to the 0.75 power of mass, and the slope in this figure matches this predicted value perfectly (confidence interval: 0.64–0.86). Large residuals indicate evolutionary changes in basal metabolic rate that are either larger (positive residuals) or smaller (negative residuals) than predicted by changes in mass, and these deviations may be associated with specific ecological factors.[113] Basal metabolic rate data are those listed as meeting reliability criteria in Ross.[113]

ally difficult. It might seem more intuitive to reconstruct the ancestral nodes of a phylogeny and then, for each branch in the phylogeny, calculate change from deeper to shallower nodes or to species values. One problem with this approach, however, is that for a fully resolved phylogeny with $n$ species, there are $2n$-2 branches, therefore giving more degrees of freedom than there are species values. In other words, two nonindependent observations would be calculated for each node. By comparison, contrasts give $n$-1 degrees of freedom with a fully resolved tree. Moreover, only one degree of freedom is lost in regression and correlation analyses of contrasts because a contrasts plot does not have an intercept (see Box 2). Thus, a contrasts analysis loses two degrees of freedom, one for calculating contrasts and one for estimation of the slope, which is identical to analysis of species values. (In regression of species values, for example, one degree of freedom is lost for the slope estimate and one degree of freedom for the intercept). However, additional degrees of freedom may be lost in contrasts analysis when the phylogeny is incompletely resolved. This is true, for example, for "soft" polytomies[47] that reflect ignorance of the true branching pattern.[48,49]

The method of independent contrasts can also help control for confounding variables shared through common descent, although it cannot be assumed to do so completely. For example, when the relationship between group size and body mass is examined with contrasts (Fig. 3), the statistical results differ sharply from the strong and highly significant relationship calculated from species values ($b = 0.16$, $F_{1,94} = 2.94$, $P = 0.09$; see also Barton[50]). Visual inspection of the contrasts reveals that at least one data point, an outlier, may exert excessive leverage on the slope estimate and statistical results (Fig. 3). We hypothesized earlier that diurnality and terrestriality explain the statistical association in Figure 1. When we control for the effects of these variables by identifying and excluding contrasts with evolutionary transitions to diurnality and terrestriality, the body mass-group size relationship weakens further ($F_{1,77} = 1.42$, $P = 0.24$). Hence, this example shows that by examining evolutionary change, contrasts analysis helps deal with unmeasured confounding variables, but that consideration of these confounding variables when they are known and measurable further reduces their effects.[51]

**When there is only one evolutionary transition between ecological categories, the true sample size is effectively one, reducing large numbers of species values to a single degree of freedom. In human evolution, for example, the correlation between less-wooded habitats and bipedalism may be entirely dependent on one evolutionary transition at the root of the hominid clade. This is why it is difficult to make statistical inferences about the adaptive significance of bipedalism or, indeed, any unique hominid trait, perhaps explaining why "just so" stories abound.**

Another source of scatter arises from how the contrasts are standardized,

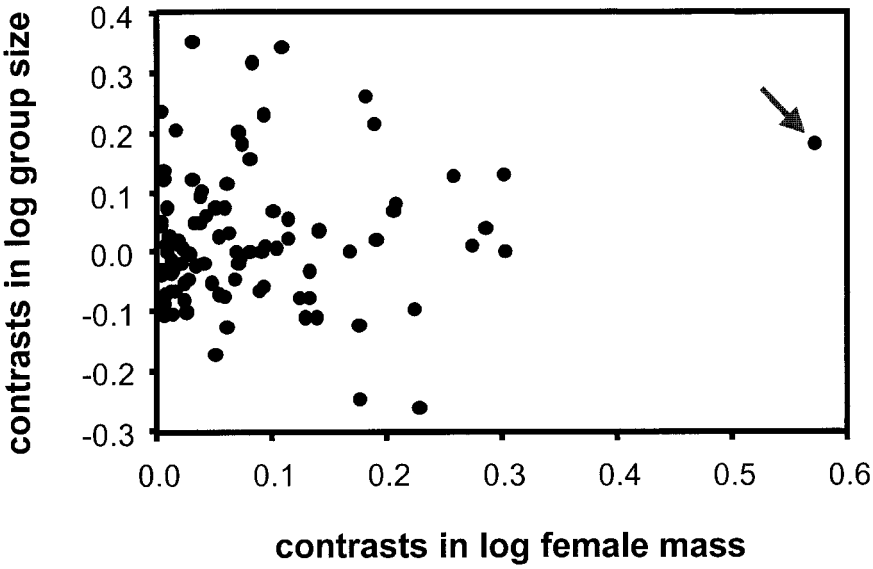



Figure 3. Association between body mass and group size using independent contrasts. When the data from Figure 1 are examined using independent contrasts, a nonsignificant positive slope is found (see text for details). This plot uses contrasts calculated assuming equal branch lengths (a "speciational" model), with the data log-transformed. The arrow points to the outlier discussed in the text.

which raises issues regarding the evolutionary model that implicitly underlies most phylogenetic comparative methods. Computer programs typically standardize contrasts by dividing each contrast by the sum of its branch lengths. These branches are further adjusted by increasing the lengths of branches deeper in the tree.[12] Dividing by branch length effectively changes the placement of the contrasts relative to the origin. For example, dividing $X$ and $Y$ contrasts by a larger value, representing longer branches, places the data point in a bivariate plot closer to the origin, giving it less statistical leverage on the overall relationship. The underlying assumption is that on longer branches more evolution is likely to take place, leading to greater change in $X$ and $Y$. These contrasts, therefore, would lead to problems of heteroscedasticity (different variances) if they were not standardized. Lengthening branches to deeper nodes also places these contrasts closer to the origin, effectively giving them less leverage because the line is forced through the origin, which may be appropriate if reconstruction of ancestral character states is less certain for these deeper nodes.[12] It should be noted, however, that such "reconstructions" are really weighted averages of descendent taxa, which are necessary to maintain the independence of the contrasts.

Hence, the main reasons to standardize independent contrasts involve the statistical problem of heteroscedasticity and reduced certainty of ancestral states. The usual means of standardization is based on a model of evolution known as Brownian motion.[12,45] This model assumes that for each branch on the tree evolutionary change is independent of previous change and drawn from a normal distribution with a mean of zero, such that positive and negative changes are equally likely. The variance of this distribution is proportional to the length of the branch on which change took place, allowing standardization by Brownian motion to control for greater change on longer branches.

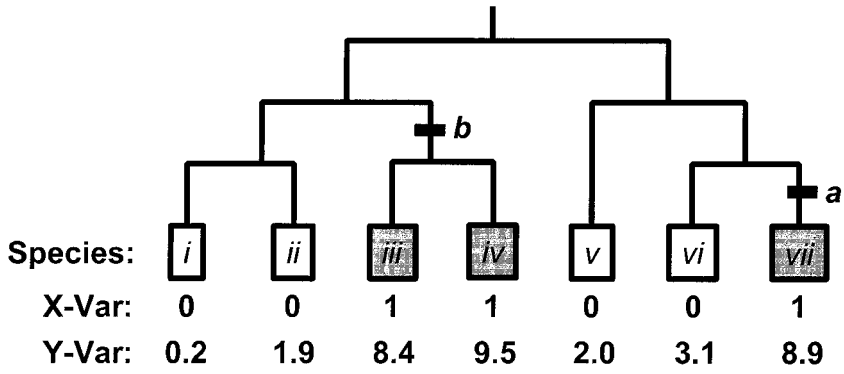
We can empirically test the assumptions to determine whether contrasts are properly standardized under different transformations of the data and branch lengths.[52,53] This is similar to testing the assumptions of parametric statistical tests such as the normality assumption, in that appropriate transformation of the data and branch lengths can often be used to meet the assumptions.[35,36] Common transformations include taking logarithms of the data or branch lengths and seemingly more radical transformations of branch lengths, such as the assumption of "speciational" change (branch lengths set equal throughout the tree[44]), or transformation of branches as a function of the number of species below each node.[54] In the contrasts plot in Figure 3, for example, we found that log-transformed data and a speciational model of evolution best met the assumptions of contrasts. To illustrate the effect of using another assumption, we repeated the analysis with the raw data and a "gradual" evolutionary model (branch lengths set proportional to time since divergence[55]). These assumptions result in greater scatter of some points (Fig. 4), but the conclusion that no relationship exists is maintained.

## Combination of Continuous and Discrete Data

Comparative analysis commonly involves a combination of continuous and discrete data. A good example is a comparative study of home-range size in carnivores and ungulates by Garland et al.[44] that tested the hypothesis that species of carnivores, as predators, require larger ranging areas than do closely related nonpredatory mammals. These authors found a highly significant difference when using species values and controlling for body mass by analysis of covariance ($F_{1,46} = 23.97$, $P < 0.001$), but the difference in home range size was not significant when phylogeny was incorporated into the analysis ($P = 0.15$ to 0.29, range based on simulation tests using different assumptions). This example therefore illustrates the fundamental difficulty in analyzing discrete data: When there is only one evolutionary transition between ecological categories, the true sample size is effectively one, reducing large numbers of species values to a single degree of freedom. In human evolution, for example, the correlation between less-wooded habitats and bipedalism may be entirely dependent on one evolutionary transition at the root of the hominid clade. This is why it is difficult to make statistical inferences

### Box 3. Combinations of Discrete and Continuous Data: The BRUNCH Algorithm

The CAIC computer program can be used to examine combinations of continuous and discrete data using the BRUNCH algorithm.[53] In the following example, there are two reconstructed transitions, or contrasts, in the discrete variable, labeled *a* and *b*. Contrasts in this variable are forced to be positive; thus, for contrast *a*, the trait value for species *vi* (= 0) is subtracted from species *vii* (= 1), giving an unstandardized contrast value of +1.




Contrasts in the continuous variable are calculated only over branches with transitions in the discrete variable. The direction of subtraction is retained for the continuous (dependent) variable, giving an unstandardized contrast of 5.8 for the *vii* − *vi* contrast. A similar procedure is followed for the other discrete transition, at *b*, which involves a higher node.

The hypothesis to test in this case is whether transitions from 0 to 1 result in consistent positive or negative changes in the dependent variable, as indicated by their contrasts. The text describes two statistical tests, one based on the nonparametric sign test and the other based on a *t*-test, to determine whether the mean contrast differs significantly from zero. There are too few contrasts to obtain a significant result with a sign test in this case, although both contrasts are positive (at *a*, after standardization assuming equal branch lengths, $\Delta Y = 2.9$, at *b*, $\Delta Y = 3.23$). A *t*-test, however, gives a significant result (mean of *a* and *b* = 3.065, $t = 18.84$, $P = 0.03$).

about the adaptive significance of bipedalism or, indeed, any unique hominid trait, perhaps explaining why "just so" stories abound. There are ways to examine continuous traits statistically over a single transition in a discrete character, but extremely large differences may be needed to obtain significant results.[44]

In many cases involving continuous and discrete data, however, more than one transition exists between classes of discrete data. Each of these transitions potentially provides an independent data point, and the average effect can be tested statistically. One commonly used approach in this situation is a variant on independent contrasts known as BRUNCH.[53] This method identifies the discrete trait as the independent variable. It then calculates contrasts with the direction of subtraction set such that the discrete variable is positive, retaining the direction of subtraction for contrasts in the continuous variable (Box 3). The prediction to test is whether contrasts in the continuous dependent variable are consistently positive or negative over evolutionary increases in the discrete independent variable.

For example, we proposed earlier that transitions to diurnality and terrestriality result in increased body mass and group size, leading to a spurious relationship between these two continuous variables. With a three-part discrete scoring of nocturnal species (= 1), diurnal-arboreal species (= 2), and diurnal-terrestrial species (= 3), we used the CAIC computer program[53] to examine transitions in body mass and group size over evolutionary shifts to diurnality and terrestriality. These discrete values can therefore be considered a ranked score relative to our expectations of group size. We identified a total of 17 contrasts with reconstructed changes in the discrete variable. The direction of subtraction for both variables is set such that the discrete variable is positive (for example, subtracting a nocturnal clade's value from a diurnal-arboreal clade's value). The expected mean of contrasts in the dependent (continuous) variable is zero, but we find a mean increase in body mass of 0.10 and a mean increase in group size of 0.11 (using log-transformed values). These values indicate that evolutionary transitions to diurnality and terrestriality are associated with evolutionary increases in body mass and group size. We need to test, however, whether these means differ significantly from expectations under the null hypothesis of no consistent pattern of change (that is, a mean of 0).

The statistical test typically is conducted in one of two ways.[53] First, a nonparametric test, the sign test, can be used to determine if observed increases are more common than expected by chance. In this case, the null hypothesis is an equal number of gains and losses distributed binomially.[56] The null hypothesis can be rejected: 15 of 17 contrasts in body mass are positive ($P < 0.02$), and 13 of 17 contrasts in group size are positive ($P < 0.05$). Second, the observed means can be tested versus the null hypothesis of 0 using a t-test. Again, the null hypothesis can be rejected (group size: $t = 3.12$, $P = 0.007$; female body mass: $t = 3.40$, $P = 0.004$). In choosing among these tests, the nonparametric test has the advantage that

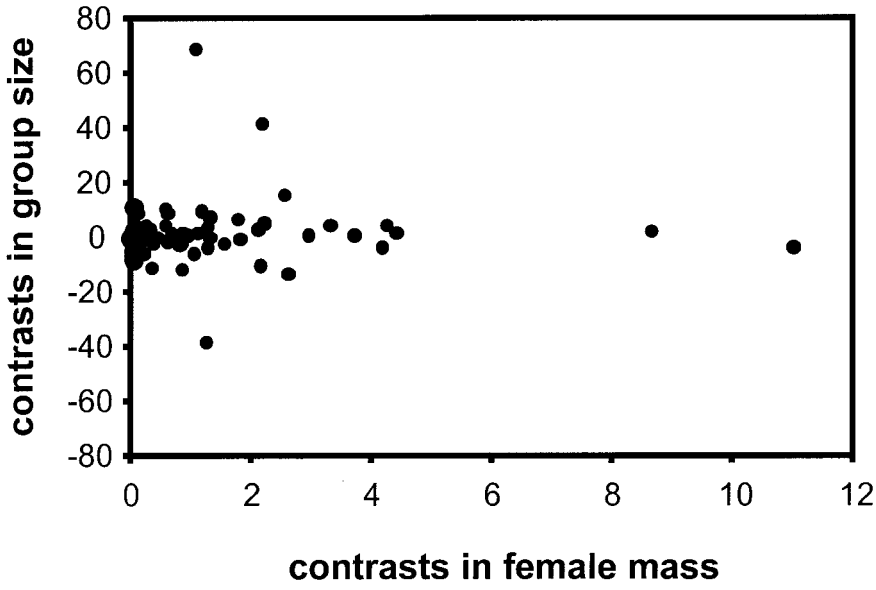



Figure 4. Improperly standardized contrasts. Using the same data as in Figures 1 and 3, contrasts were calculated assuming that branch lengths were proportional to time[55] and using raw (unlogged) data. The scales of the X and Y-axes therefore differ from those in Figure 3. These changes resulted in a different spatial distribution of the data points, with more outliers, but the general conclusion remains: There is no relationship between body mass and group size once phylogeny is taken into account. This example highlights the need to test the assumptions of independent contrasts.[52,53]

fewer assumptions are made. However, at least six gains are required to show significance.[56] By comparison, the t-test can work with smaller samples and may have higher statistical power, although a low variance among contrasts can lead to significant results with surprisingly small *P*-values (for an example, see Box 3).

### Discrete Variables

Ridley[16] provided the first method for explicitly incorporating phylogeny in the study of discrete traits. His method has been used to address questions in primate biology such as the association between breeding seasonality and single-male groups.[59] Other methods, however, have been used more commonly since Ridley's pioneering research.

Maddison's[13] concentrated changes test is probably the most commonly used method for testing statistical associations between two discrete variables. It is described thoroughly elsewhere[10,13] and is implemented in the computer program MacClade.[60] Given a phylogeny and the distribution of two traits, one identified as independent and the other dependent, the method calculates the probability of getting a certain number of gains in the dependent trait on branches with a particular character state ("distinguished branches") in the independent trait (Fig. 5). This probability, which serves as a *P*-value for statistically testing the association of the two characters, can be solved analytically for small numbers of species, as in Figure 5. For larger phylogenies and more complicated hypotheses of trait evolution, a simulation procedure is easily implemented in MacClade.[60] The concentrated changes test can potentially identify causality; that is, the temporal order of trait changes. However, this requires that the two traits are not reconstructed as changing simultaneously. It also requires accurate estimation of ancestral character states.[61]

To illustrate how the concentrated changes test is used, we return to the preceding example, focusing on the two discrete variables of activity period and substrate use. No nocturnal primate species are classified as terrestrial,[29] explaining why we combined these two discrete variables into one variable with three, rather than four, character states (nocturnal, diurnal-arboreal, and diurnal-terrestrial; see Fig. 1). If each species is treated as independent, as in the "old" comparative approach, activity period and terrestriality are significantly associated (Chi-square = 8.20, degrees of freedom = 1, $P = 0.004$; 188 species, data updated from Nunn and van Schaik[29]). But this test ignores the fact that activity period is reconstructed as having changed only four times in the evolutionary history of primates. Thus, we used the concentrated changes method to test whether terrestrial substrate use is significantly concentrated on branches of the tree characterized by diurnality.

**We agree that more progress is possible if hypotheses are tested with variation at all possible levels, including variation among species, populations, groups, and individuals. In general, however, intraspecific variation tends to obscure comparative trends rather than to create spurious ones. By this logic, intraspecific variation may explain why some results are nonsignificant, but it is not clearly a legitimate criticism when a well-supported pattern is found.**

To implement the concentrated changes test, we first mapped the two characters, activity period and sub-

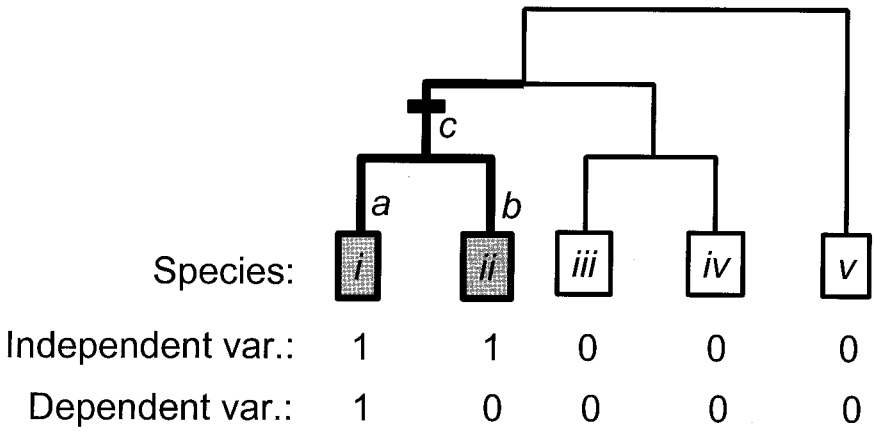



Figure 5. Maddison's concentrated changes test. This test[13] begins by identifying traits as independent (X) and dependent (Y), and the character states of interest for each trait (in this case, 1 for both traits). With this information, one calculates the probability of a particular number of gains or losses in the "distinguished branches" of the tree, those with the character state of interest in the independent variable. In this simple example, we are interested in the probability of one gain in character state 1 for the dependent variable on branches reconstructed as having character state 1 in the independent variable; that is, branches a, b, and c. To meet this condition, character state 1 in the dependent variable could therefore evolve on any of these three branches. Because this tree has eight total branches, excluding the root, the probability is 3/8, or P = 0.375.

strate use, onto a phylogeny. MacClade requires a fully bifurcating tree to implement this test, yet the phylogeny most commonly used to conduct comparative studies of primates has several unresolved nodes.[53] These polytomies were therefore randomly resolved using MacClade, although we created ten different random resolutions to examine the effect of different tree topologies on the results. To apply these tests for each tree, we set the distinguished branches as diurnal (character state = 1), because we were interested in testing whether terrestriality is significantly associated with this independent character state. We ran 1,000 simulations, with nocturnal as the ancestral character state, and simulated 7 gains of the dependent variable (that is, the number of gains reconstructed in MacClade). We are not concerned with losses of terrestriality, so the losses dialog box was not altered.

From these simulations, we calculated the probability that seven gains occur on the diurnal branches. With the different phylogenies, the range of results was remarkably narrow but not significant ($P$ = 0.22 to 0.26). Thus, in comparison to nonphylogenetic analysis, we found that diurnality and terrestriality are not statistically associated. For nonsignificant statistical results, it is useful to consider how many gains of the independent variable are required to obtain significance, which is effectively a form of power analysis. In this case, the answer turns out to be about 14 gains, which we determined by running the simulations with progressively more gains in the dependent variable.

A recent computer simulation study has shown that the concentrated changes test has acceptable Type I error rates,[62] but the method has other shortcomings. For example, it requires a bifurcating tree and does not incorporate branch length information (see Sillén-Tullberg[63] and Werdelin and Tullberg[64]). Although these problems are minor relative to the benefits, including ease of implementation, several other methodological options are available for discrete traits. For example, Read and Nee[65] provide an alternative method that avoids many of the assumptions involved in the analysis of discrete comparative data, including the model of character evolution. This method examines pair-wise comparisons in taxa that differ in the independent variable such that comparisons are genuinely independent of one another. Their approach makes the fewest possible assumptions about evolutionary change. At the other extreme is a method developed by Pagel,[66] which explicitly models trait evolution using a Markov model. Statistical significance is determined using maximum likelihood methods. Pagel's method also makes use of branch lengths and, in theory, can identify causality even when two traits are reconstructed as changing simultaneously. A related version for analysis of continuous characters[15,67] is available.

### Multivariate Analyses

Contrasts can be examined easily using multivariate statistical methods, including principal components analysis.[68] The CAIC computer program[53] allows a user to select multiple variables for calculating contrasts. Even in multivariate analyses, however, the regression line is constrained to pass through the origin.[52,68] An important statistical issue in multivariate analyses is a possible reduction in sample sizes, as only species with information on all variables can be included in the analysis.[53] Thus, variables should be chosen wisely to maximize the number of contrasts.

**Computer simulation has shown that incorrectly specifying the topology or branch lengths leads to higher Type I error rates. These simulations have also shown, however, that failure to incorporate phylogeny, and thereby assuming a star phylogeny, produces an even less desirable outcome in that Type I error rates are usually increased.**

For combined discrete and continuous data, several multivariate approaches are possible. First, one can treat an independent discrete variable as continuous and enter it along with the other variables. In multiple regression analysis of the example given earlier, activity period and substrate use account for significant variance in group size when treated as continuous variables ($b = 0.33$, $F_{1,92} = 4.82$, $P = 0.03$ and $b = 0.23$, $F = 8.81$, $P = 0.004$, respectively), but female mass is not significant ($b = -0.006$, $F = 0.003$, $P = 0.96$). These results therefore confirm that variables other than body mass account for variation in group size. Second, correlations among continuous characters can be investigated by examining only those contrasts with corresponding changes in discrete variables. Such a test would identify whether alternative explanations can explain significant results in a BRUNCH analysis. Thus, focusing only on contrasts in the three-character activity period-substrate codes from the earlier example, body mass and group size are not statistically significant ($F_{1,16} = 2.61$, $P = 0.13$), suggesting that activity period and substrate use are responsible for the comparative patterns. Finally, Grafen's[54] "phylogenetic regression" can be used to examine discrete and continuous independent variables in a multivariate context.

Approaches like analysis of covariance are available for three or more variables, for example by using computer simulation methods developed by Garland and colleagues.[44] These programs can be implemented in the computer package PDAP, although variants on the basic method that avoid simulations are also possible.[44] Simulation-generated null distributions also provide a means to incorporate phylogeny in novel statistical tests of comparative hypotheses.[69]

At present, however, it is not possible to analyze associations among multiple discrete traits in a phylogenetic context. One approach might be to examine subsets of the phylogeny where some potentially confounding variable is absent, but this would tend to reduce statistical power.

## COMMON MISUNDERSTANDINGS OF COMPARATIVE METHODS

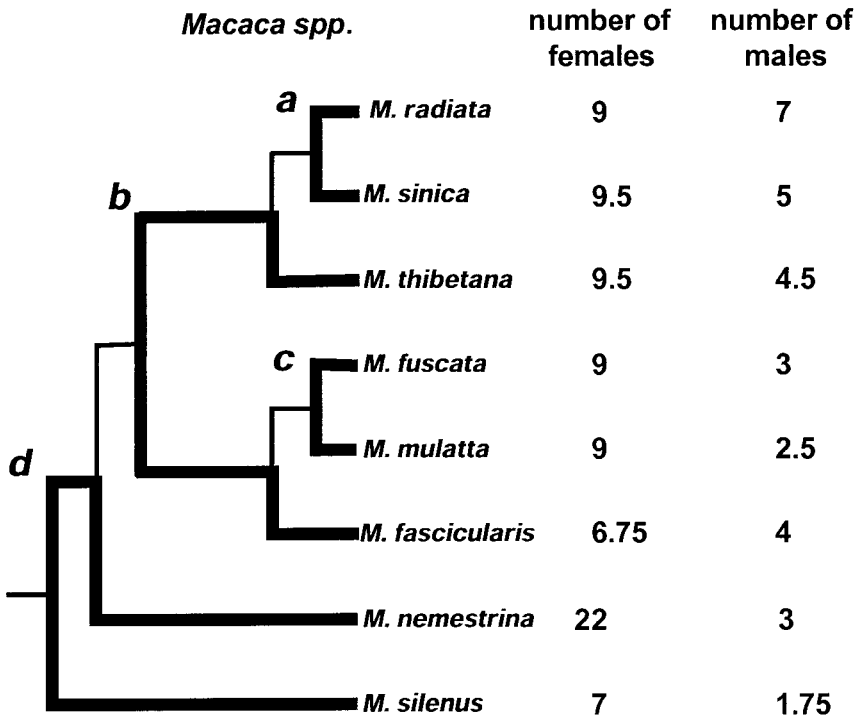
Benton[70] identified four types of critics of large-scale paleontological studies. Not surprisingly, the same types of criticisms often are leveled at comparative biologists who examine broad evolutionary patterns. Benton's[70] categories of critics, appropriately altered for comparative studies, include the proofreader, who discovers errors in the database used by a comparative study and therefore believes there are serious flaws in the conclusions; the trades unionist, who aims to protect his or her field of research and cannot bear to have an "outsider" use data secondhand; the Luddite, who cannot stand new methods and the idea of taking vast datasets and reducing them to simple graphs; and the Utopian, who suggests that we have inadequate information to conduct comparative studies, rendering the whole process pointless, although eventually such information will become available.

**. . . there is often some reluctance among biological anthropologists to rely on ancestral character state reconstruction. Incorrectly specifying ancestral states can obviously result in erroneous conclusions when examining the sequence of evolutionary events, but error in reconstructing ancestral states seems to be less of a problem for contrasts. This difference arises because contrasts involve species differences, not the actual reconstructed nodes.**

Although there is some validity to these caricatures, and some of the criticisms should be taken seriously, Benton[70] has excellent counter-arguments to each of these critics that also apply to comparative biology. Database errors noted by proofreaders, for example, are a serious matter, but they rarely alter conclusions. Following correction, in fact, the strength of patterns often improves, such that regular "sloppiness in big data sets actually strengthens the case for . . . patterns"[70] (p. 256). Similarly, some databases are too large for any one person to compile from first-hand data (the trades unionist), and we will never have sufficient data to cover all species equally (the Utopian), especially given current rates of extinction.

A final issue relating to the trade unionist should be considered. Some biologists, particularly those with active field research, often claim that the comparative approach ignores intraspecific variation, and that this somehow limits the conclusions of a comparative study when intraspecific variation is unknown. We agree that more progress is possible if hypotheses are tested with variation at all possible levels,[71,72] including variation among species, populations, groups, and individuals. In general, however, intraspecific variation tends to obscure comparative trends rather than to create spurious ones.[11] By this logic, intraspecific variation may explain why some results are nonsignificant, but it is not clearly a legitimate criticism when a well-supported pattern is found.

The development of phylogenetic comparative methods has given some of these critics new ammunition to fire at the comparative approach, including some criticisms by biological anthropologists.[20,21,73] We consider some of these issues in what follows (see also Purvis and Webster[26]). Our goal is to clarify misconceptions while




| Pairwise contrast | Females | Males |
|---|---|---|
| *a* | 0.5 | -2.0 |
| *b* | 2.75 | 0.5 |
| *c* | 0 | 0.5 |
| *d* | 15 | 1.25 |

Figure 6. Pair-wise comparisons. When the phylogeny is insufficiently resolved, or when less confidence can be placed in ancestral state reconstructions, pair-wise comparisons are sometimes used.[77] The method is illustrated here with the example of the correlation between male and female number.[46,78] With the key requirements that internal nodes are not reconstructed and branches are not used more than once, four pair-wise comparisons can be extracted from this phylogeny. It is often easiest to calculate pair-wise comparisons by hand, as computer programs will identify "mixed" contrasts (a species contrasted with a higher node; for example, *Macaca fascicularis* contrasted with the reconstructed value for *M. mulatta* and *M. fuscata*).

providing further guidelines for implementation of these methods, especially those based on independent contrasts.

## Phylogenetic Uncertainty: Tree Topology and Branch Lengths

Proofreaders find errors in phylogenetic relationships, using these errors to weaken the conclusions of a particular study, while Utopians argue that our knowledge of phylogeny, including both topology and branch lengths, is imperfect. Computer simulation has shown that incorrectly specifying the topology or branch lengths leads to higher Type I error rates. These simulations have also shown, however, that failure to incorporate phylogeny, and thereby assuming a star phylogeny, produces an even less desirable outcome in that Type I error rates are usually increased.[32–37] Branch lengths may be particularly prone to error, but it is important to remember that branch lengths mainly address the statistical issue of heteroscedasticity. With appropriate transformation of data and branch lengths[35,36,52] and analysis of outliers,[30] spurious results are less likely.

Sensitivity tests can be used to examine the effects of different branch length assumptions or topologies. A more radical approach uses computer simulation to generate a large number of random trees.[74,75] If the results remain consistent on a large proportion of these trees, then greater confidence can be placed in the conclusions, although results should be interpreted with caution.[41,76]

Another solution is to perform pair-wise comparisons that use information only from extant taxa (Fig. 6), thereby avoiding ancestral state reconstruction.[65,77–79] The problem with pair-wise comparisons is that many potential comparisons are not used, thereby reducing statistical power. In Figure 6, for example, only four pair-wise comparisons are available, whereas independent contrasts would extract seven values for analysis. The method of pair-wise comparisons is preferred for particular questions, however, such as when branch length information is critical to the hypothesis at hand.[80]

As a related issue, there is often some reluctance among biological anthropologists to rely on ancestral character state reconstruction. Incorrectly specifying ancestral states can obviously result in erroneous conclusions when examining the sequence of evolutionary events,[39] but error in reconstructing ancestral states seems to be less of a problem for contrasts.[81] This difference arises because contrasts involve species differences, not the actual reconstructed nodes. Thus, errors in some ancestral states will most likely affect the magnitudes of species differences but not their sign, and only internal contrasts will be affected. This issue has been examined empirically using a known viral phylogeny, which was experimentally generated in the laboratory so that ancestral states were known.[82] The results showed that reconstructed values failed to match known values at

internal nodes by wide margins for most traits, but contrasts analysis generally identified the evolutionary correlations among these traits.

## Phylogenetic Uncertainty: The Underlying Evolutionary Model

To standardize contrasts, the method of independent contrasts relies on an underlying model of trait evolution. In most cases, this Brownian motion model is likely to be an oversimplification. As noted, however, altering this model typically has little effect on the conclusions.

Recent work by Price[51] and Harvey and Rambaut[83] has examined a radically different evolutionary model that may have implications for the method of independent contrasts. Whereas the contrasts method assumes that traits evolve in a correlated fashion down a phylogeny, this new "niche model" assumes that the traits are correlated in niche space, with openings of new niches and invasion by similar species leading to phylogenetic topology (Fig. 7). This new model therefore focuses on the correlation of traits in niche space, not on the correlation of evolutionary change. As might be expected, then, use of species values gives expected Type I error rates in simulation of the niche model, whereas contrasts gives higher than expected error rates.[83]

The niche model makes some surprising assumptions in that it plays down the importance of phylogenetic history, assumes that species values are constant through time, and ignores correlated evolutionary change. Important questions remain, but several lines of evidence suggest that even this model will not lead to an abandonment of contrasts-based approaches. First, the correlation between species and contrasts values should differ if the niche model is correct, with lower correlations arising from contrasts. When correlations from species values and contrasts are compared, however, they are remarkably similar.[24,51] Second, a simulation study that used diagnostic statistics such as those discussed earlier not only detected violations of the Brownian motion model,[83] but found that increased Type I error rates arise through occasional outliers with high statistical leverage (P. Harvey and A. Rambaut, personal communications). These empirical results therefore suggest that diagnostic tests and careful exclusion of outliers may improve the statistical properties of contrasts when the underlying evolutionary model is unknown. Finally, if we are explicitly interested in correlated evolution, which is how adaptive hypotheses are often framed, then independent contrasts is the method to use, as contrasts represent evolutionary change.

One benefit arising from research on contrasts methodology is that biologists have developed a variety of underlying evolutionary models such as the niche model. At present, however, new evolutionary models are unlikely to lead to abandonment of independent contrasts. Accepting that the niche model reflects valid concerns[83] does not preclude the use of independent contrasts.[84–86] Methods based on maximum likelihood may provide a useful approach to comparative data generated under the niche model.[83,87]

**One benefit arising from research on contrasts methodology is that biologists have developed a variety of underlying evolutionary models such as the niche model. At present, however, new evolutionary models are unlikely to lead to abandonment of independent contrasts. Accepting that the niche model reflects valid concerns does not preclude the use of independent contrasts.**

## Statistical Power Is Too Low

Many researchers seem to believe that phylogenetic comparative methods, including independent contrasts, reduce the degrees of freedom, thereby eliminating data points for analysis.[20,21] Moreover, we have heard researchers say, "I ran the analysis using species values and contrasts, but got the same answer for both. I am therefore going to use the species values." The reasoning seems to involve the sentiment that phylogenetic comparative methods result in fewer degrees of freedom.

As noted earlier, however, a contrasts analysis with a fully resolved phylogeny has the same degrees of freedom as does analysis of species values.[52] While some phylogenetic comparative methods, most notably methods dealing with discrete data, have fewer degrees of freedom, we think that the reduction in degrees of freedom in such cases accurately represents the statistical realities of these evolutionary analyses.[13,16,65] More confidence should therefore be placed in the phylogenetic results.

Some biological anthropologists have noted that phylogenetic comparative methods ignore stabilizing selection.[20] It is true that most methods focus on evolutionary change to test hypotheses. In part, this is because correlated change is more convincing than correlated stasis, and it is unclear how to demonstrate "independence" in the case of stabilizing selection. Some more recent methods, however, have been developed to incorporate the effects of stabilizing selection.[87–89]

Other researchers have noted that they are not attempting to understand the evolutionary basis of some trait, but rather to describe the patterns among species. Examples of this from morphological studies are particularly common, including reconstruction of an extinct taxon's body size using morphological variation in extant taxa.[90] Any statistical results, however, will be affected by the nonindependence of species data points. For example, Garland et al.'s[44] analysis of carnivore and ungulate home ranges, which might be used to infer the behavior of a fossil taxon, was essentially a descriptive study. However,

as shown earlier, the statistical results differed greatly when using nonphylogenetic statistical tests versus controlling for phylogeny. Likewise, it might seem that the analysis of group size and body mass in Figure 1 could be used to reconstruct the behavior of extinct primates: A large-bodied species is more likely to have lived in a large group. A more valid approach, however, would reconstruct group size based on knowable features of the fossil taxon with statistical linkage to group size, such as activity period and substrate use, possibly with body mass as a covariate in the analysis.[29]

We suspect that most, if not all, cases where investigators claim to examine trait associations at a descriptive level, an underlying evolutionary question is implicit in the analysis. Moreover, these analyses are likely to be plagued by spurious statistical results due to sampling of a small number of hierarchically structured taxa, including multiple exemplars of closely related and similar species. Only phylogeny-based methods can be assured of providing independent data points for any sort of statistical analysis. Efforts to avoid these methods are statistically unjustified.

### Traits Vary in How Strongly They Correlate With Phylogeny

Recent interest has emerged in diagnostic tests that examine whether phylogeny "needs" to be incorporated in a comparative study.[91] The basic approach is to determine whether the continuous traits in question are significantly correlated with phylogeny.[92–96] If so, then a phylogenetic comparative method should be used; otherwise, it has been proposed, the investigator can treat species values as if they were independent.[94,96]

Although we are skeptical that any trait is truly uncorrelated with phylogeny, traits with large amounts of intraspecific variation, such as group size, may be diagnosed as having low phylogenetic correlations.[94] Two important issues, however, have yet to be fully addressed regarding these tests. First, the statistical effects of using contrasts analysis when species values are independent have not been fully investigated, although at least one study suggested that contrasts give higher Type I error rates in this situation.[93] Second, the diagnostics rely on acceptance of the null hypothesis that there is no correlation with phylogeny, rather than rejection of a hypothesis. Other tests of statistical assumptions have this characteristic, but the consequences of incorrectly accepting the null hypothesis in comparative tests may be more severe (for example, a spurious relationship, as in Fig. 1). Thus, we recommend that results from contrasts be given greatest confidence, although it is often useful to compare the results to those obtained from species values.

**Some biological anthropologists have noted that phylogenetic comparative methods ignore stabilizing selection. It is true that most methods focus on evolutionary change to test hypotheses. In part, this is because evolutionary change is more clearly independent, whereas it is unclear how to demonstrate "independence" in the case of stabilizing selection, or stasis.**

This discussion also brings up the issue of intraspecific variation and measurement error more generally, which has recently become an issue with independent contrasts[24] and may be a particular problem with behavioral data. In effect, intraspecific variation adds a "burst" of evolution onto the terminal tips across the tree, which may therefore violate the assumptions of independent contrasts (C. Janson, personal communication). Moreover, contrasts from tips of the tree may be more strongly affected by sampling and measurement error than are deeper contrasts.[24,26,30] One solution is to exclude these sister-taxa contrasts on the tips of the tree, but this is directly opposite to suggestions by advocates of pair-wise comparisons (Fig. 6). Further research is needed to address this issue, but the simplest solution may involve appropriate transformation of branch lengths[24] (for example, by lengthening terminal branches).

### Contrasts Are Inappropriate for Allometric Studies

It is not always appreciated that inference about scaling relationships is simply a specific case of the general problem of how two variables evolve together. Interspecific allometry, as distinct from growth allometry, is, or should be, the study of the quantitative change in $Y$ with change in $X$. As with the detection of correlated evolution, determination of the precise form of an allometric relationship must exclude phylogeny-based confounding variables.

We think that three general reasons explain why many biologists incorrectly argue against using contrasts for allometric questions. First, it is not immediately clear to some investigators how slope estimates from contrasts, which represent evolutionary change, relate to allometry, which is usually thought of as the relationship among species values at the tips of the tree. In fact, statistical measures of association calculated from contrasts and species values are equivalent, as based on analytical[10,73,97] and empirical investigation.[24,51] Thus, contrasts are fully appropriate for estimating allometric slopes, as discussed in detail by Harvey and Pagel.[10] Also, different methods of slope estimation, including least squares, major axis, and reduced major axis, can be used with contrasts.[10,30,52,98]

Second, it is sometimes argued that the allometric coefficient (the intercept, or elevation, in a species plot) cannot be determined using contrasts,

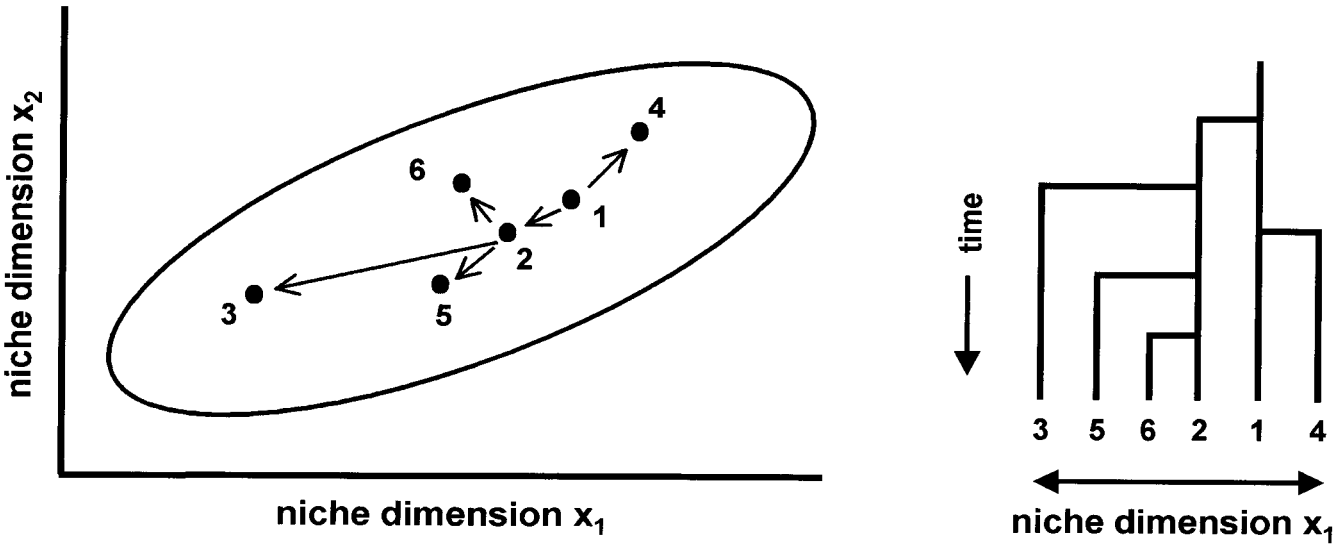



Figure 7. An alternative evolutionary model. The "niche" model assumes that two or more traits ($X_1$ and $X_2$) are correlated in niche space. The niche is originally filled by one species, labeled 1. Then, as subsequent niches open randomly in niche space, the species closest in this bivariate space undergoes a speciation event to fill the new niche. Thus, niche 1 is filled by species 2, niche 3 by species 2 (because species 5 and 6 have not yet arisen), and so on. In this model, species maintain "trunk" lineages that form the material for new lineages in adjacent niches later in time.

as a regression of contrasts is constrained to pass through the origin. While this is technically true,[67] it is possible to use the reconstructed values of the root node to position the regression line and calculate the intercept.[40] Confidence intervals can be placed on ancestral estimates, although some recent studies have suggested that such intervals can be wide.[40,81,91,99] In any case, the phylogeny-free regression line and its intercept can be forced onto the raw data for detailed study of factors that cause deviations from this line.

Finally, it has been argued that the contrasts method is inappropriate for studies of allometry due to the confounding effect of grade shifts.[73] A grade shift occurs when some variable, shared through common descent, produces a shift in the relationship between the main variables with no change in their slopes (Fig. 8). Thus, the allometric exponent is the same in the two groups, but the allometric coefficient differs. A common example involves brain size in mammals. For example, Figure 9 shows that anthropoid primates have larger neocortices relative to the rest of the brain than do prosimians, and that primates in general have larger relative neocortices than do other mammals, such as insectivores.[84]

Contrary to the argument, however, the contrasts method tends to do better than analysis of species values in this situation.[26,30,100] This is because multiple data points in a species analysis are treated incorrectly as independent but, with the method of independent contrasts, only a single contrast is calculated for each grade shift, and this single data point will produce less bias (Fig. 8). Moreover, the method of independent contrasts can be used to provide a rigorous diagnostic for grade shifts. Following the logic in Figure 8, this is done by demonstrating that the slopes for the taxa do not differ, and next by showing that the contrast between them is significantly greater than expected.[44] Thus, grade shifts cause no problems for calculating allometric slopes with the contrasts method, and the general contrasts approach actually allows formal analysis of such grade shifts.

An alternative to using independent contrasts in allometry uses the species values but adjusts the degrees of freedom.[21] This "degrees-of-freedom approach" has been used in biological anthropology,[101–104] but we think its use should be reconsidered for several reasons. First, computer simulation has shown that this approach is conservative,[34] which has been construed to mean that the method is acceptable. By comparison, the method of independent contrasts gives expected Type I error rates; that is, it behaves statistically as expected. The contrasts method should therefore be preferred over other methods.[34] Second, nonindependence is not simply about the total degrees of freedom, but rather how those degrees of freedom are partitioned among the available species data points.[97] The degrees-of-freedom method reduces the degrees of freedom, but it does not partition the data into truly independent components. Finally, when applying the method to allometric questions, a large number of data points from speciose clades may bias the allometric estimate. This occurs, for example, if there are many closely related species giving data points on extreme edges of the *X*- or *Y*-axes (that is, high leverage points). Despite these concerns, nested analysis of variance methods may have heuristic value for assessing the degree of phylogenetic correlation, although more direct approaches are available that avoid problems of unequal sample sizes in nested classes[56] (for example, using Moran's I[92] or simulation methods[96]).

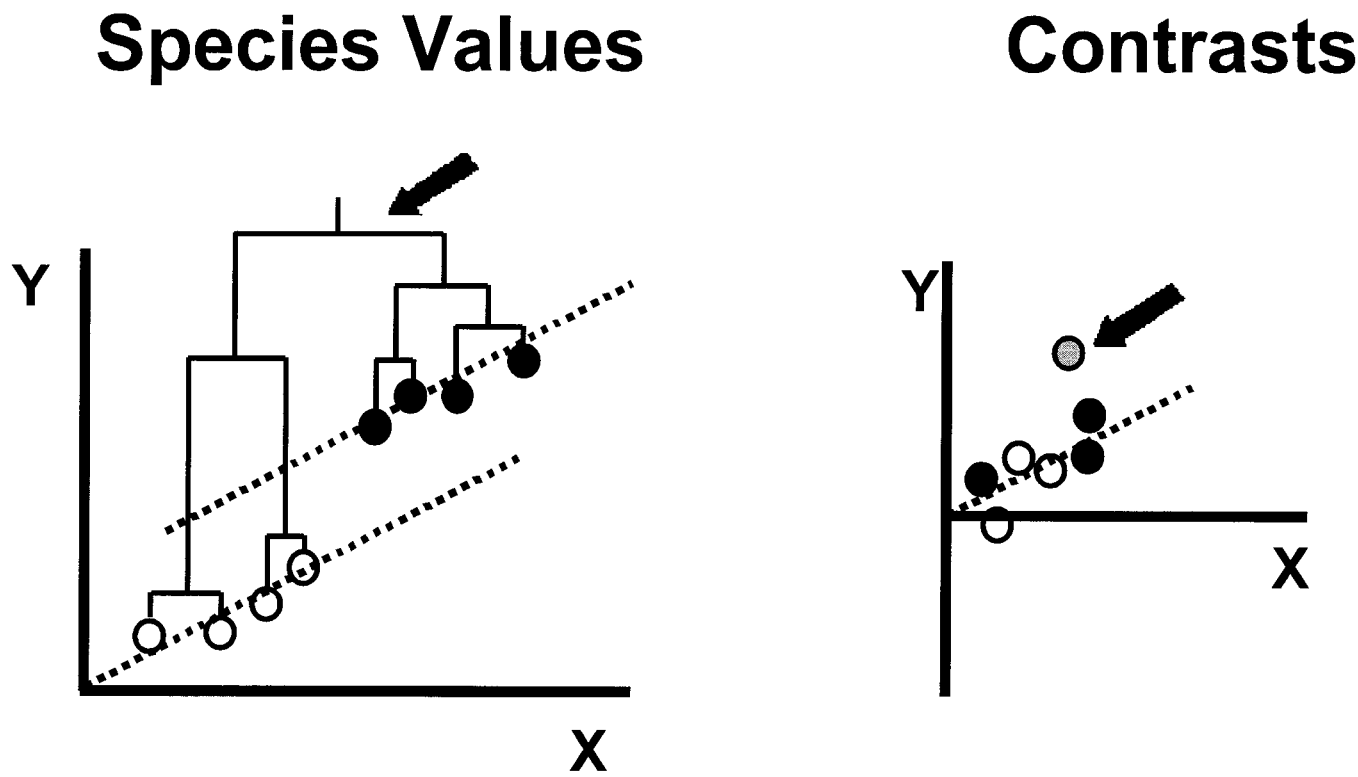



Figure 8. Grade shifts and contrasts. In a plot of species values, a grade shift tends to bias the allometric exponent (the slope). In this example, the allometric exponent is biased upward, although biases downward are also possible.[30] In a contrasts plot, however, the contrast corresponding to the difference between these grades shows up as a single outlier, indicated by the arrow. While this outlier, if not excluded, will still exert positive bias, the resulting bias is less than the multiple degrees of freedom in a species analysis. This figure illustrates, in the case of allometry, how phylogenetic relations can result in biased analysis of species values. (Figure taken, with permission, from Nunn and Barton,[30] copyright 2000 by The University of Chicago.).

## CONCLUSIONS AND FUTURE DIRECTIONS

The comparative method has played a pivotal role in generating and testing adaptive hypotheses. Because the comparative approach is nonexperimental, comparative methods control for confounding variables statistically rather than directly.[25,26] This review shows that phylogenetic comparative methods have distinct advantages over previous approaches to comparative analysis. We do not recommend, however, that investigators simply apply a phylogenetic comparative method without first testing its assumptions.[52,53,105] In addition, differences between results calculated from phylogenetic and nonphylogenetic analyses can often be informative. For example, if correlations are lower in analysis of independent contrasts than species values, this may indicate the existence of confounding variables.[51] Thus, phylogenetic analysis should often go hand-in-hand with analysis of species data points, although greater confidence should generally be placed in the phylogenetic results.

**Although we are skeptical that any trait is truly uncorrelated with phylogeny, traits with large amounts of intraspecific variation, such as group size, may be diagnosed as having low phylogenetic correlations. Two important issues, however, have yet to be fully addressed regarding these tests.**

Many areas for future research on phylogenetic comparative methods remain. Because primates are such a well-studied group of organisms, primate researchers, with their large comparative datasets, have much to offer toward improvements in these methods. For example, we investigated the scaling of home range size with group metabolic needs in primates[30] and discovered that the application of different line-fitting models to contrasts data was not so clear-cut.[10] We therefore outlined steps for conducting such analyses. Computer simulation of particular unresolved issues, such as the effect of using contrasts with traits that are uncorrelated with phylogeny, are particularly suited to primatologists, as there are numerous cases available for empirically testing the simulation results (see similar work on carnivores by

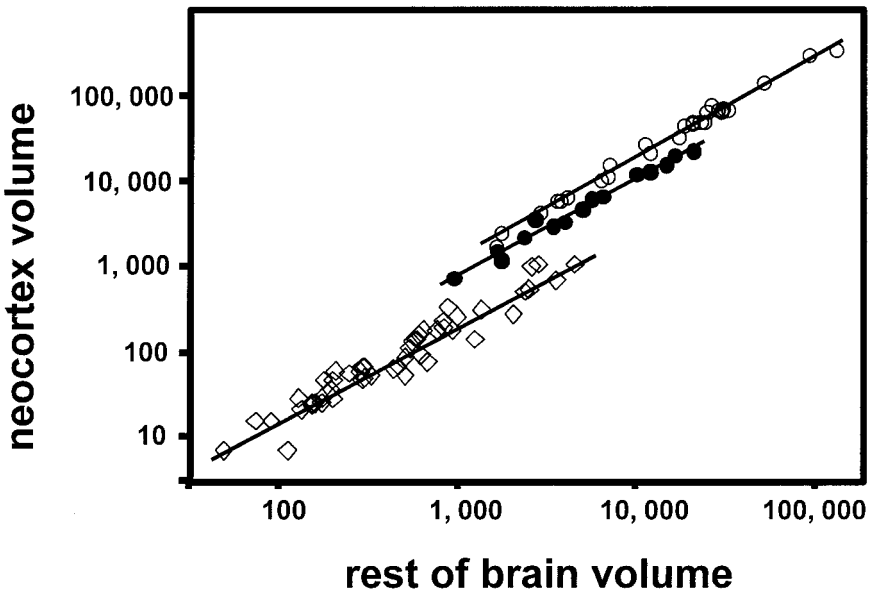



Figure 9. Grade shifts: the example of brain size. Grade shifts in neocortex size relative to the rest of the brain in insectivores and primates. Diamonds are insectivores, filled circles are strepsirhine primates, and open circles are haplorhine primates. The graph suggests that a similar scaling relationship exists within each grade. Because of these grade shifts, the slope calculated across all species is much greater (1.5) than slopes within these clades (1.1 to 1.2). Hence, an analysis that ignored phylogeny would yield a spuriously high scaling exponent. The slope based on independent contrasts is 1.2, suggesting that the contrasts analysis removes the confounding effect of the grade shifts evident in this plot. Reprinted by permission from Nature,[84] copyright 2000, Macmillion Magazines, Ltd.

Gittleman et al.[94]). Thus, traits with extensive intraspecific variation in primates, such as group size, can be examined to evaluate conditions under which a phylogenetic comparative method should be used.

Finally, it is possible to alter the basic phylogenetic approach to address totally new questions.[5,15,17–19,69,106,107] For example, Deaner and Nunn[80] used a method based on independent contrasts to test whether brain size is subject to evolutionary lag relative to body size, but found no support for this hypothesis as an explanation for residuals in brain-body size plots. Thus, phylogenetic approaches to comparative biology have the potential to offer important insights into biological anthropology. We hope that this review results in greater exploration of comparative patterns in a phylogenetic context.

## ACKNOWLEDGMENTS

We thank Dan Blumstein, John Gittleman, Charles Janson, Bill Jungers, Terry Ord, Andy Purvis, Karen Strier, Carel van Schaik, and several anonymous reviewers for helpful discussion. During the writing of this paper, CLN was supported by an NSF Post-Doctoral Research Fellowship in Biological Informatics. Additional support was provided for travel by the NSF, the Duke University Graduate School, the Duke chapter of Sigma Xi (Sally-Hughes Schrader Travel Grant), and the Evolutionary Anthropology Research Group at the University of Durham.

## REFERENCES

**1** Crook JH, Gartlan JC. 1966. Evolution of primate societies. Nature 210:1200–1203.

**2** Clutton-Brock TH, Harvey PH. 1977. Primate ecology and social organization. J Zool Lond 183: 1–39.

**3** Harcourt AH, Harvey PH, Larson SG, Short RV. 1981. Testis weight, body weight and breeding system in primates. Nature 293:55–57.

**4** Martin RD. 1990. Primate origins and evolution. London: Chapman and Hall.

**5** Lee P. 1999. Comparative primate socioecology. Cambridge: Cambridge University Press.

**6** Clutton-Brock TH, Harvey PH. 1976. Evolutionary rules and primate societies. In: Bateson PPG, Hinde RA, editors. Growing points in ethology. Cambridge: Cambridge University Press. p 195–237.

**7** Nunn CL. 1999. The evolution of exaggerated sexual swellings in primates and the graded signal hypothesis. Anim Behav 58:229–246.

**8** Harcourt AH, Purvis A, Liles L. 1995. Sperm competition: mating system, not breeding system, affects testes size of primates. Funct Ecol 9:468–476.

**9** Dunbar RIM. 1998. The social brain hypothesis. Evol Anthropol 6:178–190.

**10** Harvey PH, Pagel MD. 1991. The comparative method in evolutionary biology. Oxford: Oxford University Press.

**11** Clutton-Brock TH, Harvey PH. 1984. Comparative approaches to investigating adaptation. In: Krebs JR, Davies NB, editors. Behavioural ecology. Oxford: Blackwell. p 7–29.

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

**13** Maddison WP. 1990. A method for testing the correlated evolution of two binary characters: are gains or losses concentrated on certain branches of a phylogenetic tree? Evolution 44: 539–557.

**14** Martins EP, Hansen TF. 1996. The statistical analysis of interspecific data: a review and evaluation of phylogenetic comparative methods. In: Martins EP, editor. Phylogenies and the comparative method in animal behavior. New York: Oxford University Press. p 22–75.

**15** Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

**16** Ridley M. 1983. The explanation of organic diversity: the comparative method and adaptations of mating. Oxford: Clarendon.

**17** Harvey PH, Brown AJL, Smith JM, Nee S, editors. 1996. New uses for new phylogenies. Oxford: Oxford University Press.

**18** 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.

**19** Agapow P-M, Isaac NJB. 2000. MacroCAIC: correlations of species richness revealed by comparative analysis. Animal Conservation, in review.

**20** Anthony MRL, Kay RF. 1993. Tooth form and diet in ateline and alouattine primates: reflections on the comparative method. Am J Sci 293A: 356–382.

**21** Smith RJ. 1994. Degrees of freedom in interspecific allometry: an adjustment for the effects of phylogenetic constraint. Am J Phys Anthropol 93:95–107.

**22** Harvey PH, Read AF, Nee S. 1995. Why ecologists need to be phylogenetically challenged. J Ecol 83:535–536.

**23** Westoby M, Leishman MR, Lord JM. 1995. On misinterpreting the phylogenetic correction. J Ecol 83:531–534.

**24** Ricklefs RE, Starck JM. 1996. Applications of phylogenetically independent contrasts: a mixed progress report. Oikos 77:167–172.

**25** 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.

**26** Purvis A, Webster AJ. 1999. Phylogenetically independent contrasts and primate phylogeny. In: Lee P, editor. Comparative primate socioecology. Cambridge: Cambridge University Press. p 44–68.

**27** Di Fiori A, Rendall D. 1994. Evolution of social organization: a reappraisal for primates by using phylogenetic methods. Proc Natl Acad Sci USA 91:9941–9945.

**28** Pagel MD. 1994. The adaptationist wager. In:

Eggleton P, Vane-Wright RI, editors. Phylogenetics and ecology. London: Academic Press. p 29–51.

**29** Nunn CL, van Schaik CP. n.d. Reconstructing the behavioral ecology of extinct primates. In: Plavcan JM, Kay RF, Jungers WL, van Schaik CP, editors. Reconstructing behavior in the fossil record. New York: Plenum Press, in press.

**30** Nunn CL, Barton RA. 2000. Allometric slopes and independent contrasts: a comparative study of Kleiber's law in primate ranging patterns. Am Nat 156:519–533.

**31** Smith RJ, Jungers WL. 1997. Body mass in comparative primatology. J Hum Evol 32:523–559.

**32** Martins EP, Garland T. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. Evolution 45: 534–557.

**33** Purvis A, Gittleman JL, Luh H. 1994. Truth or consequences: effects of phylogenetic accuracy on two comparative methods. J Theor Biol 167: 293–300.

**34** Nunn CL. 1995. A simulation test of Smith's "degrees of freedom" correction for comparative studies. Am J Phys Anthropol 98:355–367.

**35** Diaz-Uriarte R, Garland T. 1996. Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. Syst Biol 45:27–47.

**36** Diaz-Uriarte R, Garland T. 1998. Effects of branch length errors on the performance of phylogenetically independent contrasts. Syst Biol 47: 654–672.

**37** Harvey PH, Rambaut A. 1998. Phylogenetic extinction rates and comparative methodology. Proc R Soc Lond B 265:1691–1696.

**38** Barton RA. 2000. Primate brain evolution: cognitive demands of foraging or of social life? In: Boinski S, Garber PA, editors. On the move: how and why animals travel in groups. Chicago: University of Chicago Press. p 204–237.

**39** Brooks DR, McLennan DA. 1991. Phylogeny, ecology, and behavior. Chicago: University of Chicago Press.

**40** Garland T, Midford PE, Ives AR. 1999. An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral values. Am Zool 39:374–388.

**41** Donoghue M, Ackerly D. 1996. Phylogenetic uncertainties and sensitivity analyses in comparative biology. Philos Trans R Soc Lond B 351: 1241–1249.

**42** de Queiroz K. 1996. Including the characters of interest during tree reconstruction and the problems of circularity and bias in studies of character evolution. Am Nat 148:700–708.

**43** Losos JB. 1995. Community evolution in Greater Antillean anolis lizards: phylogenetic patterns and experimental tests. Philos Trans R Soc Lond B 349:69–75.

**44** Garland TJ, Dickerman AW, Janis CM, Jones JA. 1993. Phylogenetic analysis of covariance by computer simulation. Syst Biol 42:265–292.

**45** Felsenstein J. 1988. Phylogenies and quantitative characters. Ann Rev Ecol Syst 19:445–471.

**46** Nunn CL. 1999. The number of males in primate social groups: a comparative test of the socioecological model. Behav Ecol Sociobiol 46: 1–13.

**47** Maddison W. 1989. Reconstructing character evolution on polytomous cladograms. Cladistics 5:365–377.

**48** Pagel MD. 1992. A method for the analysis of comparative data. J Theoret Biol 156:431–442.

**49** Purvis A, Garland TJ. 1993. Polytomies in comparative analyses of continuous characters. Syst Biol 42:569–575.

**50** Barton RA. 1996. Neocortex size and behavioural ecology in primates. Proc R Soc Lond B 263:173–177.

**51** Price T. 1997. Correlated evolution and independent contrasts. Philos Trans R Soc Lond B 352:519–529.

**52** Garland TJ, Harvey PH, Ives AR. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. Syst Biol 4:18–32.

**53** Purvis A, Rambaut A. 1995. Comparative analysis by independent contrasts (CAIC): an Apple Macintosh application for analysing comparative data. Comp Appl Biosci 11:247–251.

**54** Grafen A. 1989. The phylogenetic regression. Philos Trans R Soc Lond B 326:119–157.

**55** Purvis A. 1995. A composite estimate of primate phylogeny. Philos Trans R Soc Lond B 348: 405–421.

**56** Sokal RR, Rohlf FJ. 1995. Biometry. New York: W. H. Freeman.

**57** Barton RA, Purvis A, Harvey PH. 1995. Evolutionary radiation of visual and olfactory brain systems in primates, bats and insectivores. Philos Trans R Soc Lond B 348:381–392.

**58** Nunn CL, Gittleman JL, Antonovics J. 2000. Promiscuity and the primate immune system. Science 290:1168–1170.

**59** Ridley M. 1986. The number of males in a primate troop. Anim Behav 34:1848–1858.

**60** Maddison WP, Maddison DR. 1992. Macclade: analysis of phylogeny and character evolution. Sunderland, MA: Sinauer Associates.

**61** Cunningham CW, Omland KE, Oakley TH. 1998. Reconstructing ancestral character states: a critical reappraisal. Trends Ecol Evol 13:361–366.

**62** Lorch PD, Eadie JM. 1999. Power of the concentrated changes test for correlated evolution. Syst Biol 48:170–191.

**63** Sillén-Tullberg B. 1993. The effect of biased inclusion of taxa on the correlation between discrete characters in phylogenetic trees. Evolution 47:1182–1191.

**64** Werdelin L, Tullberg BS. 1995. A comparison of two methods to study correlated characters on phylogenetic trees. Cladistics 11:265–277.

**65** Read AF, Nee S. 1995. Inference from binary comparative data. J Theor Biol 173:99–108.

**66** Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc R Soc Lond B 255:37–45.

**67** Pagel M. 1997. Inferring evolutionary processes from phylogenies. Zool Scripta 26:331–348.

**68** Ackerly DD, Donoghue MJ. 1998. Leaf size, sapling allometry, and corner's rules: phylogeny and correlated evolution in maples (Acer). Am Nat 152:767–798.

**69** Nunn CL, Smith KK. 1998. Statistical analyses of developmental sequences: the craniofacial region in marsupial and placental mammals. Am Nat 152:82–101.

**70** Benton MJ. 1999. The history of life: large databases in paleontology. In: Harper DAT, editor. Numerical paleontology. Chichester: John Wiley & Sons. p 249–283.

**71** Strier KB. 1997. Behavioral ecology and conservation biology of primates and other animals. Adv Study Behav 26:101–158.

**72** Struhsaker TT. 2000. Variation in adult sex ratios of red colobus monkey social groups: implications for interspecific comparisons. In: Kappeler PM, editor. Primate males. Cambridge: Cambridge University Press. p 108–119.

**73** Martin RD. 1996. Scaling of the mammalian brain: the maternal energy hypothesis. News Physiol Sci 11:149–153.

**74** Losos J. 1994. An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. Syst Biol 43:117–123.

**75** Martins EP. 1996. Conducting phylogenetic comparative studies when the phylogeny is not known. Evolution 50:12–22.

**76** Abouheif E. 1998. Random trees and the comparative method: a cautionary tale. Evolution 52: 1197–1204.

**77** Møller AP, Birkhead TR. 1992. A pairwise comparative method as illustrated by copulation frequency in birds. Am Nat 139:644–656.

**78** Mitani JC, Gros-Louis J, Manson JH. 1996. Number of males in primate groups: comparative tests of competing hypotheses. Am J Primatol 38:315–332.

**79** Maddison WP. 2000. Testing character correlation using pairwise comparisons on a phylogeny. J Theor Biol 202:195–204.

**80** Deaner RO, Nunn CL. 1999. How quickly do brains catch up with bodies? A comparative method for detecting evolutionary lag. Proc R Soc Lond B 266:687–694.

**81** Schluter D, Price T, Mooers AO, Ludwig D. 1998. Likelihood of ancestor states in adaptive radiation. Evolution 51:1699–1711.

**82** Oakley TH, Cunningham CW. 2000. Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny. Evolution 54:397–405.

**83** Harvey PH, Rambaut A. 2000. Comparative analyses for adaptive radiations. Proc R Soc Lond B 355:1–7.

**84** Barton RA, Harvey PH. 2000. Mosaic evolution of brain structure in mammals. Nature 405: 1055–1058.

**85** Harvey PH. 2000. Why and how phylogenetic relationships should be incorporated into studies of scaling. In: Brown JH, West GB, editors. Scaling in biology. Oxford University Press. p 253–265.

**86** Morand S, Harvey PH. 2000. Mammalian metabolism, longevity and parasite species richness. Proc R Soc Lond B 267:1999–2003.

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

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

**89** Hansen TF, Armbruster WS, Antonsen L. 2000. Comparative analysis of character displacement and spatial adaptations as illustrated by the evolution of Dalechampia blossoms. Am Nat 156:S17–S34.

**90** Gingerich PD, Smith BH, Rosenberg K. 1982. Allometric scaling in the dentition of primates and prediction of body weight from tooth size in fossils. Am J Phys Anthropol 58:81–100.

**91** Losos JB. 1999. Uncertainty in the reconstruction of ancestral character states and limitations on the use of phylogenetic comparative methods. Anim Behav 58:1319–1324.

**92** Gittleman JL, Kot M. 1990. Adaptation: statistics and a null model for estimating phylogenetic effects. Syst Zool 39:227–241.

**93** Gittleman JL, Luh HK. 1994. Phylogeny, evolutionary models, and comparative methods: a simulation study. In: Eggleton P, Vane-Wright RI, editors. Phylogenetics and ecology. London: Academic Press. p 103–122.

**94** Gittleman JL, Anderson CG, Kot M, Luh H-K. 1996. Phylogenetic lability and rates of evolution:

a comparison of behavioral, morphological and life history traits. In: Martins EP, editor. The comparative method in animal behavior. Oxford University Press. p 166–205.

**95** Bjorklund M. 1997. Are "comparative methods" always necessary? Oikos 80:607–612.

**96** Abouheif E. 1999. A method for testing the assumption of phylogenetic independence in comparative data. Evol Ecol Res 1:895–909.

**97** Pagel M. 1993. Seeking the evolutionary regression coefficient: an analysis of what comparative methods measure. J Theor Biol 164:191–205.

**98** Berrigan D, Charnov EL, Purvis A, Harvey PH. 1993. Phylogenetic contrasts and the evolution of mammalian life histories. Evol Ecol 7:270–278.

**99** Martins EP. 1999. Estimation of ancestral states of continuous characters: a computer simulation study. Syst Biol 48:642–650.

**100** Barton RA. 1999. The evolutionary ecology of the primate brain. In: Lee P, editor. Comparative primate socioecology. Cambridge: Cambridge University Press. p 167–194.

**101** Runestad JA, Ruff CB. 1995. Structural adaptations for gliding in mammals with implications for locomotor behavior in paromomyids. Am J Phys Anthropol 98:101–119.

**102** Runestad JA. 1997. Postcranial adaptations for climbing in loridae (Primates). J Zool 242: 261–290.

**103** Paine RR, Godfrey LR. 1997. The scaling of skeletal microanatomy in non-human primates. J Zool 241:803–821.

**104** Strait DS. 1999. The scaling of basicranial flexion and length. J Hum Evol 37:701–719.

**105** Freckleton RP. 2000. Phylogenetic tests of ecological and evolutionary hypotheses: checking for phylogenetic independence. Funct Ecol 14:129–134.

**106** Purvis A, Agapow PM, Gittleman JL, Mace GM. 2000. Nonrandom extinction and the loss of evolutionary history. Science 288:328–330.

**107** Purvis A, Gittleman JL, Cowlishaw G, Mace GM. 2000. Predicting extinction risk in declining species. Proc R Soc Lond B 267:1947–1952.

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

**109** Sanderson MJ, Purvis A, Henze C. 1998. Phylogenetic supertrees: assembling the trees of life. Trends Ecol Evol 13:105–109.

**110** Corbet GB, Hill JE. 1991. A world list of mammalian species. Oxford: Oxford University Press.

**111** Bininda-Emonds ORP, Gittleman JL, Purvis A. 1999. Building large trees by combining phylogenetic information: A complete phylogeny of the extant carnivora (Mammalia). Biol Rev 74: 143–175.

**112** Kleiber M. 1961. The fire of life: an introduction to animal energetics. New York: John Wiley & Sons.

**113** Ross C. 1992. Basal metabolic rate, body weight and diet in primates: an evaluation of the evidence. Folia Primatol 58:7–23.



## *Articles in Forthcoming Issues*

- Socioecology of Hanuman Langurs: The Story of Their Success
  *Andreas Koenig and Carola Borries*
- Metric Variation and Species Recognition in the Fossil Record
  *J. Michael Plavcan and Dana Cope*
- Evolution of Human Growth
  *Steven Leigh*
- History of Health in the Western Hemisphere: 4000 B.C. to the Present
  *Richard Steckel, Jerome Rose, Clark Spencer Larsen, and Phillip Walker*
- What Can Developmental Defects of Enamel Reveal About Physiological Stress in Non-Human Primates?
  *Debbie Guatelli-Steinberg*
- Rethinking Anthropoid Origins
  *Christopher Heesy*
- The Sensory Ecology of Primate Food Perception
  *Nathaniel Dominy, Peter Lucas, Daniel Osorio, and Nayuta Yamashita*
- Dmanisi and Dispersal
  *Leo Gabunia, Susan Anton, David Lordkipanidze, Abesalom Vekua, Antje Justus, and Carl Swisher III*
- Showing-off, Handicap Signaling, and the Evolution of Men's Work
  *Kristen Hawkes and R. Bliege Bird*
- Recent Developments in the Study of Wild Chimpanzee Behavior
  *John Mitani, David Watts, and Martin Muller*