Modeling Brain Evolution from Behavior: A Permutational Regression Approach

Pierre Legendre; Francois-Joseph Lapointe; Philippe Casgrain

MODELING BRAIN EVOLUTION FROM BEHAVIOR:
A PERMUTATIONAL REGRESSION APPROACH

PIERRE LEGENDRE, FRANÇOIS-JOSEPH LAPOINTE, AND PHILIPPE CASGRAIN[1]
*Département de sciences biologiques, Université de Montréal,
C.P. 6128, Succ. Centre-Ville, Montréal, Québec, Canada H3C 3J7*

Abstract.—This paper has two complementary purposes: first, to present a method to perform multiple regression on distance matrices, with permutation testing appropriate for path-length matrices representing evolutionary trees, and then, to apply this method to study the joint evolution of brain, behavior and other characteristics in marsupials. To understand the computation method, consider that the dependent matrix is unfolded as a vector **y**; similarly, consider **X** to be a table containing the independent matrices, also unfolded as vectors. A multiple regression is computed to express **y** as a function of **X**. The parameters of this regression (R^2 and partial regression coefficients) are tested by permutations, as follows. When the dependent matrix variable **y** represents a simple distance or similarity matrix, permutations are performed in the same manner as the Mantel permutational test. When it is an ultrametric matrix representing a dendrogram, we use the double-permutation method (Lapointe and Legendre 1990, 1991). When it is a path-length matrix representing an additive tree (cladogram), we use the triple-permutation method (Lapointe and Legendre 1992). The independent matrix variables in **X** are kept fixed with respect to one another during the permutations. Selection of predictors can be accomplished by forward selection, backward elimination, or a stepwise procedure. A phylogenetic tree, derived from marsupial brain morphology data (28 species), is compared to trees depicting the evolution of diet, sociability, locomotion, and habitat in these animals, as well as their taxonomy and geographical relationships. A model is derived in which brain evolution can be predicted from taxonomy, diet, sociability and locomotion ($R^2 = 0.75$). A new tree, derived from the "predicted" data, shows a lot of similarity to the brain evolution tree. The meaning of the taxonomy, diet, sociability, and locomotion predictors are discussed and conclusions are drawn about the evolution of brain and behavior in marsupials.

Key words.—Behavior, brain, distance matrices, double-permutation test, evolution, Mantel test, marsupials, multiple regression, permutation test, triple-permutation test.

Received April 19, 1993. Accepted January 12, 1994.

Relating brain evolution to the evolution of behavior in a group of animals is a difficult task for the neuroanatomist (Pirlot 1987); it is also a classical problem in behavioral studies (Masterton et al. 1976a,b). However, "predicting" brain evolution from behavioral characteristics is even more difficult biologically and methodologically. It implies the use of statistical methods designed specifically for evolutionary studies. The brain, behavior, and evolution problem has already been approached in several ways, either by studying whole-brain data with encephalization quotients (Jerison 1973) or progression indices (Stephan 1967), or by dividing the brain into a number of components and studying their relative development (Pirlot and Stephan 1970; Stephan and Pirlot 1970). Different statistical solutions are available. On the one hand, *raw quantitative data*

(whole brain, or components) may be compared to one or several behavior variables using classical statistical tools such as correlation, regression, or canonical analysis (Baron and Jolicoeur 1980; Jolicoeur and Baron 1980; Jolicoeur et al. 1984). On the other hand, one may tackle the more difficult problem of how the volumetric *evolution* of brain components may have accompanied or followed the evolution of behavior (Lapointe 1992; Lapointe and Legendre 1994)).

This second approach seems more interesting to us because one can directly compare the evolution of brain to the evolution of behavioral characteristics. It poses some new statistical problems, however: How to estimate brain evolution? How to estimate the evolution of the behavioral characters? Does the brain component data set contain phylogenetic information? Is it possible to model the evolution of brain from the evolution of the behavioral characteristics?

This paper has two purposes: (1) to introduce a new modeling method based on distance ma-

trices and appropriate for path-length matrices representing evolutionary trees; and (2) to apply this method to study the joint evolution of brain, behavior, and other characteristics in marsupials.

The new method basically consists of a multiple regression on distance matrices, with permutation testing of the regression parameters. When the dependent matrix variable represents a simple distance or similarity matrix, permutations are performed in the same manner as the Mantel (1967) permutational test. When it is an ultrametric matrix representing a dendrogram, the double-permutation method (Lapointe and Legendre 1990, 1991) is used. When it is a path-length matrix representing an additive tree (cladogram), we use the triple-permutation method (Lapointe and Legendre 1992). The independent matrix variables are kept fixed with respect to one another during the permutations. Selection of predictors can be accomplished by forward selection, backward elimination, or a stepwise procedure.

We show that this method may be used to model brain evolution in mammals. We estimate the best possible regression model to "predict" brain evolution from taxonomic and behavioral characters and use this model to explain quantitative brain differences among ecologically different species of the same taxonomic group.

MATERIALS AND METHODS
Neurological Data

For 28 species of marsupials, Pirlot (1981) determined the volume of each of 11 brain components (in cm³): bulbus olfactorius (BO), neocortex (NX), rhinencephalon (RH), schizocortex (SZ), septum (SE), striatum (ST), hippocampus (HI), diencephalon (DI), mesencephalon (ME), cerebellum (CE), and medulla oblongata (MO). Pirlot (1981) gives a full account of the laboratory methodology. These 28 species represent all three living orders of marsupials (Paucituberculata, Polyprotodonta, and Diprotodonta) and 8 families (Caenolestidae, Didelphidae, Dasyuridae, Peramelidae, Vombatidae, Macropodidae, Phalangeridae, and Petauridae), following Kirsch's (1977) classification. Because of the large amount of time and effort required, only a single brain has been analyzed per species. In addition, brain morphology data for two edentates (Pirlot 1980; Pirlot and Kamiya 1983) and two rodents (Ouedraogo 1974) were used as *outgroups* in the

phylogenetic analysis to determine the position of the "root" of the marsupial brain tree. To standardize the data for size, brain-component volumes were divided by the total brain volume. The choice of this standardization method is discussed elsewhere (Lapointe and Legendre 1995); because we wanted to compare brain organization, using relative brain volumes seemed appropriate to our purpose.

Estimating Brain Evolution

Several methods are available for reconstructing (estimating) phylogenies, using either presence-absence or quantitative data (Felsenstein 1982; Swofford and Olsen 1990; Penny et al. 1992). In the present example, Euclidean distances were computed among species based on the standardized brain component volumes, and the distance-based phylogenetic tree reconstruction method of De Soete (1983a) was applied to this matrix to obtain an additive tree. This method was preferred to other "distance methods" for tree reconstruction (such as those of Fitch and Margoliash or Cavalli-Sforza and Edwards), for it had been shown to produce better results (De Soete 1983b) because it constructs the topology of the tree and optimizes branch lengths simultaneously (Swofford and Olsen 1990). The root of the marsupial tree is located at the point where all marsupials separate from the members of the outgroup on the tree; the outgroup data will not be used any further, but the root of the tree was added to the data set and used in the following analysis. Notice that a phylogenetic tree can be represented as an additive tree (see Lapointe and Legendre 1992). It can also be represented by a path-length matrix, also called a "patristic matrix"; in such a matrix, the distance between any two taxa is the sum of the character-state changes along the path of branches connecting these two taxa ("patristic distance"). In our example, we will analyze path-length matrices of the 28 species of marsupials, plus the root of the tree considered as a twenty-ninth taxon; we will consider the path-length matrix representing the brain phylogenetic tree, on the one hand, and also path-length matrices representing the evolution of the various types of behavior, as described in the next paragraph.

Estimating Behavior Evolution

Data describing the evolution of four categories of behavior are used: habitat, diet, locomotion, and sociability (Lapointe and Legendre

1995). Before modeling, each behavioral variable had to be transformed into a tree, and then into a distance matrix; this implies rooting the tree, finding the sequence in which the states appeared during evolution, and finally, writing a full distance matrix for the species under study. This was accomplished as described in the following three steps.

1. Each behavioral characteristic was mapped onto a reference phylogeny of the marsupials, using program MacClade 3.0 (Maddison and Maddison 1992). Our reference phylogeny was the one proposed by Kirsch (1977), and completed (as in Lapointe and Legendre 1995) to include finer relationships proposed by various authors. Following Eisenberg (1980), the ancestral states of the behavioral characters chosen were: forest dweller, insectivore, low sociability, and climber-walker. Notice that the reference phylogeny has been formulated independently from our brain data; this is important to insure the independence of the independent (i.e., explanatory) and dependent (i.e., target) variables in our model (below).

2. The order in which the behavior states descend from one another along the trees was noted and transcribed in the form of the four following behavior-character-state trees (Lapointe and Legendre 1995):

Habitat selection: R → 1 → (2, 3); R = root; 1 = forest; 2 = ubiquitous; 3 = prairie.

Diet: R → 1 → (2, 3, 4 → 5); R = root; 1 = insectivorous; 2 = carnivorous; 3 = omnivorous; 4 = herbivorous; 5 = frugivorous-folivorous.

Sociability: R → 1 → 2; R = root; 1 = low; 2 = high.

Locomotion: R → 1 → [2a, 2b → (3, 4)]; R = root; 1 = climber-walker; 2a = walker (American); 2b = walker (Australian); 3 = climber; 4 = jumper.

In this simple coding, the locomotion-state tree, for instance, reads as follows: the root (R) leads to state 1, which in turn leads to states 2a or 2b; state 2b leads either to state 3 or to state 4.

3. The character-state trees were transformed into full patristic species matrices for the given behavior characteristics, using the method described by Legendre and Lapointe (1995) and illustrated in figure 1 for habitat selection.

Taxonomic and Geographic Data

Marsupials are sometimes classified into three living orders (Kirsch 1977): the Polyprotodonta, Diprotodonta, and Paucituberculata. We will use this marsupial classification (three orders) as an independent (i.e., explanatory) variable in our model. The hypothesis behind this inclusion is that brain evolution may have been influenced both by genetic pressures (represented by the major adaptations found in the order lineages) and by behavioral adaptations. To code taxonomy (three orders) into a distance matrix, members of the same order were given a distance of zero, whereas members of different orders were given a distance of one; the root of the tree was given a distance of one to members of all three living orders in the study.

Geographical information also will be used to explain brain evolution. Marsupials are found both in America and in Australia. Species found on the same continent were given a distance of zero, whereas species from different continents received a one. The ancestor of all marsupials was hypothesized in America (Clemens 1968, 1977) and distances between species and the root of the tree were coded accordingly.

Does the Brain Component Data Set Contain Phylogenetic Information?

Searching for relationships between brain and behavior evolution can succeed only if there is phylogenetic information to be found in brain morphology data. A simple way to reassure ourselves that such is indeed the case is to test for a significant relationship between the phylogenetic tree obtained from the brain data and the reference phylogeny of the marsupials on which we mapped the behavioral traits (above); as already noted, these two trees have been obtained independently. This was accomplished by way of the triple-permutation test for relationships between independently obtained phylogenetic trees of Lapointe and Legendre (1992).

Modeling the Evolution of Brain from the Evolution of the Behavioral Characteristics

As we have seen above, path-length distance matrices can be derived from the brain as well as from the behavior characteristic evolutionary trees. We propose here to use a regression model, based upon these path-length matrices, to "predict" distances among species brains along the evolutionary tree, from distances along the be-

Hypothesis for marsupial habitat evolution: R → 1 → (2, 3)

2 - Ubiquitous   3 - Prairie

1 - Forest

Root

Behavioral data:

Species	Habitat
1	1
12	2
15	2
16	1
20	3
21	1
23	3

Hypothesis in coded path-length form:

	Root	1	2	3
Root	0	1	2	2
1		0	1	1
2			0	2
3				0

Path-length matrix among species, describing the hypothesis for the evolution of habitat selection in marsupials

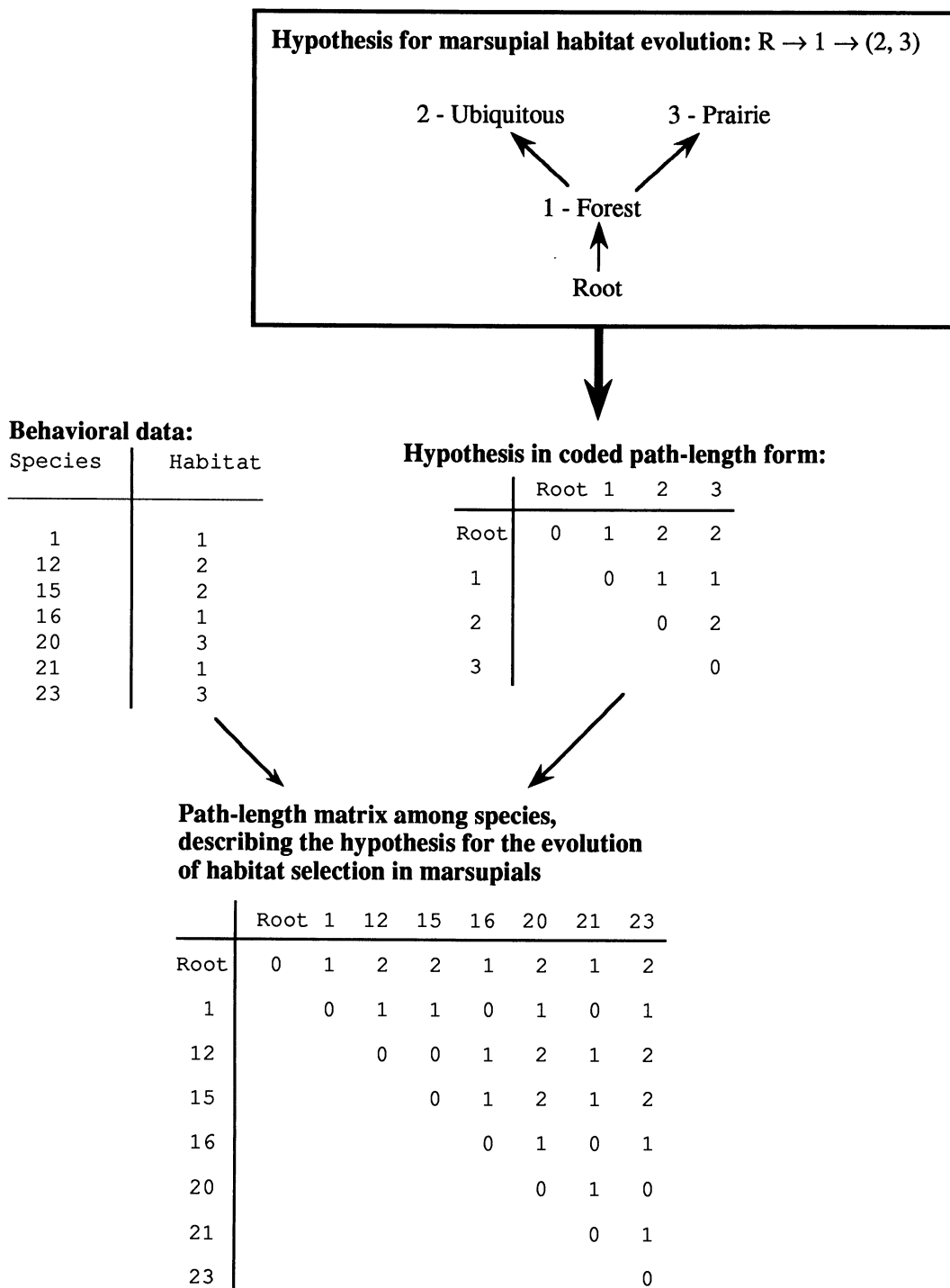|	Root	1	12	15	16	20	21	23
Root	0	1	2	2	1	2	1	2
1		0	1	1	0	1	0	1
12			0	0	1	2	1	2
15				0	1	2	1	2
16					0	1	0	1
20						0	1	0
21							0	1
23								0

FIG. 1. Coding of hypotheses of behavioral characteristic evolution (top) into path-length matrices (bottom) is illustrated using habitat preference characteristics for 7 of the 28 species of marsupials. Path lengths are measured by counting the number of arrows separating states in the hypothesis.
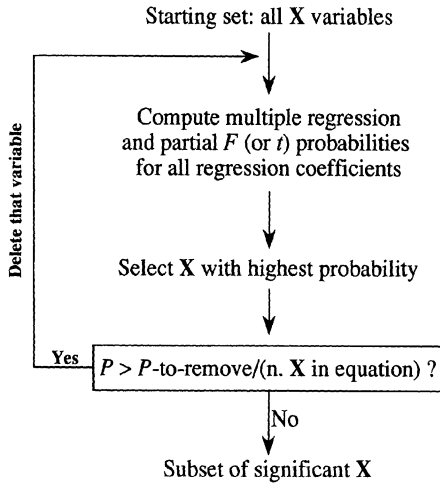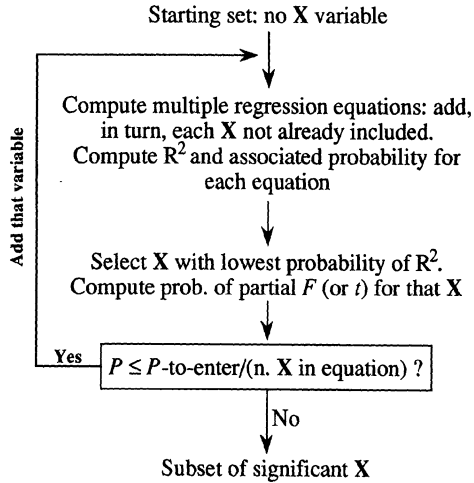
Backward elimination

Forward selection

FIG. 2. Backward and forward methods for selecting a subset of predictors. Selection is based on permutational probabilities.

havioral, taxonomic, and geographic trees; "prediction" has the same meaning here as in regression modeling. The brain evolution pathlength matrix is the dependent "variable," whereas the behavior evolution path-length matrices, taxonomic, and geographical distance matrices play the role of the independent "variables." Since all these matrices are usually symmetrical, the upper (or lower) half of each matrix, diagonal excluded, contains all the usable information, so that the computations can be limited to the upper (or lower) triangular parts. If this is not the case, the procedure can be performed on the full square matrix, excluding the diagonal, which contains, by definition, distances of zero between species and themselves. As in ordinary regression modeling, the model can be used to predict the position of new species on the brain evolution tree, if these species can first be positioned on the predictive behavioral trees.

Multiple Regression and Parameter Testing

Consider a series of distance matrices of size $(p \times p)$, p being the number of species, and assume that we want to predict (model) the distances in one of them (the dependent matrix Y) from the distances in the other matrices (the independent matrices X_1, X_2, ..., X_n). Although similarity matrices could be used just as well, we

will restrict the following development to distance (or dissimilarity) matrices for convenience. As we will see later, special types of distance matrices, such as ultrametric or patristic distances, will require special treatments. Assume also for convenience that each of these distance matrices is unfolded, row by row (or column by column), to form a long vector, excluding the diagonal terms which contain, by definition, distances of zero; if all matrices are symmetric, it is sufficient to unfold the upper (or lower) triangular parts only. We can call y the dependent unfolded matrix-variable Y, and X the table containing in successive columns the various independent matrix-variables X_1 to X_n. Multiple regression can be computed by least squares, using the standard equations to find the regression coefficients as well as the coefficient of multiple determination R^2. If the values in all matrix-variables have been standardized before the analysis, so that each has a mean of zero and a variance of one, then there is no intercept in the result; if this is not the case but standard partial regression coefficients are needed nevertheless, they can be computed from the conventional partial regression coefficients (e.g., Sokal and Rohlf 1981, chap. 16). Several authors have proposed in the past to compute multiple regression from distance matrices, and among them Hubert

and Golledge (1981), Smouse et al. (1986), Manly (1986), and Krackhardt (1988). The specific contribution of the present paper is to show how to test the significance of these parameters under the null hypothesis for different types of dependent matrix-variables; the corresponding permutation procedures are discussed below.

We may be interested in entering in the regression only the independent matrix-variables that contribute significantly to the explanation of the dependent matrix-variable, with due allowance for collinearity. As in ordinary multiple regression, the independent matrix-variables with the strongest contributions can be chosen either by forward selection or backward elimination, the two strategies not necessarily leading to the same selection of predictors. Here again, the procedure is complicated by the need for a permutation test that takes into account the distance-matrix nature of the data (fig. 2).

The *backward elimination procedure* is the easiest. All variables are initially included, and at each step, the independent matrix-variable whose partial regression coefficient has the highest P value is dropped, provided that this probability (Cooper 1968; Miller 1977) is also higher than a predetermined and Bonferroni-corrected, P-to-remove value. How to test the significance of the partial regression coefficients is discussed below. A *forward-selection* procedure can be defined as follows. Given an initial list of variables already in the model, multiple-regression equations are computed, adding each of the remaining variables in turn. The variable is selected whose multiple-regression equation provides the most significant R^2 coefficient. At the beginning of the selection process, the likelihood is high of encountering R^2 probabilities that are tied at the lowest probability value permitted by the number of permutations of the test (e.g., $P = 0.001$ after 1000 permutations). In such cases, an additional criterion is used: the variable is selected whose partial regression coefficient has the lowest P value. If a tie occurs again in the permutation probabilities, the *value* of the increment in R^2 is used as the final selection criterion. Selection of variables stops when the probability of the partial regression coefficient of the variable to be added exceeds a predetermined Bonferroni-corrected, P-to-enter value. (Notice that we cannot use here the usual forward-selection criterion based on the variable with the highest partial correlation, which is computationally simpler. The relation between the values of the partial

correlations and their associated probabilities may not hold, because of the special ways the permutations are performed in matrix regression; see below.) Forward selection, however, does not allow the removal of matrix-variables entered at a previous step that may turn nonsignificant at some later step following the inclusion of some other variable. Thus, a *stepwise procedure* should be preferred. It can be obtained by combining the forward and backward procedures: after each step of forward inclusion, the significance of all the variables included in the model is tested, and those that are not significant at the Bonferroni-corrected, P-to-remove significance level are excluded before the next forward selection step.

Testing the significance of the model parameters (the regression coefficients b_i, as well as the coefficient of multiple determination R^2) is the part where regression on distance matrices differs from ordinary multiple regression. We describe three cases, corresponding to different types of dependent matrix variables. In all cases, we consider the relations among the independent matrix-variables in the model fixed, so that they should not be permuted with respect to one another. A permutational procedure will be used.

In multiple regression on ordinary variables, the traditional tests of significance could be replaced by a permutation test, for instance, to solve problems of frequency distributions of the variables (lack of normality). In that case, repeatedly randomizing the values of the dependent variable, and recomputing the model and its parameters, would provide null distributions against which the significance of the parameters of the actual model could be tested. Notice the following computational shortcut: since matrix y only needs to be permuted in each iteration, while the independent vectors x_i remain fixed with respect to one another, then the expression $[X'X]^{-1}$ can be computed only once, and premultiplied to the product $[X'y]$ after each permutation of the dependent matrix to obtain vector b of the regression coefficients ($b = [X'X]^{-1} X'y$); this shortens the computations of a permutational test in an appreciable manner.

Matrix Case 1: Plain Dissimilarity Matrices.—In this paper, we are modeling from dissimilarity matrices. In the first situation that we consider, matrix variable y represents a plain dissimilarity possessing the properties of symmetry, positiveness and definiteness; if it also obeys the triangle inequality, it may be called a "distance" or a "metric-distance" matrix. See

Lapointe and Legendre (1992) for the properties of dissimilarity, distance, ultrametric, and patristic matrices. Parameters of the regression model could not be tested for significance in the traditional parametric way because matrix-variable data do not obey the most fundamental condition of parametric testing, which is the independence of the observations (distances, in the present case). Indeed, values in a distance matrix are not independent, but are related to one another, for instance, by the triangle's inequality with metric distances, or by the four-point metric with patristic distances. Therefore, a permutation test is in order. The one we propose is an extension of the Mantel (1967) test of matrix correspondence (e.g., for a detailed description, see Legendre and Fortin 1989); this extension was briefly suggested by Hubert and Golledge (1981) and by Smouse et al. (1986). In the Mantel test, a cross-product or correlation statistic is computed between the values in two square matrices, and its significance is tested by repeatedly permuting the objects (e.g., species) in one of the matrices—which amounts to simultaneously permuting the rows and columns—and recomputing the statistic; the significance of the statistic is assessed by comparing it to the distribution of values obtained from the permutations, which are considered as so many realizations of the null hypothesis. Notice that the permutation set—that is, the set of all distinguishable permutations according to the matrix permutation rule described above—is much smaller for the Mantel test ($p!$ for p species) than the set of all distinguishable random permutations of the $p(p-1)/2$ *values* in that distance matrix if they were free to be moved independently from one another between rows and columns (their number is

$$[p(p - 1)/2]!).$$

In the present case, the values of the independent matrix variable are considered fixed with respect to each other, but the dependent matrix variable is considered random with respect to them under the null hypothesis of exchangeability of rows and columns in the dependent distance matrix. Thus, a realization of the null hypothesis is obtained by permuting the objects at random in the dependent matrix variable and recomputing the regression and its coefficients; each such realization corresponds to the null hypothesis that the object values that gave rise to the dependent matrix variable are unrelated to the values that gave rise to the independent matrix variables taken together.

In practice, this amounts to computing first the real regression between matrix-variable brain and behavior evolution path-length matrices, as well as the F (or t) statistics associated with the partial regression coefficients. Then the species are permuted at random in the brain path-length matrix (in other words, the rows and columns of that matrix are permuted to the new order of the species), that matrix is unfolded again, and the regression equation is recomputed, as well as the partial F's (or t's). The permutation and recomputation steps are repeated many times (typically, 99, 999, or 9999 times), and a distribution is constructed from the values of R^2 and of the F's (or t's) associated with the partial regression coefficients. Following Hope (1968), the actual value of each statistic (R^2, or F's or t's of the partial regression coefficients) is added to the corresponding null distribution, which makes each test slightly conservative; its significance is assessed in the usual way: when a parameter value is so extreme by comparison with its null distribution that it is unlikely to have been generated under the null hypothesis (e.g.,, if it is as extreme as or more extreme than 95% of the randomly obtained values), it is labeled as significantly different from zero. Note that the permutation test on the R^2 coefficient can be conducted indifferently on the R^2 value itself, or on the associated multiple F statistic; the result is the same because F is a monotonically increasing function of R^2 for a constant number of data values and constant number of independent variables in the model. Such is not the case for the F (or t) statistics associated with the partial regression coefficients, however; this is why the probabilities are estimated from distributions of the F (or t) statistics in that case. To simplify computations, one can choose to compute a pseudo-t statistic at the end of each permutation, obtained by dividing the given partial regression coefficient by $\sqrt{(1 - R^2)}$. Since all the other terms in the expression of the usual t statistic are constant among permutations, the pseudo t's are linear transformations of the usual t's, which guarantees that the statistical decision is the same.

Matrix Case 2: Ultrametric Matrices. — When the dependent matrix-variable y is an ultrametric, representing a dendrogram whose ultrametric nature is believed to be an integral part of the problem, then the randomizations of matrix vector y should involve not only the position of the object labels of the dependent matrix Y, but also its topology (or *shape*) as given by the

fusion levels of the dendrogram. A double-permutation procedure (proposed by Lapointe and Legendre 1990, 1991) allows one to do that. The two randomizations are conducted as follows: first, the vector of fusion levels of the dendrogram is permuted at random and, using these fusion levels, the new matrix is filled with values in such a way that it is ultrametric; then the object labels are permuted at random and repositioned onto the rows and columns of the new ultrametric matrix. In the above-mentioned papers, we have shown that this constitutes a "uniform random-generation algorithm" *sensu* Furnas (1984), which means that all distinguishable ultrametric matrices are generated in an equally likely manner. After each such permutation of **y**, the regression is recomputed as in the plain dissimilarity case. For $p > 4$, the permutation set of a double permutation

$$\text{(size} = p! \, (p - 1)!/2^{p-1}:$$

Frank and Svensson 1981) is larger than the set of distinguishable permutations under matrix case 1 above because it combines a Mantel-like permutation of the object labels and a permutation of the dendrogram topology (Lapointe and Legendre 1991).

Matrix Case 3: Path-Length Matrices.—The third situation that we consider now is the case in which the dissimilarity matrix Y is a path-length matrix, obeying the four-point condition (Buneman 1971; Lapointe and Legendre 1992) and representing an additive tree (cladogram or phylogeny). Again, if we believe that its cladogram nature is an integral part of the problem, then the randomizations of matrix-vector **y** should involve all aspects of its structure. Lapointe and Legendre (1992) have shown that an equally likely permutation algorithm can be derived because a path-length matrix can be decomposed into an ultrametric and a star component; this last component contains the different branch lengths, which are unequal in a cladogram, leading to the various objects labeling the additive tree. Thus, the dependent matrix-vector **y** can be fully randomized by decomposing it into ultrametric and star components, separately randomizing the ultrametric component (as described in the previous paragraph) and the star component (which is a simple vector of values), reassembling the two randomized components into a new path-length matrix, to which randomized object labels are added. After each such permutation of **y**, the regression is recomputed as in the two previous cases; the permutation probabilities are obtained by comparing the actual val-

ues of the statistics (R^2, and F or t statistics associated with the partial regression coefficients) to distributions of parameter values obtained from a large number of these permutations. The permutation set of a triple permutation is larger than the set of distinguishable permutations under matrix cases 1 and 2 above (size = $(p!)^2 \, (p - 1)!/2^{p-1}$), because it actually combines a Mantel-like permutation of the object labels with a permutation of the star component and a permutation of the ultrametric component.

Following the traditional use of the Mantel test for comparing distance matrices, we have emphasized multiple linear regression and OLS parameter estimation as the mathematical instrument for studying the relation between a dependent distance matrix Y and several independent matrices X_1 to X_n. Before using this method, however, one should check the linearity of the relations between Y and each of the X_i matrices, because multiple linear regression is a linear model. This can easily be done by plotting scatter diagrams for pairs of unfolded distance matrices. In case of nonlinearity, the present permutational approach could readily be adapted to nonparametric or nonlinear regression methods.

RESULTS

The phylogenetic tree (fig. 3) derived from the brain volumetric data can be fully represented by a path-length matrix, which becomes our dependent matrix Y. The exercise then consists of modeling that tree using the behavioral, taxonomic, and geographic data described above, which also represent character evolution trees. In this example, each matrix X_i contains information about a single variable, but this need not be the case. The root of each tree is explicitly coded as an additional object in our data tables. The triple-permutation algorithm of Lapointe and Legendre (1992) does not require that, but we believed that this was a better way of handling the present example, and we have designed our permutational computer program accordingly. Selecting any other node of the tree for the decomposition (which is done only once), instead of the root, would lead to exactly the same probabilities (Lapointe and Legendre 1992). In this example, we are not only interested in getting a significant forecasting model (which would put the emphasis on the R^2 value only), but also in the significance of the matrix variables in the model; thus, a subset of predictors was selected using the various procedures described above
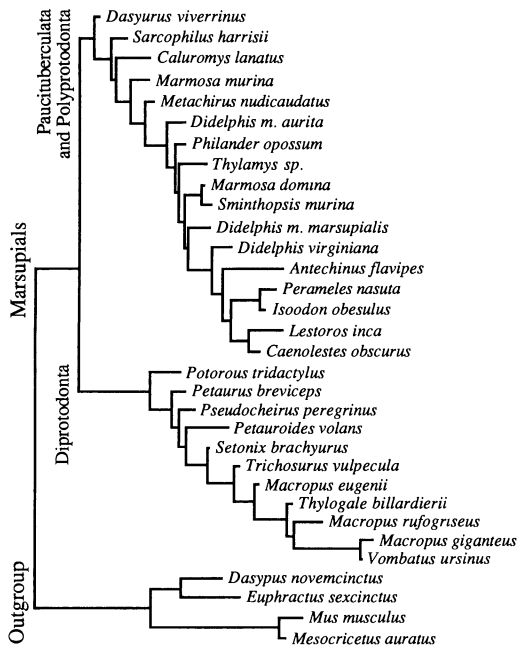
FIG. 3. Phylogeny of marsupials derived from 11 size-corrected brain component volumes.

FIG. 4. Phylogeny of marsupials derived from "predicted" distances computed from the multiple regression model.

(backward, forward, and stepwise), in the hope of obtaining an optimal set of variables for the evolution of brain in marsupials. All three selection procedures (backward, forward, stepwise) led to the same final selection of four variables; this need not always be the case. The steps of the backward selection procedure are presented in table 1. Thus the final model in standard form (where all matrix variables have been standardized before modeling) is the following:

$$Brain\ evolution = 0.629 \cdot Taxonomy$$
$$+ 0.192 \cdot Diet + 0.200 \cdot Sociability$$
$$+ 0.153 \cdot Locomotion\ R^2 = 0.7549$$

All partial regression coefficients of this model, as well as the coefficient of multiple determination (R^2), are significant. The model may then be used to "predict" the brain phylogeny of marsupials given information on their taxonomy, diet, social level, and mode of locomotion. However, one should not rely on this regression equation to fully reconstruct the values of the variables from which the model was derived; this exercise is done here simply to compare the actual brain phylogeny (fig. 3) to the tree derived from the values predicted by our regression model (fig. 4) and get a feeling of how good the model really
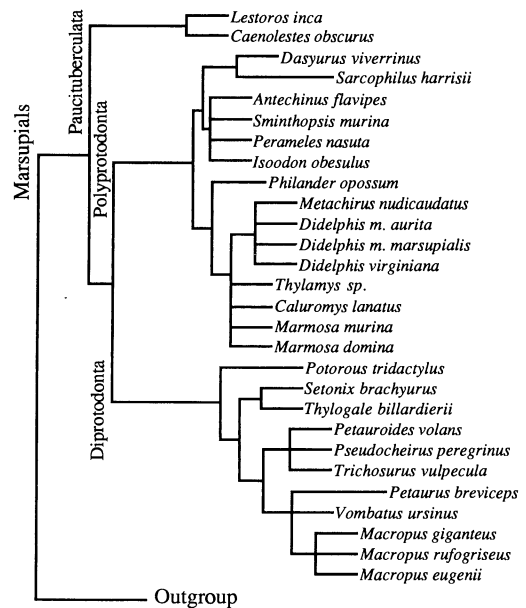
is. The matrix correlation between the path-length matrices associated to both trees is equal to 0.864 ($R^2 = 0.746$). This is different from the R^2 obtained from the regression model because the predicted distance values have been fitted to a tree before the comparison. Nevertheless, the trees are quite similar; this exercise clearly illustrates that one could use behavioral and taxonomic distance matrices to predict distances on a brain phylogeny.

DISCUSSION

We know that specific cortical patterns resulting from ontogeny in individual species are strictly controlled by genetics. This control apparently sets the rule of a primary repertoire of recognition by the cortical cell group through determining the basic intrinsic and extrinsic circuitries of such ensembles (Pirlot 1987). Selection and adaptation to particular environmental inputs may lead, however, to the formation of a secondary repertoire (Edelman and Mountcastle 1982). That is, the brain is not only the end-product of genetic pressures but also the result of environmental selective forces, and the prime instrument by which an animal can adapt to novel environmental conditions.

Because there exist reciprocal adaptations be-

TABLE 1. Backward elimination procedure for selecting an optimal subset of explanatory matrix variables for the marsupial brain evolutionary tree. The model parameters (Std. b) are standard partial regression coefficients. All probabilities (P) are one-tailed, computed after 9999 random triple permutations of the dependent matrix-variable; with this number of permutations, no probability can be smaller than $1/(9999 + 1) = 0.0001$. At each step, the variable chosen for elimination is marked with a dagger (†); the variable with the largest probability is eliminated if its probability is larger than the Bonferroni-corrected significance level $\alpha' = 0.05/[$(number of variables in the model at the given step)].

Matrix-variables	Step 1		Step 2		Step 3	
	Std. b	P	Std. b	P	Std. b	P
Taxonomy (orders)	0.626	0.0001	0.627	0.0001	0.629	0.0001
Diet	0.193	0.0001	0.192	0.0001	0.192	0.0001
Sociability	0.182	0.0001	0.178	0.0001	0.200	0.0001
Locomotion	0.172	0.0001	0.150	0.0001	0.153	0.0001
Habitat	0.060	0.1229	0.061	0.1227†	—	—
Geography	−0.033	0.2986†	—	—	—	—
R^2	0.7585		0.7580		0.7549	
Probability of R^2	0.0001		0.0001		0.0001	

tween an organ and its functions, one should expect a functional relation between brain and behavior. Different life habits imply specific sensory motor developments that modify brain size. Since the brain is the central organ controlling all relationships between an organism and its environment, it is very likely to be influenced by external forces. Behavioral drive can therefore really shape the brain of distant taxa to look similar. Species of mammals with similar habits (Baron 1977; Harvey et al. 1980; Harvey and Bennett 1983; Hofman 1983; Bennett and Harvey 1985; Gittleman 1986), diets (Pirlot and Stephan 1970; Stephan and Pirlot 1970; Pirlot and Pottier 1977; Eisenberg and Wilson 1978; Pirlot 1981), habitats (Bauchot et Stephan 1968; Baron 1981; Pirlot and Kamiya 1982), or demographic strategies (Eisenberg and Wilson 1981) are known indeed to exhibit similar brain volumes. But is the behavioral drive responsible by itself for brain development? A correlation between a specific behavioral trait and a particular brain type does not imply causality. Both can be related to a common covariate: phylogeny.

Because functional adaptations to the environment result from past evolutionary events, including selection (Atchley et al. 1984), it would be very restrictive to invoke a single cause to explain brain development. One must consider, as we have shown, both extrinsic (i.e., environmental) and intrinsic (i.e., genetic) pressures when studying brain development. The model that we arrived at illustrates this fact: the volumetric evolution of the marsupial brain, as captured by our 11 neuroanatomical variables, is significant-ly related to the taxonomy, as well as to the evolution of diet, sociability, and locomotion; these descriptors represent either adaptations to the environment or genetic constraints.

Taxonomy is obviously the most important feature ($b = 0.629$) related to brain differences among marsupials. The phylogeny reconstructed from brain data clearly separates diprotodonts, on the one hand, from polyprotodonts and paucituberculates on the other (fig. 3). This shows that the brain has not evolved according to what would be expected from the marsupial phylogeny (in which some polyprotodonts are members of the same clade as the diprotodonts) and illustrates the convergence of brain organization between American and Australian polyprotodonts.

Dietary information then becomes important to explain smaller differences ($b = 0.192$) not accounted for by taxonomy. We can interpret the separation of marsupials on the basis of brain characters as related to dietary differences; diprotodonts are herbivorous or frugivorous, whereas polyprotodonts and paucituberculates are either insectivorous, omnivorous, or carnivorous. The finer relationships within orders of marsupials are probably more correlated to diet than to taxonomy, which explains only the major dichotomy at the base of the brain tree (fig. 3); it is also reflected in the "predicted" phylogeny (fig. 4).

Social level is also significantly improving the regression model at a smaller scale ($b = 0.200$). Low sociability is a characteristic of primitive American marsupials of the Polyprotodonta and Paucituberculata orders. Among the Australian

polyprotodonts, social species are grouped to-gether in the brain tree (fig. 3); this is also the case for diprotodont marsupials. (Details are giv-en in Lapointe and Legendre 1995.)

Finally, locomotion has a significant contri-bution to the regression model ($b = 0.15$). An-cestors of the marsupials were probably climber-walkers (Eisenberg 1980). Primitive paucituber-culates and American polyprotodonts (Didel-phidae) are semiarboreal (climber-walkers) or terrestrial animals (walkers), whereas the Aus-tralian polyprotodonts (Dasyuridae) evolved into walking marsupials. Diprotodont radiation led to jumping kangaroos, walking semifossorial wombats, and climbing or even gliding possums.

The standard partial regression coefficients, as-sociated with the three significant independent matrix variables, seem to be related to the hi-erarchical level at which those characters explain brain similarities. The major dichotomy on the brain phylogeny (fig. 3) is taxonomic, whereas the finer relationships are more likely to be re-lated to diet and social level. At low taxonomic level, differences between species are accounted for mostly by behavioral adaptations in response to environmental pressures. Among-order com-parisons, however, are probably reflecting ge-netic differences among marsupials. This general fact, that closely related species are more likely to exhibit similar development than distant spe-cies, is a consequence of the "taxon-level effect" (Pagel and Harvey 1988). It implies that the brain-behavior evolution cannot be properly studied without considering also the phylogenetic prox-imities among species. This is what the "com-parative method" is trying to accomplish (Har-vey and Pagel 1991); it considers the phylogeny fixed and controls for its effect to study the cor-relation among traits. For instance, the focus of a comparative study might be to test that eco-logical and brain characters remain correlated to each other, even when one corrects for phylo-genetic distances. The question that the present paper tries to answer is quite different: our null hypothesis is that phylogenetic and ecological distances can be combined to predict the dis-tances depicting the evolution of the brain. The structure of the actual phylogeny is not preserved through our permutational procedure; on the contrary, a new (permuted) phylogenetic struc-ture is generated every time, under the null hy-pothesis. With the double-permutation method, the permuted phylogenies are generated under a molecular-clock hypothesis (all branches are as-

sumed to evolve at the same rate), whereas in the triple-permutation method, the permuted phylogenies have unequal rates of evolution in the different branches; the Mantel permutation method does not preserve nor create any phy-logenetic structure, of course.

The methodological contribution of this paper provides a new way to evaluate the joint evo-lution of structure and function. This method may also be useful for other problems, where several types of distance matrices are simulta-neously considered: coevolution problems, such as hosts and parasites, with due allowance for environmental and geographic information; in ecology, epidemiology or behavior studies, ex-planation of the community, disease, or behavior structure by different types of biotic, abiotic and spatial variables; in population studies, expla-nation of the genetic structure of populations by environmental, linguistic and geographic factors. Problems of the same type are found in geog-raphy, sociology, psychometry, political science, econometrics, and so on.

A computer program is available from Phi-lippe Casgrain to carry out the three types of regression modeling from dissimilarity matrices described in this paper.

LITERATURE CITED

Atchley, W. R., B. Riska, L. A. P. Kohn, A. A. Plum-mer, and J. J. Rutledge. 1984. A quantitative ge-netic analysis of brain and body size associations, their origin and ontogeny: data from mice. Evolu-tion 38:1165–1179.

Baron, G. 1977. Allometric development of the ves-tibular nuclei in relation to flight behavior among Chiroptera. Zoologischer Anzeiger 199:227–250.

———. 1981. Niches écologiques et développement morphométrique des centres visuels chez les pri-mates. Revue Canadienne de Biologie 40:85–89.

Baron, G., and P. Jolicoeur. 1980. Brain structure in Chiroptera: some multivariate trends. Evolution 34:386–393.

Bauchot, R., and H. Stephan. 1968. Etudes des mod-ifications encéphaliques observées chez les insec-tivores adaptés à la recherche de nourriture en mi-lieu aquatique. Mammalia 32:228–275.

Bennett, P. M., and P. H. Harvey. 1985. Relative

brain size and ecology in birds. Journal of Zoology (London) Series A 207:151–169.

Buneman, P. 1971. The recovery of trees from measures of dissimilarity. Pp. 387–395 in: F. R. Hodson, D. G. Kendall, and P. Tautu, eds. Mathematics in archeological and historical sciences. Edinburgh University Press, Edinburgh.

Clemens, W. A. 1968. Origin and evolution of marsupials. Evolution 22:1–18.

———. 1977. Phylogeny of the marsupials in Stonehouse. Pp. 51–68 in: B. Gilmore and D. Gilmore, eds. The biology of marsupials. University Park Press, Baltimore.

Cooper, D. W. 1968. The significance level in multiple tests made simultaneously. Heredity 23:614–617.

De Soete, G. 1983a. A least squares algorithm for fitting additive trees to proximity data. Psychometrika 48:621–626.

———. 1983b. On the construction of "optimal" phylogenetic trees. Zeitschrift für Naturforschung, Section C (Biosciences) 38:156–158.

Edelman, G. M., and V. B. Mountcastle. 1982. The mindful brain. MIT Press, Cambridge, Mass.

Eisenberg, J. F. 1980. Biological strategies of living conservative mammals. Pp. 13–30 in: K. Schmidt-Nielsen, L. Bolis, and C. R. Taylor, eds. Comparative physiology: primitive mammals. Cambridge University Press, Cambridge.

Eisenberg, J. F., and D. E. Wilson. 1978. Relative brain size and feeding strategies in the Chiroptera. Evolution 32:740–751.

———. 1981. Relative brain size and demographic strategies in didelphid mammals. American Naturalist 118:1–15.

Felsenstein, J. 1982. Numerical methods for inferring evolutionary trees. Quarterly Review of Biology 57:379–404.

Frank, O., and K. Svensson. 1981. On probability distributions of single-linkage dendrograms. Journal of Statistical Computation and Simulation 12:121–131.

Furnas, G. W. 1984. The generation of random, binary unordered trees. Journal of Classification 1:187–233.

Gittleman, J. L. 1986. Carnivore brain size, behavioral ecology, and phylogeny. Journal of Mammalogy 67:23–36.

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

Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford.

Harvey, P. H., T. H. Clutton-Brock, and G. M. Mace. 1980. Brain size and ecology in small mammals and primates. Proceedings of the National Academy of Sciences, USA 77:4387–4389.

Hofman, M. A. 1983. Energy, metabolism, brain size and longevity in mammals. Quarterly Review of Biology 58:495–512.

Hope, A. C. A. 1968. A simplified Monte Carlo significance test procedure. Journal of the Royal Statistical Society, Series B 30:582–598.

Hubert, L. J., and R. G. Golledge. 1981. A heuristic method for the comparison of related structures. Journal of Mathematical Psychology 23:214–226.

Jerison, H. J. 1973. Evolution of the brain and intelligence. Academic Press, New York.

Jolicoeur, P., and G. Baron. 1980. Brain center correlations among Chiroptera. Brain, Behavior and Evolution 17:419–431.

Jolicoeur, P., P. Pirlot, G. Baron, and H. Stephan. 1984. Brain structure and correlation patterns in Insectivora, Chiroptera and Primates. Systematic Zoology 33:14–29.

Kirsch, J. A.W. 1977. The comparative serology of Marsupialia, and a classification of marsupials. Australian Journal of Zoology, Supplementary Series 52:1–152.

Krackhardt, D. 1988. Predicting with networks: nonparametric multiple regression analysis of dyadic data. Social Networks 10:359–381.

Lapointe, F.-J. 1992. On the congruence of brain evolution with taxonomic distances and eco-ethological affinities in bats: a statistical evaluation. Ph.D. diss. Département de sciences biologiques, Université de Montréal.

Lapointe, F.-J., and P. Legendre. 1990. A statistical framework to test the consensus of two nested classifications. Systematic Zoology 39:1–13.

———. 1991. The generation of random ultrametric matrices representing dendrograms. Journal of Classification 8:177–200.

———. 1992. A statistical framework to test the consensus among additive trees (cladograms). Systematic Biology 41:158–171.

———. 1995. Evolution of the marsupial brain: Does it reflect the evolution of behavior? In T. Cabana, ed. The animal and its environment: a tribute to Paul Pirlot. Orbis, Frelighsburg. In press.

Legendre, P., and M.-J. Fortin. 1989. Spatial pattern and ecological analysis. Vegetatio 80:107–138.

Legendre, P., and F.-J. Lapointe. 1995. Assessing hypotheses of behavioral evolution against brain morphology data. Brain, Behavior and Evolution 45:110–121.

Maddison, W. P. and D. R. Maddison. 1992. MacClade—Analysis of phylogeny and character evolution, Version 3. Sinauer, Sunderland, Mass.

Manly, B. F. J. 1986. Randomization and regression methods for testing for associations with geographical, environmental, and biological distances between populations. Researches on Population Ecology (Kyoto) 28:201–218.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209–220.

Masterton, R. B., M. E. Bitterman, C. B. G. Campbell, and N. Hotton, eds. 1976a. Evolution of brain and behavior in vertebrates. Lawrence Erlbaum, Hillsdale, N.J.

Masterton, R. B, W. Hodos, and H. J. Jerison, eds. 1976b. Evolution, brain and behavior: persistent problems. Lawrence Erlbaum, Hillsdale, N. J.

Miller, R. G., Jr. 1977. Developments in multiple comparisons. Journal of the American Statistical Association 72:779–788.

Ouedraogo, M. 1974. Morphologie comparée du cerveau de quelques rongeurs. M.S. thesis. Université de Montréal, Montréal.

Pagel, M. D., and P. H. Harvey. 1988. The taxon-level problem in the evolution of mammalian brain size: facts and artifacts. American Naturalist 132: 344–359.

Penny, D., M. D. Hendy, and M. A. Steel. 1992. Progress with methods for constructing evolutionary trees. Trends in Ecology & Evolution 7:73–79.

Pirlot, P. 1980. Quantitative composition and histological features of the brain in two South American edentates. Journal für Hirnforschung (Berlin) 21:1–9.

———. 1981. A quantitative approach to the marsupial brain in an eco-ethological perspective. Revue Canadienne de Biologie 40:229–250.

———. 1987. Contemporary brain morphology in ecological and ethological perspectives. Journal für Hirnforschung (Berlin) 28:145–211.

Pirlot, P., and T. Kamiya. 1982. Relative size of brain and brain components in three gliding placentals (Dermoptera: Rodentia). Canadian Journal of Zoology 60:565–572.

———. 1983. Quantitative brain organization in anteaters (Edentata-Tubilidentata). Journal für Hirnforschung (Berlin) 24:677–689.

Pirlot, P., and J. Pottier. 1977. Encephalization and quantitative brain composition in bats in relation to their life-habits. Revue Canadienne de Biologie 36:321–336.

Pirlot, P., and J. Stephan. 1970. Encephalization in Chiroptera. Canadian Journal of Zoology 48:433–444.

Smouse, P. E., J. C. Long, and R. R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Systematic Zoology 35:627–632.

Sokal, R. R., and F. J. Rohlf. 1981. Biometry. W. H. Freeman, San Francisco.

Stephan, H. 1967. Zur Entwicklungshöhe der Insektivoren nach Merkmalen der Gehrins und die Definition des "basalen Insektivoren". Zoologischer Anzeiger 179:177–199.

Stephan, H., and P. Pirlot. 1970. Volumetric comparisons of brain structures in bats. Zeitschrift für Zoologische Systematik und Evolutionsforschung 8:200–236.

Swofford, D. L., and G. J. Olsen. 1990. Phylogeny reconstruction. Pp. 411–501 in D. M. Hillis and C. Moritz, eds. Molecular systematics. Sinauer Associates, Sunderland, Mass.

Corresponding Editor: J. B. Walsh