Review

Problems of allometric scaling analysis: examples from mammalian reproductive biology

Robert D. Martin[1],*, Michel Genoud[2] and Charlotte K. Hemelrijk[3]

[1]*Academic Affairs, The Field Museum, 1400 S. Lake Shore Drive, Chicago, IL 60605-2496, USA,* [2]*Département d'Ecologie et d'Evolution, Université de Lausanne, Bâtiment de Biologie, CH-1015 Lausanne, Switzerland* and [3]*Centre for Ecology and Evolutionary Studies, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands*

*Author for correspondence (e-mail: rdmartin@fieldmuseum.org)

Summary

Biological scaling analyses employing the widely used bivariate allometric model are beset by at least four interacting problems: (1) choice of an appropriate best-fit line with due attention to the influence of outliers; (2) objective recognition of divergent subsets in the data (allometric grades); (3) potential restrictions on statistical independence resulting from phylogenetic inertia; and (4) the need for extreme caution in inferring causation from correlation. A new non-parametric line-fitting technique has been developed that eliminates requirements for normality of distribution, greatly reduces the influence of outliers and permits objective recognition of grade shifts in substantial datasets. This technique is applied in scaling analyses of mammalian gestation periods and of neonatal body mass in primates. These analyses feed into a re-examination, conducted with partial correlation analysis, of the maternal energy hypothesis relating to mammalian brain evolution, which suggests links between body size and brain size in neonates and adults, gestation period and basal metabolic rate. Much has been made of the potential problem of phylogenetic inertia as a confounding factor in scaling analyses. However, this problem may be less severe than suspected earlier because nested analyses of variance conducted on residual variation (rather than on raw values) reveals that there is considerable variance at low taxonomic levels. In fact, limited divergence in body size between closely related species is one of the prime examples of phylogenetic inertia. One common approach to eliminating perceived problems of phylogenetic inertia in allometric analyses has been calculation of 'independent contrast values'. It is demonstrated that the reasoning behind this approach is flawed in several ways. Calculation of contrast values for closely related species of similar body size is, in fact, highly questionable, particularly when there are major deviations from the best-fit line for the scaling relationship under scrutiny.

Key words: best-fit lines, brain size, gestation period, grade shift, maternal energy hypothesis, metabolic scaling, non-parametric line fitting, neonatal body mass, outliers, phylogenetic inertia.

Introduction

Analysis of scaling relationships between individual biological features and body size across species (interspecific allometry) has a long history and has made a particularly valuable contribution to studies of animal physiology (Schmidt-Nielsen, 1972, 1984). Scaling of basal metabolic rate (BMR) has been a central concern (Brody and Procter, 1932; Brody, 1945; Kleiber, 1932, 1947, 1961; Hemmingsen, 1950, 1960; McNab, 2002). Reference is often made to the influence of Huxley (1932), but his interest focussed on scaling within species (intraspecific allometry), notably with respect to growth. In fact, it seems that interspecific scaling analyses were initiated considerably earlier in studies of vertebrate brain size (e.g. Snell, 1891; Dubois, 1897a,b, 1913). There has also been considerable interest in the scaling of variables in reproductive biology, notably among mammals (e.g. Portmann, 1941, 1965), and these have been shown to connect up with brain development and hence with the scaling of brain size (Portmann, 1962; Sacher and Staffeldt, 1974; Sacher, 1982). This review focuses on two examples taken from mammalian reproductive biology and on exploration of potential connections with the development and completed size of the brain. It follows a 'frequentist approach', in which the probability of the data having occurred is estimated given a particular hypothesis. An alternative approach taken by some authors is Bayesian inference, which uses information available prior to the study to generate a probability distribution (Ellison, 2004).

Bivariate allometric analyses of the relationship between any chosen individual biological dimension (Y) and body size (X, usually body mass) have generally used the empirical scaling formula $Y=kX^{\alpha}$, in which k is the allometric coefficient and α

the allometric exponent (Gould, 1966; Martin, 1989). It is standard practice to convert data to logarithmic form for analysis, as this linearizes the allometric formula ($\log Y = \alpha \cdot \log X + \log k$), making it more amenable to statistical treatment and interpretation. The exponent α is directly indicated by the slope of the best-fit line and the coefficient k by the intercept. This widely used approach is mainly applied to identify positive or negative deviations of individual species from the overall scaling relationship (their residual values) and to detect grade shifts between groups of species (Martin, 1989). One widely used application of this approach has been examination of the size of the brain (or parts thereof) relative to body size in a given sample of species and to seek potential links with behavioural and/or ecological features (e.g. Jerison, 1963, 1973; Eisenberg and Wilson, 1978; Clutton-Brock and Harvey, 1980; Harvey and Bennett, 1983; Gittleman, 1986; Sawaguchi, 1992; Barton et al., 1995; Dunbar, 1995; Allman, 1999; Barton and Harvey, 2000). Although allometric analysis can also permit identification and subsequent interpretation of the scaling exponent (α), as has been common in physiological studies (Schmidt-Nielsen, 1972, 1984), this aspect has generally received far less attention. One striking exception has been determination and interpretation of the scaling exponent for the relationship between basal metabolic rate and body mass in mammals. Debate about whether the 'true' value of the interspecific scaling exponent is 0.67 or 0.75 has recently been fuelled by the observation that the empirically determined value may be biased upwards by inclusion of large-bodied herbivores with marked digestive fermentation (White and Seymour, 2003, 2005). There have been increasingly sophisticated attempts to provide a valid theoretical explanation for the commonly accepted scaling exponent value of 0.75 (e.g. West et al., 1997, 1999; Darveau et al., 2002; Bejan, 2001, 2005; Dawson, 2001, 2005), but these are bedevilled both by uncertainty about the actual scaling exponent for BMR and by the unresolved conflict between competing explanations. Furthermore, questions have been raised about the reliability of simple power laws in this context (Weibel, 2002).

Problems in allometric analysis

The seeming simplicity of the allometric equation and the apparent ease with which a line can be fitted to logarithmically converted data are deceptive. Complex statistical and logical problems inherent in such bivariate analysis have been progressively recognized (Fig. 1). To make matters worse, these problems interact in ways that hinder straightforward interpretation of allometric analyses. The most immediately obvious problem in allometric analysis, already debated extensively, is the choice of an appropriate best-fit line (e.g. see Harvey and Mace, 1982; Martin, 1989; Martin and Barbour, 1989; Harvey and Pagel, 1991; Riska, 1991). Most published allometric analyses have used least-squares regression (Model I regression) to determine a best-fit line. This is simple to calculate because it only takes into account unidirectional deviations of species from the best-fit line (those relative to the Y-axis).
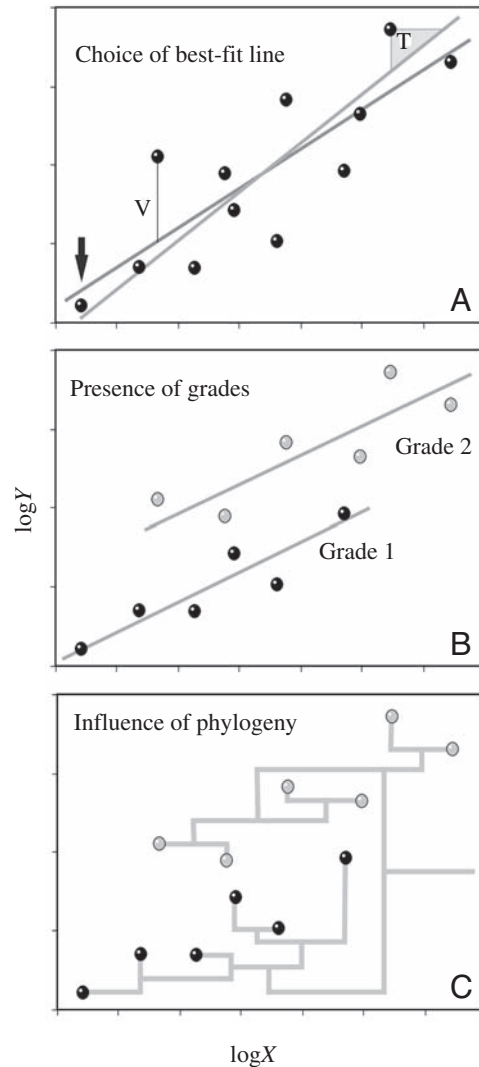

Fig. 1. Illustration of three fundamental statistical problems involved in allometric scaling analyses. Problem 1 (A): results can be affected by the choice of line-fitting technique. Even in this case, where the correlation coefficient is relatively high (r=0.96), the least-squares regression (darker line) and the reduced major axis (lighter line) yield different results. The point indicated by the arrow lies below the regression but above the reduced major axis. The least-squares regression minimizes only deviations along the y-axis (as shown by the distance V for one point), whereas the reduced major axis minimizes deviations along both axes (area of triangle T shown for one point). Problem 2 (B): subsets of species may be separated by grade shifts, following the same scaling principle (common slope value) but differing in intercept value. In this case, two separate lines can be fitted to the dataset whereas an overall best-fit line yields a quite different result. Problem 3 (C): individual taxa in the comparison may not be statistically independent because of phylogenetic relationships within the tree to which they belong. (After Martin, 1998.)

However, this approach rests on the requirements that (a) the X-variable be measured without error, and (b) the Y-variable be clearly dependent upon the X-variable. With interspecific biological data, it is inherently unlikely that both of these criteria will be met, even if in some cases measurement error in body

mass may be minor in comparison to that in the *Y*-variable (e.g. see Taper and Marquet, 1996). In an analysis of mammalian brain mass relative to body mass, for example, there is no reason why the measurement error for brain mass should be greater than that for body mass. Indeed, for various reasons it is more likely that the converse will be true. Furthermore, it is not evident that brain mass is unidirectionally dependent on body mass in any sense. Both brain mass and body mass depend on species-specific growth processes programmed in the genome, and some evidence in fact suggests that brain growth may serve as a pacemaker for bodily growth (Sacher and Staffeldt, 1974). To escape the questionable twin assumptions underlying the use of least-squares regression in bivariate allometric analysis, various authors have instead used a Model II regression approach that allows for variation in both variables and does not require a distinction between dependent and independent variables. The major axis and the reduced major axis have both been used for this purpose. Nevertheless, it is still widely held that the least-squares regression is appropriate for any kind of prediction. Because the residual value for any species is the difference between the actual *Y*-value and that 'predicted' by the best-fit line, this implies that the least-squares regression may be the method of choice for one of the main concerns in allometric analyses. Hence, uncertainty about the correct line-fitting procedure continues. In cases where the data fit fairly closely to a single best-fit line, the choice of line-fitting technique is relatively unimportant. However, in cases where the data are widely scattered relative to the line (thus containing potentially interesting information about differential biological adaptation among species), the alternative line-fitting methods can yield very different conclusions.

Determination of a best-fit line using any of the commonly used techniques (least-squares regression, major axis or reduced major axis) is complicated by two additional factors. First, there is an underlying assumption that the *X*- and *Y*-variables are normally distributed and, indeed, that the standard model of the bivariate normal distribution is potentially applicable. However, whereas the assumption of bivariate normal distribution may be justifiable for certain intraspecific comparisons, it is rarely if ever appropriate for interspecific comparisons (Martin and Barbour, 1989). The three commonly used line-fitting techniques are all derived from the general structural relationship model and thus ideally require knowledge about the distribution of the data. In comparative analyses, estimation of error variances is problematic because scatter in the data reflects a combination of sampling error and biological variation, with no means of distinguishing between them (Riska, 1991). The second obstacle to determination of an appropriate best-fit line is that individual values deviating greatly from the line (outliers), particularly if located at the extremes of the line, can strongly influence the value obtained for the slope. Because of their mode of calculation, least-squares regression, major axis and reduced major axis are all very sensitive to outliers. To avoid the problematic standard requirements for normality of distribution and knowledge of error variances, and to achieve decreased sensitivity to outliers,

a non-parametric line-fitting technique was developed (Isler et al., 2002). This is an iterative method in which the slope of the best-fit line is obtained as the angle of rotation required to minimize a measure of the degree of dependence (*D*) between marginal values of the *X*- and *Y*-variables. *D* is obtained as the integral of the difference between the density of the common distribution of *X* and *Y* and the product of the marginal densities of *X* and *Y*. The pivotal point for rotation is provided by the median values of *X* and *Y*, and the data are subdivided into quantiles for assessment of dependence. This non-parametric 'rotation method' involves no assumptions about error distributions. Furthermore, it proved to be remarkably resistant to the influence of outliers in comparison to standard parametric techniques.

Marked dispersion of points around the best-fit line becomes even more of a problem when the data are heterogeneous, falling into distinct subgroups (grades) that arguably require determination of separate best-fit lines. Determination of a single best-fit line can yield a very misleading result in cases where grade shifts remain unrecognized. Hitherto, however, there has been no readily available technique permitting objective recognition of grade shifts and appropriate analysis of data subsets. Analysis of separate grades has been limited to cases where the investigator suspected for biological reasons of some kind that subdivision of the dataset would be appropriate. As a matter of course, it should be asked whether a single best-fit line is appropriate for any given dataset, and it is advisable to test the outcome of fitting separate lines to subsets that may be suspected to exist on taxonomic, functional or other grounds. In fact, an additional benefit of the new non-parametric 'rotation line' method (Isler et al., 2002) is that it provides an objective basis for recognition of grade shifts in datasets in cases where these are very marked. One step in applying this method is visual inspection of minima in the *D*-values indicating the degree of dependence between marginal values of *X* and *Y*. If there is a reasonable clear linear relationship between the *X*- and *Y*-values and no marked subdivision of the data into grades, a single minimum is found for the *D*-values. However, if the dataset is subdivided into distinctly separate grades, this is indicated by one or more additional local minima for the *D*-values. Unfortunately, if the sample size is small or the distributions of subsets in a given dataset overlap extensively, the existence of grades will not be detectable in this way.

There is another potential problem involved in bivariate allometric analysis in that the values for individual species included in a given sample may not meet the criterion of statistical independence because all species are connected to one another to varying degrees in a phylogenetic tree (Felsenstein, 1985). There is, in principle, a danger that closely related species do not represent independent samples from the adaptive landscape and may therefore bias the outcome of allometric analysis. If 'phylogenetic inertia' or 'phylogenetic constraint' exerts a particularly strong effect, values derived from closely related species could come close to being repeat observations. Furthermore, because different branches of a

phylogenetic tree typically differ in their species richness, particularly well-represented taxa (e.g. numerous species in a single genus) could bias the outcome of analysis even without pronounced phylogenetic inertia. Once this potential problem was recognized, one pragmatic approach that was adopted was to take average values at a higher taxonomic level (e.g. for genera or subfamilies) rather than values for individual species. However, it could still be argued that phylogenetic inertia continues to influence the outcome of analyses even at the generic or subfamilial level, and an additional drawback of the use of values averaged above the species level is that it results in a progressive reduction in sample size. For this reason, considerable attention has been devoted to alternative approaches that might eliminate the potential problem of statistical non-independence arising from phylogenetic inertia. By far the most widely used technique is calculation of 'independent contrast values' that are subjected to analysis instead of the raw data for species (Harvey and Pagel, 1991). This technique is based on the assumption that evolutionary change follows a Brownian motion model (Felsenstein, 1985). Under this assumption, the differences between the raw X and Y values for any pair of taxa (the 'contrast values') represent independent evolutionary change, whereas the raw values may themselves be subject to phylogenetic inertia. Presenting this in simplistic terms, for any pair of taxa conforming perfectly to an allometric relationship described by the standard formula, the following subtraction applies:

$$\log Y_1 = \alpha \cdot \log X_1 + \log k$$

$$\log Y_2 = \alpha \cdot \log X_2 + \log k$$

$$(\log Y_1 - \log Y_2) = \alpha \cdot (\log X_1 - \log X_2) \, .$$

Accordingly, a best-fit line determined for contrast values (e.g. log Y_1–log Y_2 *versus* log X_1–log X_2) should have a slope directly reflecting the value of the scaling exponent (α) and should pass through the origin. Because limiting the calculation of contrast values to pairs of extant species would drastically reduce the sample size, the method is extended down through the tree by averaging values above each node and then calculating contrast values between adjacent nodes as well. In a perfectly dichotomous phylogeny, this means that a dataset containing raw mean values for N extant species will yield a total of N–1 contrast values, thus barely reducing the original sample size. One important weakness of the method is that the assumption of a Brownian motion model of evolutionary change, originally proposed by Felsenstein (1985), may not always be adequate. Other more realistic modes of evolution have more recently been introduced (Hansen, 1997; Freckleton et al., 2002). Furthermore, the method also has the major drawback that calculation of contrast values requires the availability of a reliable phylogenetic tree for the taxa under comparison, ideally including information on the ages of individual nodes. However, well-resolved phylogenies are becoming increasingly available for various groups of mammals and other animals. Many recent allometric analyses have used

the standard programme CAIC (comparative analysis by independent contrasts) developed and distributed by Purvis and Rambaut (1995). For primates, such analyses have been facilitated by the availability of a tailor-made consensus phylogenetic tree (Purvis, 1995). In principle, all of the standard methods used for allometric analysis of the raw data can also be used for analysis of contrast values. However, for technical reasons relating to the obligate selection of an independent variable for calculation of contrast values, the best-fit line prescribed for their analysis is a least-squares regression forced through the origin (see also Garland et al., 1992).

It should be noted that insistence on the need for action to offset effects of phylogenetic inertia essentially concerns the issue of reliability of tests for statistical significance. For example, if it is claimed that fruit-eating primates typically have larger brains than leaf-eating primates (e.g. Allman, 1999), it is necessary to exclude the possibility that any probability value attached to this claim is not biased by the influence of recent common ancestry of fruit-eaters and leaf-eaters, respectively. It should be noted that Smith (1994) has suggested an alternative approach to eliminating the effects of phylogenetic constraint in any such comparison by reducing the number of degrees of freedom for calculation of a probability value. In his examples, approximate halving of the degrees of freedom was found to be appropriate. However, this alternative approach leaves open the question of whether, for any given comparison, phylogenetic inertia might have biased the slope of the allometric line, and it begs the question of the reliability of residual values determined for individual species.

All three problems of allometric analysis discussed thus far – choice of an appropriate best-fit line, recognition and appropriate treatment of grade shifts, and coping with potential bias arising from phylogenetic inertia – relate to determination of the allometric relationship and calculation of reliable residual values on that basis. A fourth problem concerns interpretation of these results derived from allometric analysis and arises from dangers inherent in progressing from correlations to causation. It is all too easy to jump to conclusions about underlying causal factors on the basis of just a few allometric analyses, in the extreme case relying on just one bivariate plot. It is essential to recognize that biological variables are typically linked in complex networks and that it is oversimplification to single out a pair of variables for analysis (e.g. adult brain and body mass). This is why the distinction between 'dependent' and 'independent' variables in biological systems is fraught with difficulty. It is important to be especially careful in making such a distinction because it may influence propagation of errors when allometric relationships are used to generate derived ones (e.g. see Taper and Marquet, 1996). It is therefore essential to conduct numerous allometric analyses and to focus on identifying testable hypotheses in order to move cautiously towards a causal interpretation. One technique that can be used in tackling complex networks of biological variables is partial correlation, which permits determination of the correlation

remaining between any two variables after the influence of one or more other variables has been excluded. Although direct interpretation of a partial correlation as a causal relationship should be avoided just as carefully as for a simple correlation (Sokal and Rohlf, 1981), this approach is certainly a valuable tool in attempting to progress from correlation to functional interpretation.

These four main problems in allometric analysis are further exacerbated by potential interactions between them. For instance, the choice of an appropriate line-fitting method in allometric analysis can become a secondary issue if grade shifts are present, because determination of single best-fit line of any kind for a dataset that is clearly subdivided can yield misleading results ('grade confusion'). Now it might be thought that, in addition to eliminating potential bias resulting from phylogenetic inertia, calculation of independent contrasts could eliminate effects of grade shifts, such that no special consideration of this problem is required. Indeed, this has been claimed as a benefit of the CAIC programme (Purvis and Rambaut, 1995). In principle, it might be expected that a grade shift in one group of species might arise in the common ancestor of that group and thus affect only one contrast value calculated for that ancestral node. This might in fact happen when a single grade shift is present right at the base of an evolutionary tree. In other cases, the practice of averaging values between adjacent nodes to permit calculation of contrast values within the phylogenetic tree will actually lead to diffusion of the effects of grade shifts through lower nodes. Thus, especially if there are multiple grade shifts within a given tree and if they are located well above the initial ancestral node,

a complex pattern of deviating contrast values will result. Of course, if misleading results emerge from allometric analysis because of failure to recognize and deal effectively with grade shifts in a dataset, any functional interpretation developed on that basis must also be flawed.

Examples from mammalian reproduction

Practical application of the interconnected principles of allometric analysis discussed above can be illustrated with two examples from mammalian reproductive biology that link up with inference of a possible connection between reproductive variables and the evolution of brain size. The two examples involve (1) gestation periods in mammals, and (2) neonatal body mass in primates. Before presenting the allometric analysis of these two variables, it should be noted that none of the variables involved (adult body mass, gestation period, neonatal body mass) conforms to a normal distribution (Fig. 2: Shapiro-Wilk test; for dataset (1) with $N=429$, $M_b=0.91$ and 0.96 for log gestation time and log adult body mass, respectively; for dataset (2) with $N=109$, $M_b=0.93$ and 0.95 for log neonatal body mass and log adult body mass, respectively; $P<0.01$ in all cases;), thus confirming that a non-parametric approach is preferable.

Scaling of mammalian gestation periods provides a particularly striking example of the fundamental need to recognize grade distinctions. If no attention is paid to possible grade distinctions, the resulting conclusion is that the slope for scaling of mammalian gestation periods to body size is quite steep (Fig. 3). Any single best-fit line that is determined yields
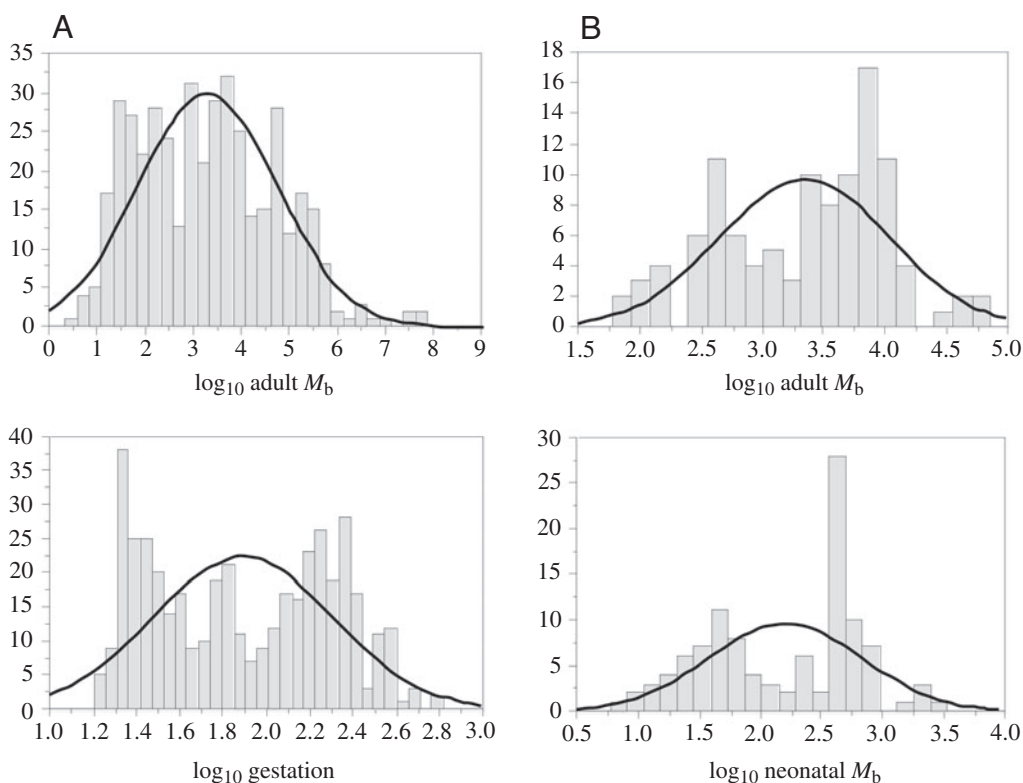

Fig. 2. Histograms showing distributions of (A) gestation periods and adult body mass for mammals (first author's dataset) and (B) neonatal mass and adult body mass for primates (data from Smith and Leigh, 1998). In both cases, it can be seen that neither variable meets the criterion of normality of distribution, as indicated by the superimposed curves. (See text for tests of significance.) Body mass (M_b; g); gestation (days).
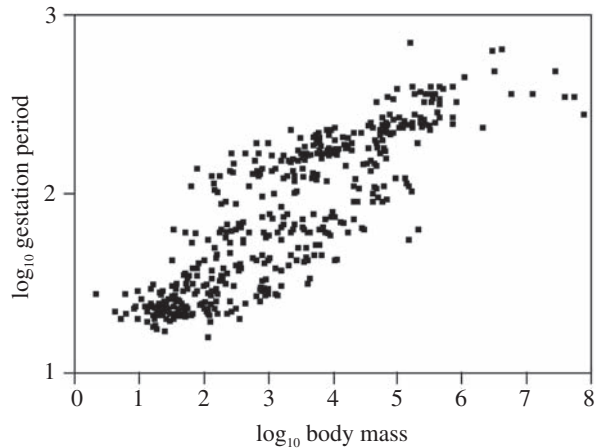
Fig. 3. Scaling of gestation periods for a sample of 429 placental mammal species (first author's dataset). Regardless of the method used, determination of a single best-fit line for the data yields a value for the scaling exponent (α) close to 0.25 (least-squares regression, 0.240; major axis, 0.255; reduced major axis, 0.285; rotation line, 0.260). (After Martin, 1989.) Gestation (days); body mass (g).

a scaling exponent value close to $\alpha=0.25$. There have been suggestions that a scaling exponent of this value is typical for individual components of mammalian life histories (e.g. developmental periods and lifespan; West et al., 1997), and it has even been suspected that there might be a connection with the scaling exponent value of $\alpha=0.75$ for basal metabolic rate in that the two exponents combined could result in a 'metabolic lifetime' with a scaling exponent of $\alpha=1$. However, all of this
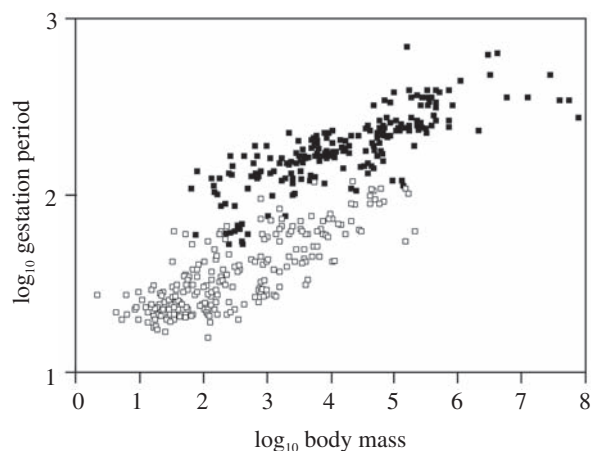

Fig. 4. Scaling of gestation periods for the same sample of 429 placental mammal species, following subdivision into species with altricial neonates and those with precocial neonates. Best-fit lines fitted separately to the two grades have approximately similar slopes that are both markedly lower than the overall slope for the single distribution shown in Fig. 3 and are generally close to 0.15. (For altricial mammals: least-squares regression, 0.158; major axis, 0.160; reduced major axis, 0.198; rotation line, 0.176. For precocial mammals: least-squares regression, 0.135; major axis, 0.136; reduced major axis, 0.168; rotation line, 0.133.) (After Martin, 1989.) Gestation (days); body mass (g).

ignores the long-established distinction between mammal species that give birth to multiple litters of poorly developed altricial neonates and those that give birth to (typically) single, well-developed precocial neonates (Portmann, 1938, 1939, 1965). It should be patently obvious that, other things being equal, development of an altricial neonate should require less time than development of a precocial neonate. Consequently, it is to be expected that there should be a distinct grade shift in a plot of gestation periods against adult body mass, with values for species with precocial neonates generally exceeding those for species with altricial neonates at any given adult body mass. This prediction is, indeed, borne out if scaling of gestation period in placental mammals is analyzed for altricial and precocial neonates as two separate grades (Martin and MacLarnon, 1985). The slope of the scaling relationship for
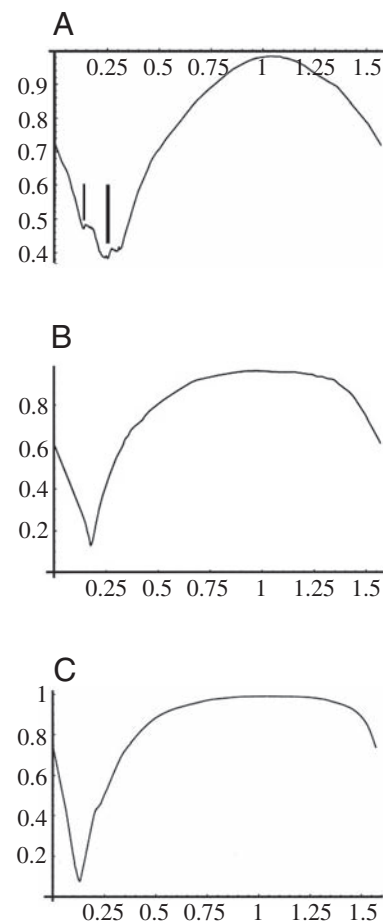

Fig. 5. Plots of D-values (degree of dependence of marginal values of X and Y) derived by applying the rotation method to scaling of gestation periods for a large sample of placental mammal species. Rotation of the data in rads is indicated on the abscissa. (A) For the entire sample ($N=429$), a global minimum (thick vertical line) is located at 0.255 rads, corresponding to $\alpha=0.260$. However, a local minimum value (thin vertical line) is located at about 0.15 rads, corresponding to $\alpha=0.151$. (B) For altricial mammals taken alone ($N=227$), there is a clear single global minimum located at 0.174 rads, corresponding to $\alpha=0.176$. (C) For precocial mammals taken alone ($N=202$), there is a clear single global minimum located at 0.132 rads, corresponding to $\alpha=0.133$.

each individual grade (altricial or precocial) is clearly less steep than for the overall dataset (Fig. 4). In fact, the scaling exponent value for each grade is almost halved, to α≈0.15. Hence, there is in fact no empirical support for the proposition that mammalian gestation periods resemble other life-history components in scaling with an exponent value close to α=0.25.

Scaling of gestation periods in placental mammals provides a good test case for exploring the grade-detecting capacity of the non-parametric 'rotation' line-fitting method (Isler et al., 2002). In fact, it had already been shown that iterative fitting of lines of different slopes to plots of gestation period against adult body mass yielded a bimodal distribution at a value of α≈0.15 (Martin, 1989). Clearly, such a result should not emerge with a homogeneous dataset conforming even approximately to a bivariate normal distribution. Following application of the rotation method to data on gestation periods for 429 placental mammal species, visual inspection of a plot of *D*-values reveals that, in addition to the global minimum value corresponding to α=0.26 there is a local minimum value corresponding to α≈0.15 (Fig. 5A). When altricial species (*N*=227) and precocial species (*N*=202) are analysed separately, in each case the plot of *D*-values exhibits a single global minimum (Fig. 5B,C). The minimum for altricial mammals corresponds to an α value of 0.176, while the minimum for precocial mammals corresponds to an α value of 0.133. The mean of these two values is α=0.155, corresponding to the local minimum value seen in Fig. 5A.

In fact, the problem posed by grade shifts in the scaling of gestation periods in placental mammals is even more complicated than the simple division between altricial and precocial species would suggest. If individual taxonomic groups within each category are examined, it can be seen that there are less pronounced grade shifts between them. Among altricial mammals, lipotyphlan insectivores and carnivores generally have relatively longer gestation periods than lagomorphs and myomorph rodents (Fig. 6A), while among precocial mammals primates tend to have relatively longer gestation periods than artiodactyls, and the latter in turn tend to have relatively longer gestation periods than hystricomorph rodents (Fig. 6B). However, there is considerable overlap between taxonomic groups within each category of neonate type, such that the curves for *D*-values do not show any local minima for either category (Fig. 5B,C). Hence, detection of subtle grade shifts still depends on careful examination of the data to seek differences between biologically meaningful groups (separate taxa or functional groupings). In the case of gestation periods, it is important to note that the ancestral condition for placental mammals was probably production of litters of altricial neonates (Portmann, 1938, 1939; Martin,
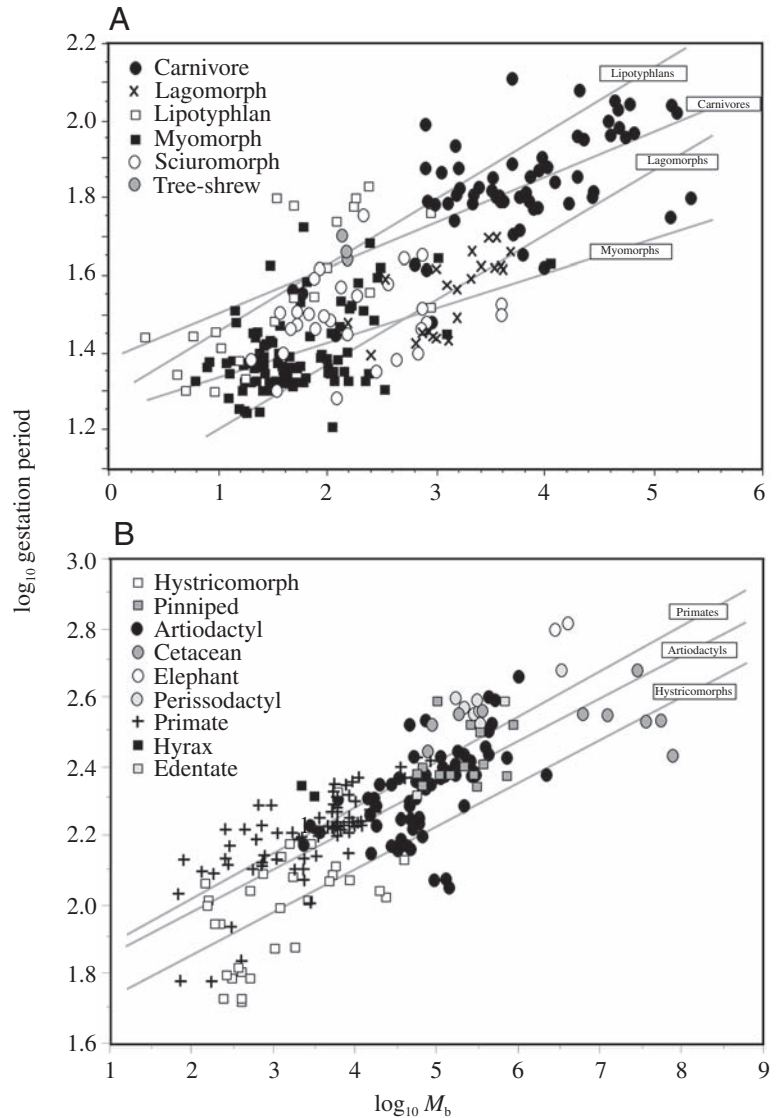

Fig. 6. Within the two major groups of altricial and precocial mammals, there are further grade distinctions in the scaling of gestation periods between taxonomic categories. (A) Among altricial mammals, lipotyphlan insectivores and carnivores generally have relatively longer gestation periods than myomorph rodents. (B) Among precocial mammals, primates generally have relatively longer gestation periods than artiodactyls, while the latter tend to have longer gestation periods than hystricomorph rodents overall. (Lines are least-squares regressions, used simply as a visual guide to differences between taxonomic groups.) Gestation (days); body mass (M_b in g).

1990). Accordingly, in addition to the minor divergences now observable among altricial mammals (Fig. 5A), there were probably several major grade shifts associated with the evolution of precocial offspring. Assuming that no reversals occurred, the molecular phylogeny for placental mammals generated by Murphy et al. (2001) would require shifts from the altricial to the precocial condition in 10 separate lineages (primates; hyraxes + elephants + sirenians; artiodactyls + cetaceans; perissodactyls; hystricomorph rodents; elephant shrews; bats; pangolins; anteaters; xenarthrans). Despite the
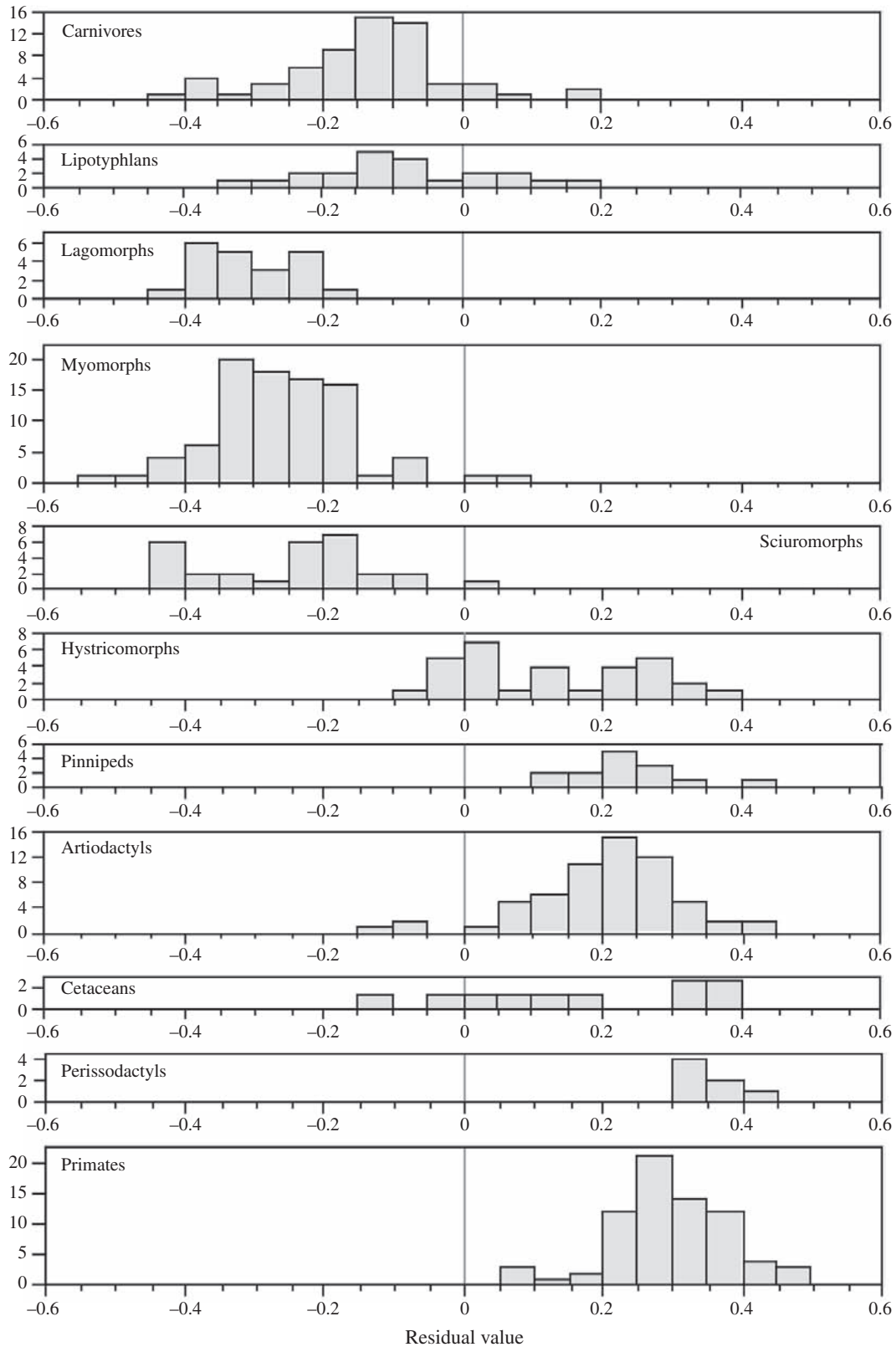
Fig. 7. Histograms showing residual values for gestation periods of placental mammals calculated using an exponent value of 0.15, subdivided into individual taxonomic groups. Although there is some variability within each taxon, it can be seen that there is a fairly clear separation between groups with altricial neonates (carnivores, lipotyphlan insectivores, lagomorphs, myomorph and sciuromorph rodents), which typically have negative residual values, and groups with precocial neonates (hystricomorph rodents, pinnipeds, artiodactyls, cetaceans, perissodactyls and primates), which typically have positive residual values.

existence of minor grade distinctions within categories, it is obvious from histograms of residual values for gestation period, calculated with an exponent value of α=0.15, that there is a fundamental dichotomy between altricial and precocial species (Fig. 7).

The second example of the need to recognize distinct grades in reproductive biology is provided by the scaling of neonatal body mass in primates. It has been known for some time that in strepsirrhine primates (lemurs and lorises) neonates are markedly smaller relative to adult body mass than in haplorhine primates (tarsiers, monkeys, apes and humans) (Leutenegger, 1973; Martin, 1990). If a single best-fit line is determined for the scaling of neonatal body mass in primates (Fig. 8), the commonly used parametric techniques all yield an exponent value (α) close to 0.90 (least-squares regression: 0.874; major axis: 0.906; reduced major axis: 0.909). However, visual inspection of a plot of *D*-values derived from application of the rotation method clearly shows that, in addition to the global minimum value corresponding to α=0.916 there is a local minimum value corresponding to α=0.624 (Fig. 9A). When strepsirrhine primates (*N*=28) and haplorhine primates (*N*=81) are analysed separately, in each case the plot of D-values exhibits a single global minimum (Fig. 9B,C). The minimum for strepsirrhines corresponds to a α value of 0.688, while the minimum for haplorhines corresponds to a α value of 0.862. The average of these two values is α=0.775, which is higher than the local minimum value seen in Fig. 9A. Given this discrepancy and the fact that the α values determined for strepsirrhines and haplorhines differ from one another, it seems likely that there are further subtle grade distinctions within the dataset, in addition to the primary distinction between strepsirrhines and haplorhines (Fig. 10).

Progressing from correlation to causation

Considerable care is needed in attempting to proceed from

the empirical results of allometric scaling analyses to inference of underlying causal connections. An illustrative example is provided by analyses of potential links between brain size and basal metabolism in placental mammals. At the simplest level, several authors noted from analyses of large datasets that the value of the scaling exponent for brain mass is closely similar to that found for basal metabolic rate, with α≈0.75 in both cases (Bauchot, 1978; Eisenberg, 1981; Armstrong, 1982; Hofman, 1982; Martin, 1990). This led to a number of proposals for some kind of connection between brain size and basal metabolic rate. However, it is at once apparent that any existing link must be indirect because there is far less residual variation relative to the best-fit line with BMR than with brain

Fig. 9. Plots of *D*-values derived by applying the rotation method to scaling of neonatal body mass for primate species (data from Smith and Leigh, 1998). Rotation of the data in rads is indicated on the abscissa. (A) For the entire sample (*N*=109), a global minimum (thick vertical line) is located at 0.742 rads, corresponding to α=0.916. However, an additional local minimum value (thin vertical line) is located at 0.558 rads, corresponding to α=0.624. (B) For strepsirrhine primates taken alone (*N*=28), the curve is somewhat irregular because of the small sample size, but there is a global minimum located at 0.603 rads, corresponding to α=0.688. (C) For haplorhine primates taken alone (*N*=81), the sample size is considerably larger and there is a clear single global minimum located at 0.711 rads, corresponding to α=0.862.
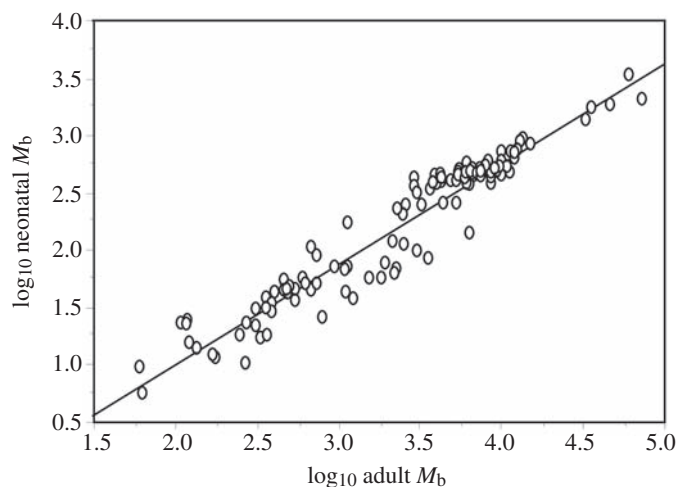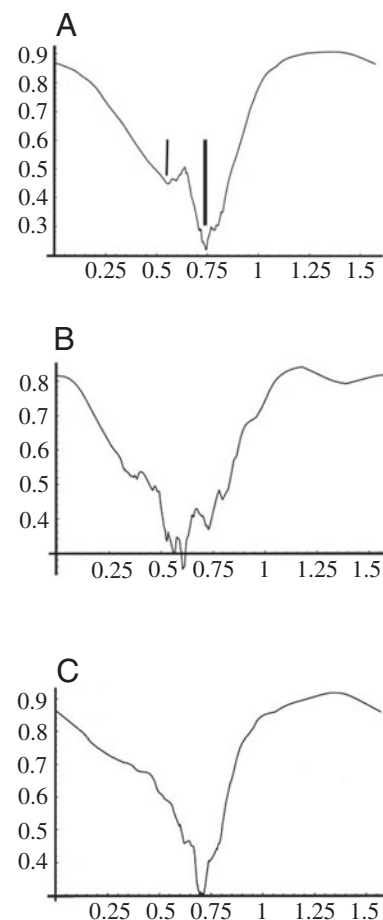
Fig. 8. Scaling of neonatal body mass for the sample of 109 primate species (data from Smith and Leigh, 1998). A single best-fit line fitted to the data is seemingly appropriate. Body mass (M_b in g).
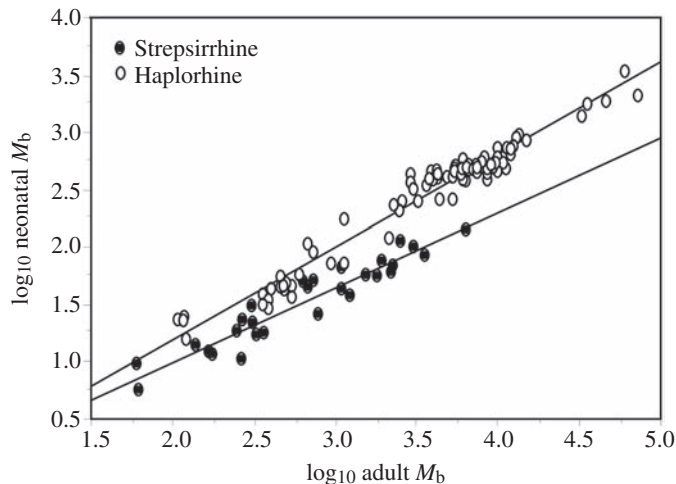
Fig. 10. Scaling of neonatal body mass for the sample of 109 primate species (data from Smith and Leigh, 1998), following subdivision into strepsirrhines (*N*=28) and haplorhines (*N*=81). Best-fit lines fitted separately to the two grades have approximately similar slopes that are both markedly lower than the slope of the single best-fit line in Fig. 8. Body mass (M_b in g).

size. Overall, BMR varies only by a factor of four relative to body size, whereas brain size shows much greater variation relative to body size, showing a 25-fold range of variation. Hence, there is considerable variation in relative brain size that cannot be explained by variation in relative BMR. This and other considerations led the first author to propose the maternal energy hypothesis (Martin, 1981, 1983, 1996, 1998). In this hypothesis, it is proposed that the mother's metabolic turnover constrains energy availability for brain development in the embryo/foetus during intrauterine development and in the offspring during postnatal life up to the time of weaning. Accordingly, differences between species in gestation period and lactation period could generate variability in completed brain size that is not directly attributable to BMR. One of
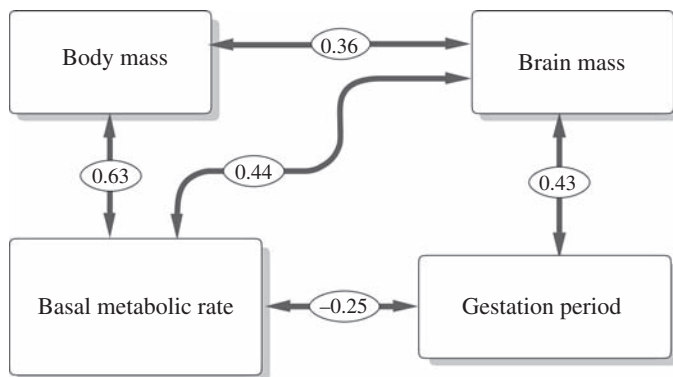

Fig. 11. Flow diagram showing partial correlations in a four-way analysis of body mass, BMR, gestation period and brain mass for a sample of 51 placental mammal species. Substantial positive partial correlations exist between all pairs of variables except for BMR and gestation period, where the partial correlation is negative.

the initial indicators of a potential link between maternal physiology and brain development is the finding that neonatal brain mass is more tightly correlated with gestation period than is neonatal body mass in placental mammals (Sacher and Staffeldt, 1974).

Because of the typical existence of complex biological networks, it is unwise to rely on individual correlations between variables. For this reason, it is very useful to employ the technique of partial correlation, as this permits examination of the correlation between any two variables after the influence of other variables has been taken into account. An example of this approach is provided by partial correlations from a four-way analysis involving adult body mass, adult brain mass, basal metabolic rate and gestation period for a sample of 51 placental mammal species. [N.B. A similar analysis for 53 mammal species was reported by Martin (1996, 1998). Data for two species were subsequently found to be questionable, and a repeat analysis with a reduced sample of 51 species yielded somewhat higher correlations.] It can be seen from Fig. 11 that, as expected, there is a strong partial correlation between BMR and body mass. However, it is also seen that adult brain mass shows substantial partial correlations with adult body mass, BMR and gestation period, indicating that all three variables are connected in some way to brain mass. As the maternal energy hypothesis seemingly provides the only potential explanation for a connection between brain size and gestation period as well as BMR, this can be viewed as supporting evidence. Perhaps the most interesting finding, however, concerns the relationship between BMR and gestation period. If the correlation of either variable with body mass is considered in isolation, a strong positive value is obtained. However, when partial correlations are considered, it emerges that the remaining correlation between BMR and gestation is negative (Fig. 11). This suggests that an increase in brain mass may be associated with an increase in either BMR or gestation period relative to body mass, but that these two variables do not increase in tandem. Hence, there is an apparent trade-off between BMR and gestation period in the development of relatively large brains.

Another way to approach the problem is through the use of path analysis, for which an underlying model of causal relationships between the variables is explicitly stated. However, there is the drawback that, when applied to comparative analysis, this technique again raises the issue of having to distinguish between 'dependent' and 'independent' variables. A preliminary path analysis was conducted, although the available dataset (51 species) is somewhat limited for this kind of approach. Adult brain mass was considered to be the dependent variable and the three other variables (adult body mass, BMR, gestation period) were treated as inter-correlated predictor (causal) variables. Using such a model, a high coefficient of determination was achieved (r^2=0.957; Sokal and Rohlf, 1981). Interestingly, the correlation between BMR and brain mass appeared mainly as the result of a strong direct correlation between the two variables (0.59, for a total

observed correlation of 0.97), indirect 'effects' (a term used here in a non-causal or predictive sense) through gestation period and body mass being relatively minor (0.38). The converse was true for the correlation between body mass and brain mass or for the correlation between gestation period and brain mass. In both cases, direct effects turned out to be minor (0.26, for a total observed correlation of 0.97, and 0.17, for a total correlation of 0.75, respectively), the major indirect effect being through BMR in each case (0.58 and 0.39, respectively). This very preliminary analysis confirmed that the relationship between body mass and brain mass is mainly indirect (here through its correlation with BMR), but also suggested that the same applies to gestation period.

It is important to note that these findings for partial correlations and for all of the other analyses discussed thus far were obtained using raw variables. Thus far, no attempt has been made to correct for possible effects of differential phylogenetic relatedness among the species examined in any sample. It is therefore necessary to turn to the issue of potential conflict between such phylogenetic relatedness and the requirement for statistical independence of data points.

The problem of phylogenetic inertia

It is undoubtedly true that failure to consider phylogenetic relationships might lead to misleading results because of the potential problem of phylogenetic inertia. At the very simplest level, over-representation of a group of closely related species in a sample could swamp data from other taxonomic groups. An apt example is that of the hominoid primates (apes), in which relatively species-rich lesser apes (gibbons) outnumber the great apes (chimpanzees, gorillas and orang-utans). One recent taxonomic revision (Groves, 2001) recognizes 14 gibbon species allocated to one genus, as opposed to six great ape species allocated to three genera. As a result, any analysis of scaling relationships among apes at the species level could be biased by over-representation of gibbons. Although this problem could be partially offset by restricting analysis to the generic level, this would reduce the sample size by 80% (from 20 species to four genera). A similar bias can also result in cases where data are more easily available for certain taxa than for others. For instance, in scaling analyses involving Old World monkeys, data are often far more readily available for macaques (*Macaca* species) than for any of the remaining 21 genera, including the highly speciose genus *Cercopithecus*. Phylogenetic inertia can also exert more subtle influences with respect to the origins of specific adaptations. For instance, in examining possible links between diet and relative brain size, it is conceivable that a small number of adaptive shifts from frugivory to folivory might account for any observed pattern. In the oft-quoted example in which fruit-eating primates are found to have relatively larger brains than leaf-eating primates, it is important to be aware of the possibility that leaf-eating species may be descended from a small number of ancestral nodes and that any correlation with brain size that is detected may be weakly supported. Hence, it is certainly important to

examine any dataset for potential sources of bias arising from imbalanced phylogenetic representation.

In the seminal paper by Felsenstein (1985) on potential bias of comparative analyses arising from phylogenetic relatedness, one example specifically cited was a study of brain size scaling in mammals conducted by the first author (Martin, 1981). It was argued that the data points for individual species might not meet the criterion of statistical independence because of their differential degrees of relatedness within the phylogenetic tree. It is, of course, conceivable that this might bias the results, although Felsenstein did not actually demonstrate that it did. It is a moot point whether or not the degree of change accompanying divergence between sister species is sufficient to dilute the effects of phylogenetic inertia to the point where conflict with the requirement of statistical independence is minor and perhaps negligible. Before pursuing this point, it should be noted that any statistical problems associated with differential degrees of relatedness applying to comparisons *within* species should be massive relative to those associated with interspecific comparisons, as differential divergence within a single gene pool over much shorter periods of time must surely entail strong relatedness effects. Curiously, however, this problem has been relatively neglected in comparison to the extensive recent literature on potential effects of phylogenetic inertia in comparisons between species.

Felsenstein's (1985) view that phylogenetic inertia could be a problem in interspecific comparisons was driven in part by the results of nested analysis of variance (ANOVA) conducted with certain variables. Similar results subsequently reported by Harvey and Pagel (1991) bolstered a belief in the necessity for measures to exclude the effects of phylogenetic inertia. Harvey and Pagel reported ANOVA results for several variables in placental mammals (adult body mass, neonatal body mass, gestation period, age at weaning, maximum reproductive lifespan, annual fecundity and annual biomass production), consistently indicating that there is relatively little variance at the level of species and genera. Only 8–20% of the variance was found between species and genera, whereas 80–92% was found at higher taxonomic levels (between families and orders). On the face of it, these figures do seem to suggest that there is relatively little evolutionary divergence between closely related species or even genera and that phylogenetic inertia may hence be a major problem. Initially, this provided a rationale for several studies in which allometric analysis was conducted at the level of the subfamily or above, and subsequently it was invoked as a justification for special techniques such as calculation of 'independent contrasts'. However, it should be emphasized that the nested ANOVAs reported by Felsenstein (1985) and by Harvey and Pagel (1991) were all conducted on the raw data. This approach is questionable because many biological parameters are highly correlated with body mass and because body mass itself provides a prime example of phylogenetic inertia, typically differing far less between closely related species than between distantly related species. For example, there is a relatively limited range of body mass values within each order of

placental mammals (Fig. 12A), and this pattern is replicated between families within any given order, such as primates (Fig. 12B). Because most features are correlated with body mass, it follows that the distribution of variance in raw values of individual biological variables (e.g. gestation period or brain mass) will generally exhibit a pattern very similar to that observed with body mass. Yet the real question of interest in scaling analyses is whether brain mass, for instance, is tightly constrained or relatively free to vary at any given body size. In other words, it is the pattern of variation of residual values rather than that of the raw values that needs examination. When nested ANOVA was conducted on adult body mass, gestation period, brain mass and basal metabolic rate for a large sample of placental mammals, the result obtained with the raw values was similar to that reported by Felsenstein (1985) and by Harvey and Pagel (1991). Only 5–16% of the variance in residual values was found between species and genera, whereas 84–95% was found between families and orders. By contrast, when nested ANOVA was conducted on the residual values for gestation period, brain mass and basal metabolic rate, more variance was detected at low taxonomic levels and there were more pronounced differences between variables. With brain mass, 34.6% of the variance in the residuals was found at the generic and specific levels, and for basal metabolic rate that figure was even higher, at 45.1% (Fig. 13). Hence, with these two variables, analysis conducted at the subfamilial level would have led to exclusion of one third to almost half of the residual variance. With gestation period, however, the picture is very different. Only 7.4% of the variance in the residuals was found at the generic and specific levels, only slightly greater than the value of 5.4% found with the raw data. Thus, it would seem that gestation period – in contrast to brain mass and basal metabolic rate – is, indeed, subject to considerable phylogenetic inertia. Such inertia had been explicitly proposed by Martin and MacLarnon (1985), who noted that gestation periods generally vary little between species within a genus.

At this point, it is necessary to reflect on the meaning of the catch-all term 'phylogenetic inertia', which can encompass several different phenomena (Fig. 14). Global inertia affecting both X and Y values in closely related species, essentially

resulting in repeat values, is perhaps the simplest form imaginable, with similar genotypes constraining organisms to fit a similar bodily pattern in all respects. However, it is also possible that inertia will primarily affect only one of the two
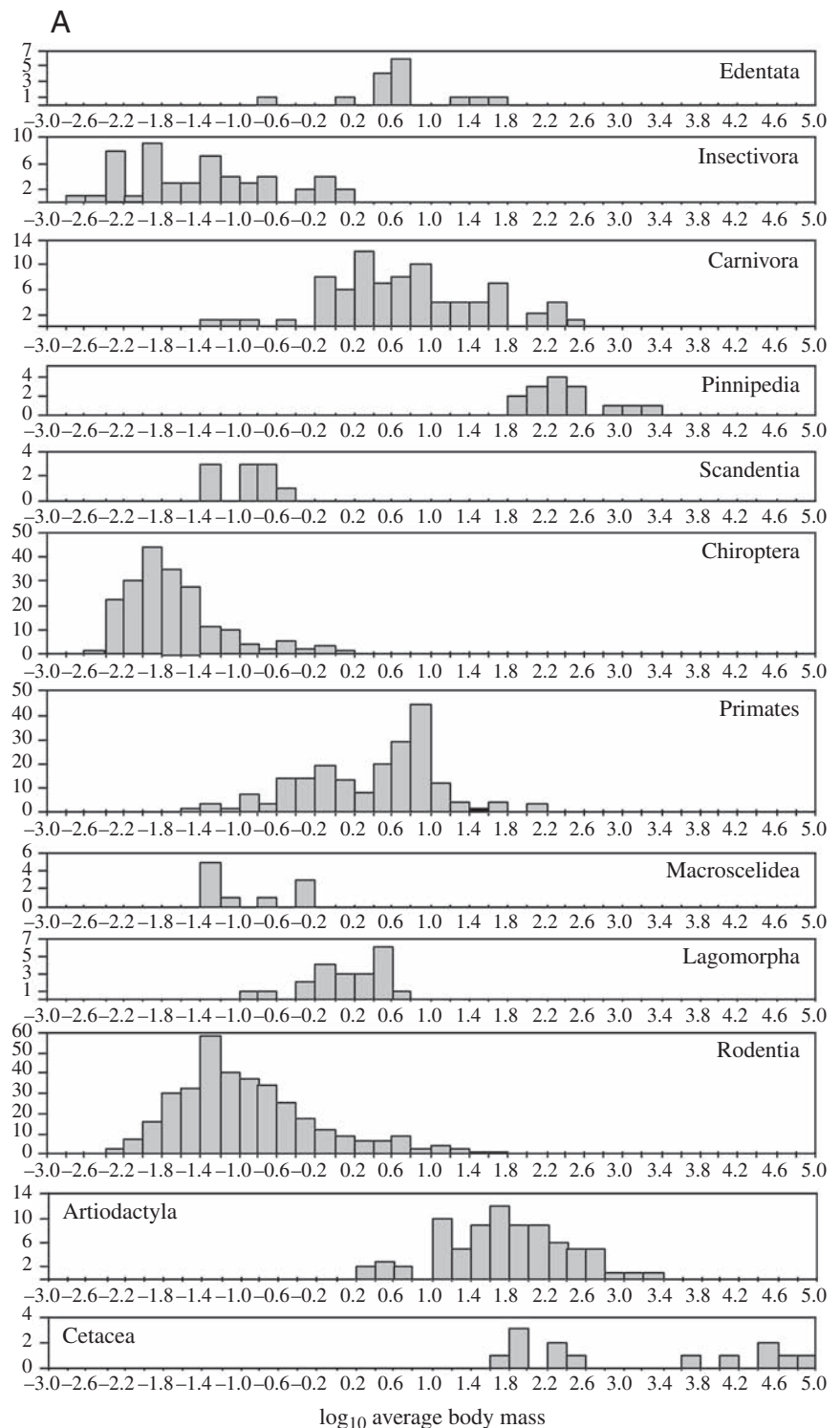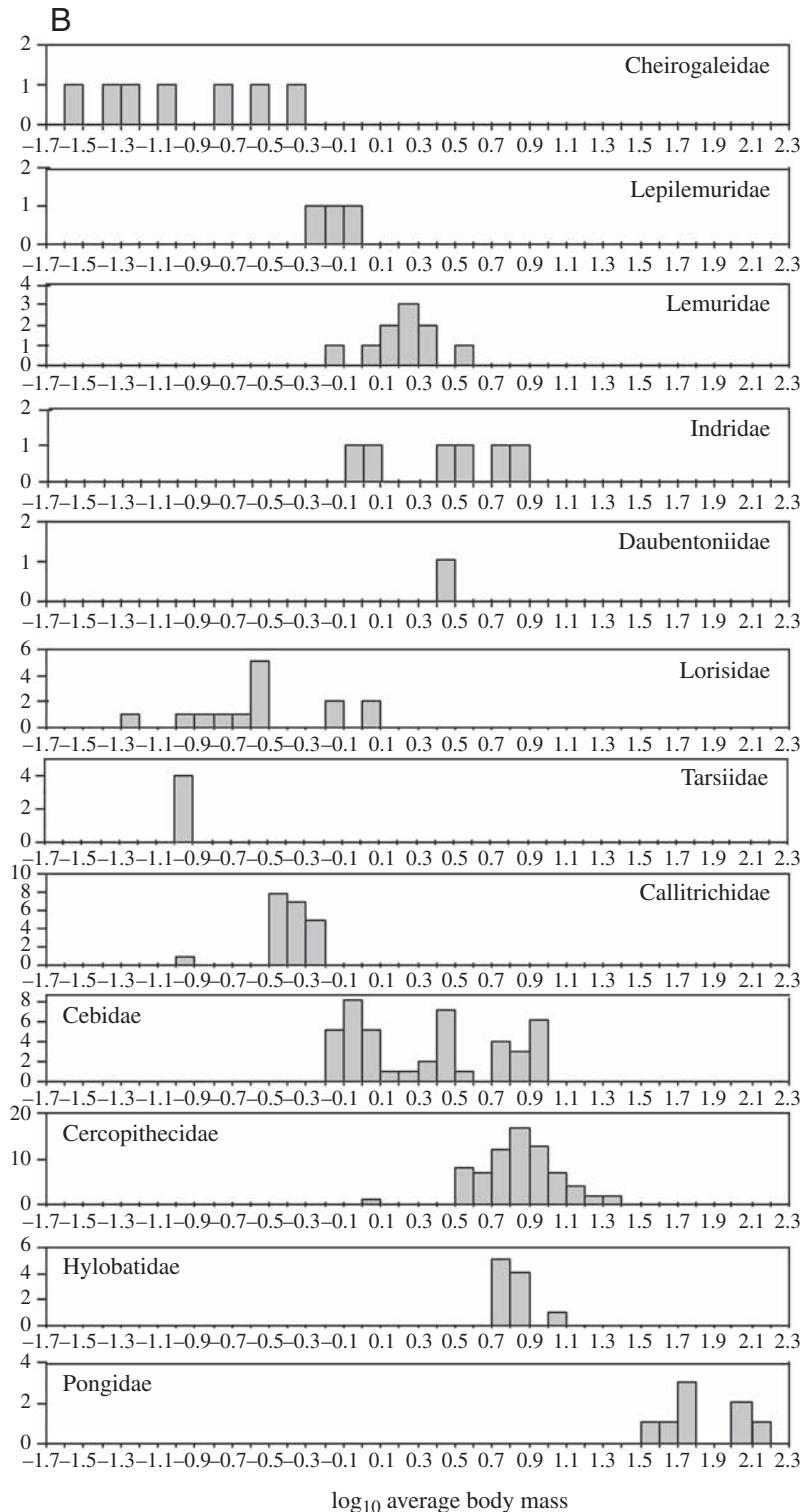

Fig. 12. Histograms showing logarithmic distributions of data for body mass within (A) orders of placental mammals, and (B) families of primates. It can be seen that variation in body mass is typically quite tightly constrained within each lower-level taxon. Body mass (kg).

Fig. 12B

size. The alternative possibility is inertia mainly restricted to *X* values (body size inertia). An example of this could be provided by relatively wide variation in metabolic rate between closely related species without any marked divergence in body size. Finally, it is possible to envisage a constrained allometric relationship between the *X* and *Y* variables as a form of inertia, assuming that close adherence to a given scaling principle might be an inherited property of closely related species (allometric inertia). From this perspective, marked departures from the scaling principle (i.e. relatively large residual values) can be seen as an escape from the general allometric constraint. Given all of these different possibilities for phylogenetic inertia, it is difficult to see how a single analytical procedure (e.g. the CAIC programme) would effectively deal with them all at once. As is seen in Fig. 14, the different kinds of inertia will exert very different effects on any 'independent contrast' values that are calculated. It is particularly important to note that inertia primarily affecting only one variable (scaling inertia or body size inertia) will generate contrast values that deviate from the best-fit line reflecting the allometric relationship. Ironically, only allometric inertia in which both variables are quite tightly constrained to a particular scaling relationship will yield values conforming closely to the best-fit line.

The problems involved in attempting to correct for the effects of phylogenetic inertia using 'independent contrasts' can be illustrated with the practical examples of basal metabolic rate (BMR) and brain size in mammals. Allometric analyses of these two variables contributed to the maternal energy hypothesis for brain size scaling in mammals, the starting point for which was the observation that the value of the scaling exponent (α) is close to 0.75 in both cases. However, whereas the exponent value of 0.75 persisted for BMR when scaling was examined using contrast values, the exponent value for scaling of contrast values for the brain declined to 0.69 (Harvey and Pagel, 1991). There are two possible explanations for this result. The first is that the exponent value of 0.75 obtained for brain size scaling by analysis of the raw data was an artefact arising from effects of phylogenetic inertia, and that application of the contrast method removed taxonomic bias to yield the correct result. The second possibility is that application of the contrast method in fact generated some distortion in the original relationship reflected by the raw data. This latter possibility is suggested by a number of findings. For instance, when analysis of brain size scaling was conducted with overall mean values calculated for individual mammalian orders, a scaling exponent value close to 0.75 was determined (Harvey and Bennett, 1983).

variables. One possibility would be inertia mainly restricted to *Y* values, such that differences in body size between closely related species are not accompanied by adjustments in the *Y* variable (scaling inertia). An example of this might be provided by mammalian gestation periods, although species within a genus also tend to be relatively similar in overall body
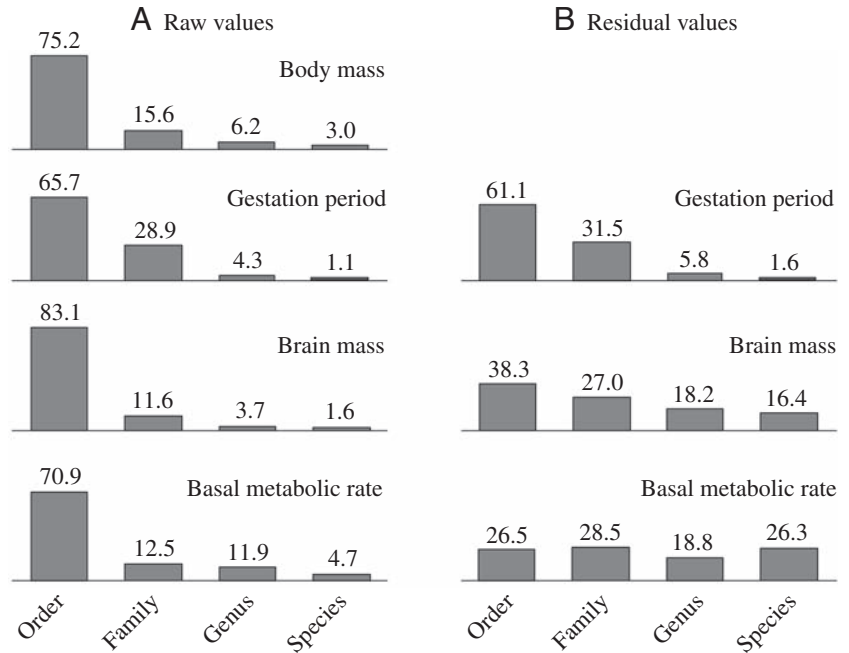
Fig. 13. Results of nested analysis of variance conducted on (A) raw values for body mass, gestation period, brain mass and basal metabolic rate (BMR) in placental mammals, and (B) the residual values for gestation period, brain mass and basal metabolic rate calculated relative to body mass. With all analyses of raw values, relatively little variance is found at the level of comparisons between species or between genera, directly reflecting the limited variation of body mass at those levels. By contrast, with analyses of residual values a substantial proportion of the variance is found at the level of comparisons between species or between genera for brain mass and BMR, but not for gestation period (which is clearly subject to marked phylogenetic inertia). This analysis replicates one previously reported by Ross (1989).

Given that the various orders of placental mammals diverged between 60 and 90 million years ago, it might be expected that phylogenetic inertia would exert relatively little effect on the result obtained for scaling of ordinal mean values. However, it could be argued that differential degrees of phylogenetic relatedness between species within each order might have had some influence on the scaling pattern observed. An alternative approach that circumvents this problem is to take just one species at random from each order of placental mammals and examine the scaling of brain size. When this is done repeatedly (analyses conducted by the first author), the average scaling exponent value found is close to 0.75. There is therefore a mismatch between the results obtained with raw values at the ordinal level and that obtained after the calculation of independent contrast values.

A clue to the reason for the lower exponent for brain size scaling obtained with contrast values is provided by a comparison with BMR, where the exponent value obtained is the

same both for raw values and for contrast values. As already noted, with the raw values for BMR, there is far less residual variation relative to the best-fit line than with the raw values for brain size. Whereas there is only a fourfold variation in relative BMR, for relative brain size the range of variation is more than 25-fold. Hence, it is possible that the greater degree of residual variation in brain size gives rise to special problems in calculation of independent contrast values. In particular, it should be noted that there are multiple grade shifts in relative brain size among mammals. The fact of the matter is that marked variation in residual values leads to conflict with the
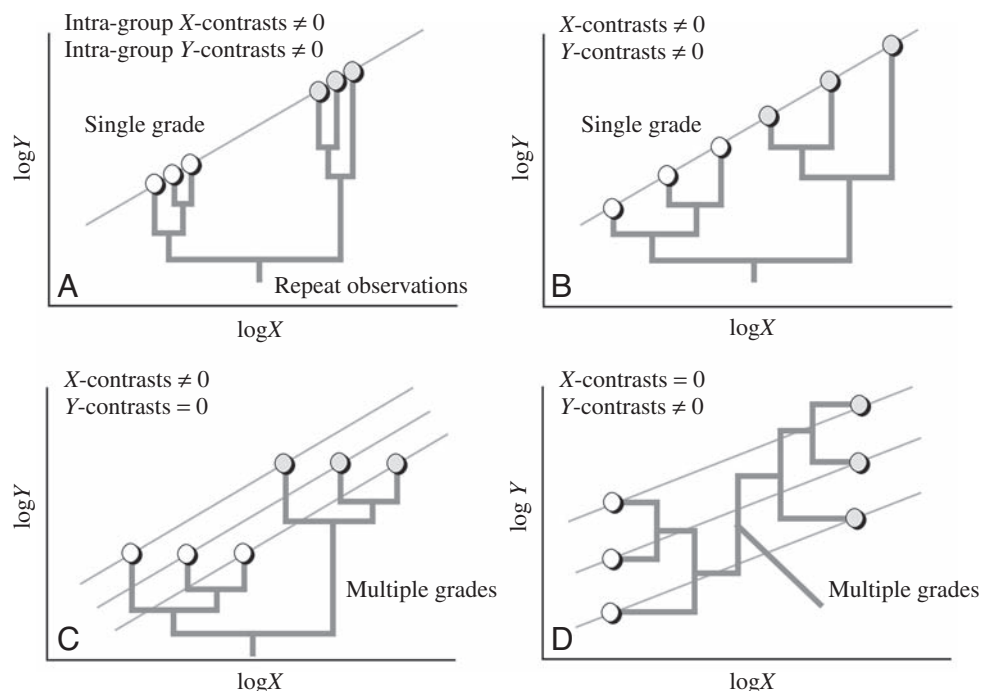
Fig. 14. Illustration of different kinds of phylogenetic inertia: (A) inertia in both X and Y values (global inertia; repeat values); (B) inertia restricted to Y values (scaling inertia); (C) inertia restricted to X values (body size inertia); and (D) constrained allometric scaling of X and Y (allometric inertia).
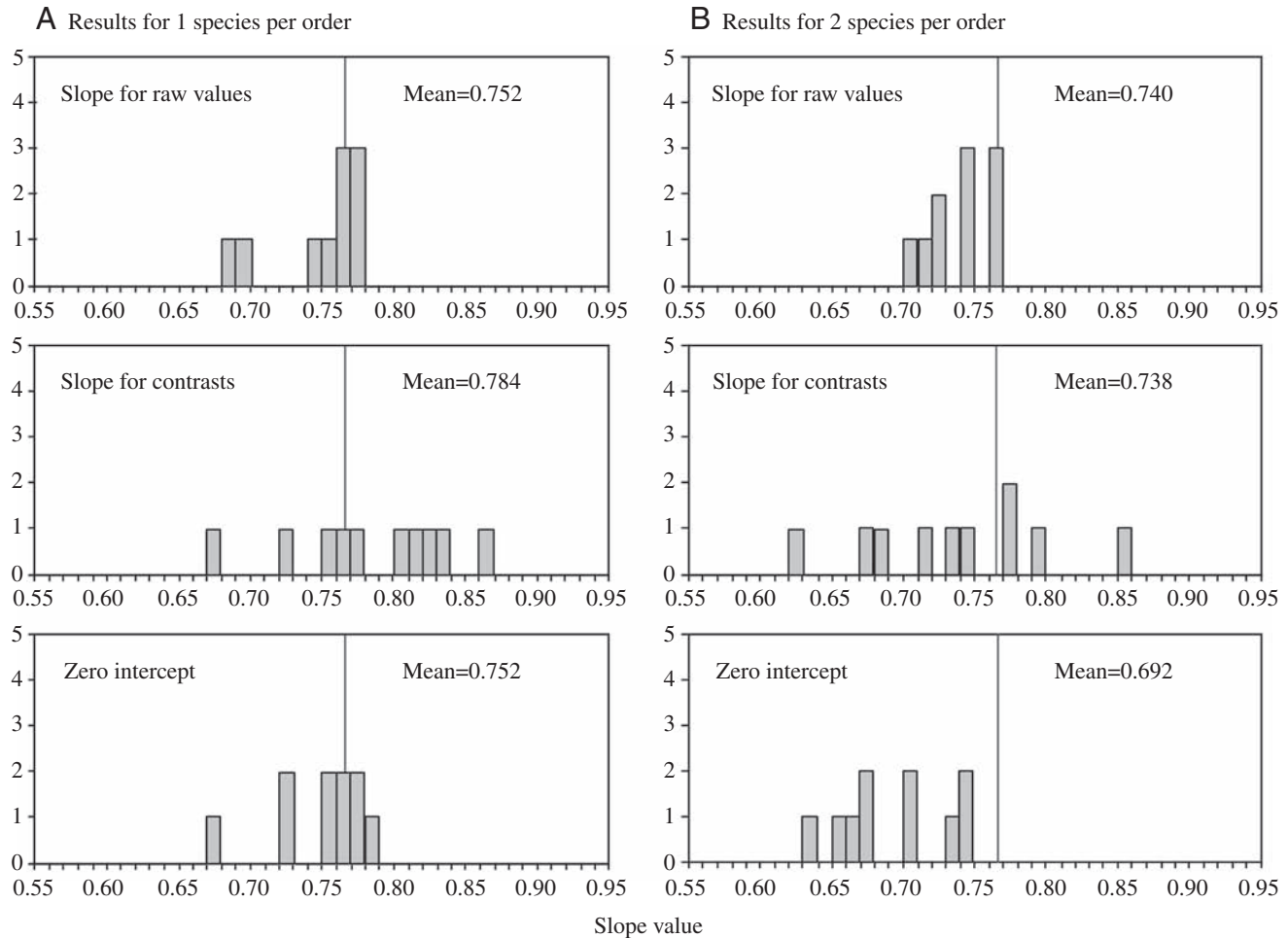
Fig. 15. Results of allometric analysis of brain:body scaling in placental mammals with raw values, contrast values and contrast values forced through the origin, randomly selecting either (A) one species from each order or (B) two species per order.

simple equations given above in support of the rationale for calculation of contrast values for use in allometric analyses. Those equations must be modified to take account of grade shifts (i.e. different values for the scaling coefficient, k) and any kind of measurement error (ε). The revised equations read as follows:

$$\log Y_1 = \alpha \cdot \log X_1 + \log k_1 + \varepsilon_1$$

$$\log Y_2 = \alpha \cdot \log X_2 + \log k_2 + \varepsilon_2$$

$$(\log Y_1 - \log Y_2) = \alpha \cdot (\log X_1 - \log X_2) + (\log k_1 - \log k_2) + (\varepsilon_1 - \varepsilon_2).$$

The implication of this is that any scaling analysis conducted with contrast values – e.g. ($\log Y_1 - \log Y_2$) *versus* ($\log X_1 - \log X_2$) – will be complicated by any difference in value for the scaling coefficient (k_1 *versus* k_2) and by any difference in the error terms (ε). Hence, greater variation in the range of the residual values (i.e. a greater range in k values) will predictably exert an influence on the scaling relationship determined with contrast values. Furthermore, there will be an unexpectedly undesirable effect of comprehensive samples. Because closely related species tend to have similar body sizes (Fig. 12), in a dataset including marked residual variation any contrast

values that are calculated will tend to be relatively small in comparison to the other terms in the equation, ($\log k_1 - \log k_2$) and ($\varepsilon_1 - \varepsilon_2$). The noise generated by these additional terms may well overwhelm the original signal in the raw data. This effect can be demonstrated with a simple example. If, as explained above, one species is selected at random from each order of placental mammals, analysis of the raw data yields an exponent value close to 0.75. If independent contrasts are calculated for such data, the exponent value remains close to 0.75, and if the regression line is forced through the origin the result remains essentially the same as with the raw data (Fig. 15A). However, if the dataset is increased to include two species instead of one from each mammalian order, thus increasing the effect of closely related species with relatively similar body sizes, very different results are obtained. Whereas the raw data still yield an exponent value close to 0.75, the value yielded for the contrasts by a regression line forced through the origin is markedly lower and close to 0.69 (Fig. 15B). Thus, merely by doubling the sample of species taken at random from each order, it is possible to replicate the result reported by Harvey and Pagel (1991) for the scaling of brain size to body size using contrast values. As such calculation of independent contrasts

is typically undertaken with no preceding attempt to identify and separate grades in the dataset, the potentially distorting effect of grade shifts on contrast calculations is effectively ignored. Given the evidence summarized here, it may be concluded that the appropriate exponent value for the scaling of brain size to body size in placental mammals is, in fact, close to 0.75 and that, in this case at least, calculation of contrast values can lead to misleading results. The take-home message is that any analysis that fails to give as much attention to choice of an appropriate line-fitting technique and to grade shifts as to the potential effects of phylogenetic inertia is unlikely to yield reliable results.

References

Allman, J. (1999). *Evolving Brains*. New York: W.H. Freeman/Scientific American.

Armstrong, E. (1982). A look at relative brain size in mammals. *Neurosci. Lett.* **34**, 101-104.

Barton, R. A. and Harvey, P. H. (2000). Mosaic evolution of brain structure in mammals. *Nature* **405**, 1055-1058.

Barton, R. A., Purvis, A. and Harvey, P. H. (1995). Evolutionary radiation of visual and olfactory brain systems in primates, bats and insectivores. *Phil. Trans. R. Soc. Lond. B* **348**, 381-392.

Bauchot, R. (1978). Encephalization in vertebrates: a new mode of calculation for allometry coefficients and isoponderal indices. *Brain Behav. Ecol.* **15**, 1-18.

Bejan, A. (2001). The tree of convective heat streams: its thermal insulation function and the predicted 3/4-power relation between body heat loss and body size. *Int. J. Heat Mass Transf.* **44**, 699-704.

Bejan, A. (2005). The constructal law: tree-shaped flows, body size and heat loss. *J. Exp. Biol.* **208**, 1677-1686.

Brody, S. (1945). *Bioenergetics and Growth*. New York: Rheinhold Publishing Company.

Brody, S. and Procter, T. C. (1932). Relations between basal metabolism and mature body weight in different species of mammals and birds. *Res. Bull. Missouri Agric. Res. Stat.* **166**, 89-101.

Clutton-Brock, T. H. and Harvey, P. H. (1980). Primates, brains and ecology. *J. Zool. Lond.* **190**, 309-323.

Darveau, C. A., Suarez, R. K., Andrews, R. D. and Hochachka, P. W. (2002). Allometric cascade as a unifying principle of body mass effects on metabolism. *Nature* **17**, 166-170.

Dawson, T. H. (2001). Similitude in the cardiovascular system of mammals. *J. Exp. Biol.* **204**, 395-407.

Dawson, T. H. (2005). Modeling of vascular networks. *J. Exp. Biol.* in press.

Dubois, E. (1897a). Sur le rapport du poids de l'encéphale avec la grandeur du corps chez les mammifères. *Bull. Soc. Anthrop. Paris* **8**, 337-376.

Dubois, E. (1897b). Über die Abhängigkeit des Gehirngewichts von der Körpergrösse bei den Säugetieren. *Arch. Anthrop.* **25**, 1-28.

Dubois, E. (1913). On the relation between the quantity of brain and the size of the body in vertebrates. *Proc. Kon. ned. Akad. Wet.* **16**, 647-667.

Dunbar, R. I. M. (1995). Neocortex size and group size in primates: a test of the hypothesis. *J. Hum. Evol.* **28**, 287-296.

Ellison, A. M. (2004). Bayesian inference in ecology. *Ecol. Lett.* **7**, 509-520.

Eisenberg, J. F. (1981). *The Mammalian Radiations: An Analysis of Trends in Evolution, Adaptation and Behaviour*. London, UK: Athlone Press.

Eisenberg, J. F. and Wilson, D. E. (1978). Relative brain size and feeding strategies in the Chiroptera. *Evolution* **32**, 740-751.

Felsenstein, J. (1985). Phylogenies and the comparative method. *Am. Nat.* **125**, 1-15.

Freckleton, R. P., Harvey, P. H. and Pagel, M. (2002). Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.* **160**, 712-726.

Garland, T., Harvey, P. H. and Ives, A. R. (1992). Procedures for the analysis of comparative data using independent contrasts. *Sys. Biol.* **41**, 18-32.

Gittleman, J. L. (1986). Carnivore brain size, behavioral ecology, and phylogeny. *J. Mamm.* **67**, 23-36.

Gould, S. J. (1966). Allometry and size in ontogeny and phylogeny. *Biol. Rev.* **41**, 587-640.

Groves, C. P. (2001). *Primate Taxonomy*. Washington, DC: Smithsonian Institution Press.

Hansen, T. F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evolution* **51**, 1341-1351.

Harvey, P. H. and Bennett, P. M. (1983). Brain size, energetics, ecology and life history patterns. *Nature* **306**, 314-315.

Harvey, P. H. and Mace, G. M. (1982). Comparison between taxa and adaptive trends: problems of methodology. In *Current Problems in Sociobiology* (ed. King's College Sociobiology Group), pp. 343-361. Cambridge: Cambridge University Press.

Harvey, P. H. and Pagel, M. D. (1991). *The Comparative Method in Evolutionary Biology*. Oxford: Oxford University Press.

Hemmingsen, A. M. (1950). The relation of standard (basal) energy metabolism to total fresh weight of living organisms. *Rep. Steno Mem. Hosp.* **4**, 7-58.

Hemmingsen, A. M. (1960). Energy metabolism as related to body size and respiratory surfaces, and its evolution. *Rep. Steno Mem. Hosp.* **9**, 1-110.

Hofman, M. A. (1982). Encephalization in mammals in relation to the size of the cerebral cortex. *Brain Behav. Evol.* **20**, 84-96.

Huxley, J. S. (1932). *Problems of Relative Growth*. London: Methuen.

Isler, K., Barbour, A. D. and Martin, R. D. (2002). Line-fitting by rotation: a nonparametric method for bivariate allometric analysis. *Biometr. J.* **44**, 289-304.

Jerison, H. J. (1963). Interpreting the evolution of the brain. *Hum. Biol.* **35**, 263-291.

Jerison, H. J. (1973). *Evolution of the Brain and Intelligence*. New York: Academic Press.

Kleiber, M. (1932). Body size and metabolism. *Hilgardia* **6**, 315-353.

Kleiber, M. (1947). Body size and metabolic rate. *Physiol. Rev.* **27**, 511-541.

Kleiber, M. (1961). *The Fire of Life: An Introduction to Animal Energetics*. New York: John Wiley.

Leutenegger, W. (1973). Maternal–fetal weight relationships in primates. *Folia Primatol.* **20**, 280-293.

Martin, R. D. (1981). Relative brain size and metabolic rate in terrestrial vertebrates. *Nature* **293**, 57-60.

Martin, R. D. (1983). *Human Brain Evolution in an Ecological Context (52nd James Arthur Lecture on the Evolution of the Human Brain)*. New York: American Museum of Natural History.

Martin, R. D. (1989). Size, shape and evolution. In *Evolutionary Studies – A Centenary Celebration of the Life of Julian Huxley* (ed. M. Keynes), pp. 96-141. London: Eugenics Society.

Martin, R. D. (1990). *Primate Origins and Evolution: A Phylogenetic Reconstruction*. New Jersey: Princeton University Press.

Martin, R. D. (1996). Scaling of the mammalian brain: the maternal energy hypothesis. *News Physiol. Sci.* **11**, 149-156.

Martin, R. D. (1998). Comparative aspects of human brain evolution: scaling, energy costs and confounding variables. In *The Origin and Diversification of Language* (ed. N. G. Jablonski and L. C. Aiello), pp. 35-68. San Francisco: California Academy of Sciences.

Martin, R. D. and Barbour, A. D. (1989). Aspects of line-fitting in bivariate allometric analyses. *Folia Primatol.* **53**, 65-81.

Martin, R. D. and MacLarnon, A. M. (1985). Gestation period, neonatal size and maternal investment in placental mammals. *Nature* **313**, 220-223.

McNab, B. K. (2002). *The Physiological Ecology of Vertebrates: A View from Energetics*. Ithaca and London: Comstock Publishing Association.

Murphy, W. J., Eizirik, E., O'Brien, S. J., Madsen, O., Scally, M., Douady, C. J., Teeling, E., Ryder, A. O., Stanhope, M. J., de Jong, W. W. et al.

(2001). Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* **294**, 2348-2351.

Portmann, A. (1938). Die Ontogenese der Säugetiere als Evolutionsproblem. II. Zahl der Jungen, Tragzeit und Ausbildungsgrad der Jungen bei der Geburt. *Biomorphosis* **1**, 109-126.

Portmann, A. (1939). Nesthocker und Nestflüchter als Entwicklungszustände von verschiedener Wertigkeit bei Vögeln und Säugern. *Rev. Suisse Zool.* **46**, 385-390.

Portmann, A. (1941). Die Tragzeiten der Primaten und die Dauer der Schwangerschaft beim Menschen: Ein Problem der vergleichenden Biologie. *Rev. Suisse Zool.* **48**, 511-518.

Portmann, A. (1962). Cerebralisation und Ontogenese. *Medizin. Grundlagenforsch.* **4**, 1-62.

Portmann, A. (1965). Über die Evolution der Tragzeit bei Säugetieren. *Rev. Suisse Zool.* **72**, 658-666.

Purvis, A. (1995). A composite estimate of primate phylogeny. *Phil. Trans. R. Soc. Lond. B* **348**, 405-421.

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

Riska, B. (1991). Regression models in evolutionary allometry. *Am. Nat.* **138**, 283-299.

Ross, C. A. (1989). Life history strategies of primates. PhD thesis, University of London, UK.

Sacher, G. A. (1982). The role of brain maturation in the evolution of the primates. In *Primate Brain Evolution* (ed. E. Armstrong and D. Falk), pp. 97-112. New York: Plenum Press.

Sacher, G. A. and Staffeldt, E. (1974). Relation of gestation time to brain weight for placental mammals. *Am. Nat.* **108**, 593-615.

Sawaguchi, T. (1992). The size of the neocortex in relation to ecology and social structure in monkeys and apes. *Folia Primatol.* **58**, 131-145.

Schmidt-Nielsen, K. (1972). *How Animals Work*. Cambridge: Cambridge University Press.

Schmidt-Nielsen, K. (1984). *Scaling: Why is Animal Size so Important?* Cambridge: Cambridge University Press.

Smith, R. J. (1994). Degrees of freedom in interspecific allometry: an adjustment for the effects of phylogenetic constraint. *Am. J. Phys. Anthrop.* **93**, 95-107.

Smith, R. J. and Leigh, S. R. (1998). Sexual dimorphism in primate neonatal body mass. *J. Hum. Evol.* **34**, 173-201.

Snell, O. (1891). Die Abhängigkeit des Hirngewichtes von dem Körpergewicht und den geistigen Fähigkeiten. *Arch. Psychiatr. Nervenkr.* **23**, 436-446.

Sokal, R. R. and Rohlf, F. J. (1981). *Biometry*. New York: Freeman and Company.

Taper, M. L. and Marquet, P. A. (1996). How do species really divide resources? *Am. Nat.* **147**, 1072-1086.

Weibel, E. R. (2002). The pitfalls of power laws. *Nature* **417**, 131-132.

West, G. B., Brown, J. H. and Enquist, B. J. (1997). A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122-126.

West, G. B., Brown, J. H. and Enquist, B. J. (1999). The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science* **284**, 1677-1679.

White, C. R. and Seymour, R. S. (2003). Mammalian basal metabolic rate is proportional to body mass(2/3). *Proc. Natl. Acad. Sci. USA* **100**, 4046-4049.

White, C. R. and Seymour, R. S. (2005). Scaling of basal metabolic rate. *J. Exp. Biol.* in press.