Estimation of Ancestral States of Continuous Characters: A Computer Simulation Study

Emília P. Martins

Department of Biology, University of Oregon, Eugene, OR 97403, USA; E-mail: emartins@evolution.uoregon.edu

In systematics, we usually estimate ancestral phenotypes of morphological and molecular characters measured as categorical or "state" variables. Ancestral estimation has also been important in the studies of animal behavior, physiology, ecology, and other areas of biology, where the characters are usually measured as continuous rather than categorical variables. In the past, the most common way of estimating the ancestral states of continuous characters was to use a parsimony algorithm (Farris, 1970; Swofford and Maddison, 1987; Maddison, 1991). Recently, two new methods (Martins and Hansen, 1997; Schluter et al., 1997) have been developed for estimating the ancestral states of continuous characters on a phylogeny. For several reasons, the theory underlying these methods provides a major leap forward from the traditional parsimony methods. In the current study, I use computer simulation to examine the statistical properties of these methods and to determine how much of an improvement these methods actually provide when applied to realistic data.

The sum of squared changes parsimony algorithm (SSP: Huey and Bennett, 1987; Maddison, 1991; McArdle and Rodrigo, 1994) estimates the phenotype of each ancestor as a weighted average of all the phenotypes measured for extant taxa. The weights used in calculating averages correspond to the phylogenetic distance between each extant taxon and the ancestor being estimated. Huey and Bennett (1987) applied this approach, iteratively calculating each ancestral state as the average of its three nearest neighbors until estimates converged on stable solutions. Maddison (1991) provided a recursive equation for getting the estimates and explored the assumptions of the method.

Specifically, he showed that SSP is a statistically reasonable approach very similar to using least-squares regression to estimate ancestral states. McArdle and Rodrigo (1994) showed how SSP could be described as a system of linear equations and thus provided a direct algorithm that can be used to reduce computational time.

Linear or Wagner parsimony (Farris, 1970; Swofford and Maddison, 1987) is a similar method that lacks the above statistical development and justification. Instead of minimizing the sum of squared evolutionary changes on the phylogeny, it minimizes the simple sum of changes. In a computer simulation study, Butler and Losos (1997) found that the two methods produce rather different results and that linear parsimony gave worse estimates under the type of phenotypic evolution (Brownian motion) considered in their simulation procedure. I do not consider linear parsimony in the current study.

One weakness of SSP is that it does not include any measure of estimation accuracy. For example, the state at the root of a phylogeny is known with less accuracy than a state nearer the tips, simply because the longer time since the root makes estimation more difficult. With SSP, we have no way of showing that difference in accuracy. Similarly, when we argue that a lot of phenotypic change followed a particular speciation event, we have no way of determining whether that evolutionary change is *significantly* large or whether the difference from other changes is small enough to be explained by estimation error. Even when we argue that one ancestor had a phenotype different from that of its descendants, we have no way to determine whether the two phenotypes are significantly different from each

other in a statistical sense. The problem appears in many contexts and when answering many types of evolutionary questions.

Schluter et al.'s (1997) maximum likelihood technique (ML) begins by yielding the same estimates of ancestral states as SSP. It does so, however, by using a rather different approach: an extension of the evolutionary model underlying Felsenstein's (1985) independent contrasts technique. The difference between SSP and ML is similar to the difference between least-squares and maximum likelihood regression, with the two providing identical answers in most situations. In this case, however, the ML method is a major improvement because it also provides estimates of the standard errors of those ancestral states. These standard errors can be used to develop hypothesis tests, confidence intervals, and other statistical tests. We can use these standard errors to determine, for example, whether a particular evolutionary change is large, whether an ancestor differed from its descendant, and roughly how much confidence we can place in our estimates of a particular ancestral state. Standard error estimates, however, are useful only if they are reasonably accurate, and we do not yet know whether the results of ML are so.

Both SSP and ML methods suffer from another potential problem—that of assumptions. SSP is really an algorithm, rather than a statistical model, and it is not immediately clear what its assumptions are. For example, SSP probably assumes that the measured variables are normally distributed—but this is not yet known. Maddison (1991) showed how SSP is similar to Felsenstein's (1985) approach, which assumes primarily that phenotypic evolution occurs as if by Brownian motion. The ML method (which, again, gives identical ancestral estimates) explicitly assumes that evolution follows a Brownian motion process. Thus, if evolution occurs as if by Brownian motion, both SSP and ML methods are likely to yield reasonable estimates of evolutionary ancestors. Brownian motion is a powerful mathematical process commonly used by population geneticists to describe characters evolving via random genetic drift or via directional selection when the direction of selection shifts back and forth at random. This process is often sufficient to describe many traits undergoing different types of evolution (Schluter et al., 1997). Unfortunately, sometimes it may not be a good model to describe the evolution of behavior, life-history traits, and many other characters that are subject to selective constraints or that respond rapidly to changes in the environment (Hansen and Martins, 1996). Whether SSP or Schluter et al.'s (1997) methods will perform well with these sorts of characters is not known.

In Martins and Hansen (1997), we developed a procedure that, like ML, also yields ancestral estimates with standard errors. We began by adopting a general linear model approach, similar to the linear equations suggested by McArdle and Rodrigo (1994), and then (among other things) developed specific procedures for estimation of ancestral states by using generalized least squares (GLS). Our GLS method is quite broad and can be used to estimate a number of things other than ancestral states. It also allows the user flexibility in the choice of microevolutionary assumptions. For example, instead of assuming the standard Brownian motion model, the GLS method gives the researcher the flexibility to choose among a family of possible microevolutionary models, including models of stabilizing selection, evolution in a fluctuating environment, and "burst-like" change (Hansen and Martins, 1996). The GLS method also allows for the incorporation of within-species variation.

In one simple form (assuming Brownian motion evolution and no within-species variation), our GLS method will produce ancestral state estimates that are identical to those found by ML and SSP. This method is termed "GLS-linear," because it assumes a linear or clock-like diversification of phenotype through time. As illustrated in Martins and Lamont (1998), however, other forms of our GLS method can produce very different results, including ancestral estimates and standard errors outside the range of the original data. In this study, we focus on an alternative form of this method, termed "GLS-exponential," which is appropriate for use with traits that are thought to be evolving under certain constraints such as weak sta-

bilizing selection (see Martins and Lamont [1998] for details).

In this study, I use computer simulation to consider the statistical performance of SSP, ML, and these two forms of GLS methods under different microevolutionary scenarios. We know from the theoretical development of the methods that, when phenotypic evolution is well described by the most commonly used model of evolutionary change (Brownian motion), then SSP, ML, and GLS-linear are expected to give reasonable estimates of the ancestral state at each node of the phylogeny. We also know that those estimates should be virtually identical. The ML and GLS-linear methods are also expected to give reasonable estimates of the standard errors about ancestral states; again, these standard errors should be virtually identical. The main difference between methods is expected when dealing with characters resulting from constrained evolution (e.g., under an Ornstein–Uhlenbeck [OU] model of stabilizing selection). In this case, GLS-exponential is expected to give better estimates, if the actual type of evolution is known, simply because we can match the assumptions of the method with the reality of the data. We do not know, however, whether these method differences are large enough to be meaningful with real data. With computer simulation, we can determine how well the four methods are likely to perform on an absolute scale and whether the theoretical improvements provided by the ML and GLS methods are likely to be useful.

METHODS

Ancestral State Estimation

Because SSP and ML always give identical estimates of ancestral states, I did not calculate SSP estimates separately. ML and SSP estimates of ancestral states were obtained by using a slightly modified version of the ANCML program (Schluter, 1997). Both methods calculate ancestral state estimates that minimize the sum of squared evolutionary changes across the phylogeny as a whole. ML does so by using techniques that also yield estimates of the standard error for those states as well as estimates of the β pa-

rameter of its underlying Brownian motion model.

The GLS method (Martins and Hansen, 1997) uses the model $\mathbf{A} = \mathbf{WY}$ to estimate ancestral states, where $\mathbf{Y}$ is a vector of the measured species data, $\mathbf{W}$ is a matrix describing the phylogeny and chosen model of phenotypic evolution, and $\mathbf{A}$ is a vector of the resulting ancestral state estimates. Standard errors for those ancestral states were obtained by using standard generalized least-squares techniques (see Martins and Lamont [1998] for a worked example) and a weighting matrix derived (as in $\mathbf{W}$ above) from the phylogeny and model of evolutionary change. The method was implemented with the "ancestor" module of COMPARE (Martins, 1998).

The GLS method was implemented with both "linear" and "exponential" versions of the method (Martins and Hansen, 1997). GLS-linear assumes that the phenotypic similarity between taxa decreases linearly with phylogenetic divergence, such as is observed under Brownian motion. Other microevolutionary models involving random fluctuations of either phenotype or selective forces can produce a similar pattern (Hansen and Martins, 1996). The results of this linear version of the method are expected to be quite similar to those obtained with the ML method. The linear model requires the estimation of only a single parameter (σ^2), which should be proportional to the β parameter estimated by the ML method. GLS-exponential assumes an exponential decrease of phenotypic similarity with phylogenetic distance, as is expected when there is a constraint on evolution (e.g., stabilizing selection; Hansen and Martins, 1996). The exponential model requires the use of a second parameter, α, which describes the strength of the restraining force. Rather than estimate this parameter, I applied a series of possible α values (e.g., as suggested in Martins and Hansen [1997] and Martins and Lamont [1998]) to determine the importance of estimating this parameter accurately. Results are thus presented for GLS-linear and for several versions of GLS-exponential, only one of which assumes the correct value of α. The GLS method also in-

corporates measures of within-taxon variability. For the sake of simplicity, within-taxon variation was assumed to equal zero in all cases.

Computer Generation of Data

Data were generated by simulating the evolution of phenotypes along a phylogeny under a specified model. All data were generated by using the "simulate" module of COMPARE (Martins, 1998). I used two possible phylogenies: (1) a phylogeny of 9 taxa of *Cyclura* iguanas (Malone and Davis, unpubl.; applied by Martins and Lamont, 1998) to estimate the ancestral states of behavioral displays, and (2) a phylogeny of 42 species of *Sceloporus* lizards (Larsen and Tanner, 1974; applied in Martins, 1993). I do not make any claims about the accuracy of these phylogenies to describe the evolutionary history of lizard taxa. They are simply examples of the types of phylogeny commonly used in comparative analyses. Together these two phylogenies provide a total of 49 putative ancestors for which states can be estimated.

In all cases, I began at the root of the tree with an ancestral phenotype of zero. At each subsequent unit of time, I chose an evolutionary change from a specified distribution of possible changes and added that value to the previous value of the ancestral state. This process was continued along the phylogeny, leading from root to tips. Whenever a speciation event occurred, the process was split, with independent changes being chosen for each of the two daughter branches.

I generated 1,000 data sets on each phylogeny under each model of phenotypic evolution. To begin, I applied the Brownian motion model of phenotypic evolution that has been used in most previous simulation studies of phylogenetic comparative methods (e.g., Martins and Garland, 1991) and which is assumed by both ML and linear-GLS. Under this model, evolutionary changes (δ) at each unit of time were chosen from a normal distribution ($\delta = N$, where N is a random normal deviate with mean = 0 and variance = σ^2). The variance (σ^2) cancels out in calculating relationships between two correlated characters evolving together on a phylogeny and has thus been ignored in previous computer simulation studies of phylogenetic comparative methods (e.g., Martins and Garland, 1991). Although it does not cancel out in estimation of ancestral states, some preliminary runs suggested that it had little if any impact on the overall results. Thus, I generated data by using an arbitrary value of $\sigma^2 = 1$.

In addition, I generated data under an OU or "rubber band" process. This process has been used previously to describe the evolution of phenotypes under weak stabilizing selection with a constant optimum (e.g., Felsenstein, 1988; Lande, 1976, 1979; Hansen and Martins, 1996). It requires the use of the above σ^2 parameter and also a second parameter, α, which describes the strength of the restraining force. As above, evolutionary changes were added to the current state of the trait at each unit of time. To model OU evolution, each change (δ) was expressed as a random number minus a factor describing the restraining force ($\delta = N - \alpha x_t$ where N is a random normal deviate with mean = 0 and variance = σ^2, and x_t is the state of the character at the preceding unit of time). To test the effects of varying the strength of the restraining force, I generated data under several different values of α (0.01, 0.05, 0.1, 0.5, 1, and 2 for the 42-taxon phylogeny; 0.01, 1, 2, and 5 for the 9-taxon phylogeny).

Statistical Analyses

Use of any particular method to estimate ancestral states for a single run of the simulation led to 1,000 estimates of the state (X) at each node of the phylogeny and 1,000 estimates of the standard error of that nodal estimate (S). To determine the relative accuracy of each method in terms of ancestral state estimation, I compared the estimates for each node with the values obtained directly from the simulation procedure (A). First, I calculated the bias ($X - A$) for each run of the simulation and node of the phylogeny and also the mean bias across 1,000 runs. I also calculated a Pearson product-moment correlation [Corr.] between the ancestral state estimated by a particular method and the ancestral state obtained from the simulation procedure: Corr[X, A].

To determine the relative accuracy of standard error estimates, I calculated the squared deviation of each estimate from the true value of the ancestor: $(X - A)^2$. The result is the true standard error of a method. I used the difference between this squared deviation and the standard error estimated by the method (S) as a measure of the bias in standard error for each method: $(X - A)^2 - S$. For simplicity, the mean of this quantity across 1,000 runs is termed the mean error bias (MEB, comparable with mean squared error). Again, I also calculated a Pearson product-moment correlation between the standard error estimated by a particular method and the true standard error obtained from the simulation procedure: $\mathrm{Corr}[(X - A)^2, S]$.

RESULTS

As expected, for data generated under a Brownian motion model of evolution, ancestral state estimates were virtually identical for linear-GLS, ML, and SSP (Fig. 1, Table 1). These estimates were reasonably good in an absolute sense, especially for very recent ancestors. The mean bias for all three methods was usually negligible, but occasionally became quite large, ranging up to 4.0 times the magnitude of the ancestral state. Similarly, the estimated states for almost all the ancestors were highly correlated with the true value for the ancestor ($r = 0.90$–0.96), except for two nodes on the 9-taxon tree (the two most distant from all extant taxa), which were poorly estimated ($r = 0.50$ and 0.54).

Estimates of error provided by GLS-linear and ML were substantially worse, with the squared difference between estimated and true values for either method (MEB) ranging up to 223 and correlation coefficients between estimated and known standard errors ranging between 0.30 and 0.40 (Fig. 2, Table 1). Moreover, the true standard error was always underestimated by the phylogenetic methods, such that values of MEB were always positive. Although error estimates provided by the two methods occasionally differed from each other for a particular data set, these differences were trivial when summed across 1,000 runs (Table 1).

MEB values of the two methods were always within 0.2 of each other, and differences between methods in correlations with the true values were always > 0.01.

GLS-linear, ML, and SSP all performed substantially worse with data generated under the constrained OU model (Table 1, Fig. 1). Mean bias in estimating ancestral states occasionally reached as high as 5 times the magnitude of the state itself, and correlations with the true ancestral states were never far from zero (Table 1). Estimates of error were less biased, being sometimes underestimated and sometimes overestimated. Nevertheless, correlations with truth were close to zero, suggesting that the error estimates were quite poor (Table 1, Fig. 2).

Unfortunately, use of GLS-exponential did little, if anything, to improve the situation with OU-generated data. Given the correct value of α and OU-generated data, GLS-exponential tended to give better estimates of ancestral states (Table 1, Fig. 1). Mean bias was slightly lower and the correlation between estimated and true ancestral states was somewhat higher. These differences, though, were exceedingly small for the 42-taxon tree and negligible for the 9-taxon tree. Moreover, although estimates of error were better for GLS-exponential on the 9-taxon tree, they were actually worse for this method on the 42-taxon tree (Table 1, Fig. 2).

Unfortunately, given incorrect values of α, the exponential-GLS method could also give very poor results for the 42-taxon tree (see Table 1, GLS-exponential performing with Brownian motion data, Figs. 1 and 2). With incorrect values of α, mean bias was still as much as 5 times the magnitude of the ancestral state, and correlations with true ancestral states usually ranged between -0.42 and 0.56 (not better than SSP, ML, or GLS-linear). MEB ranged up to 3,100, and correlations with the true error ranged between -0.46 and 0.46 (the same as when α was known). Surprisingly, the correlation between the true and the estimated standard errors was occasionally substantially greater for these methods than for the linear methods. Again, virtually no difference was found between results for GLS-
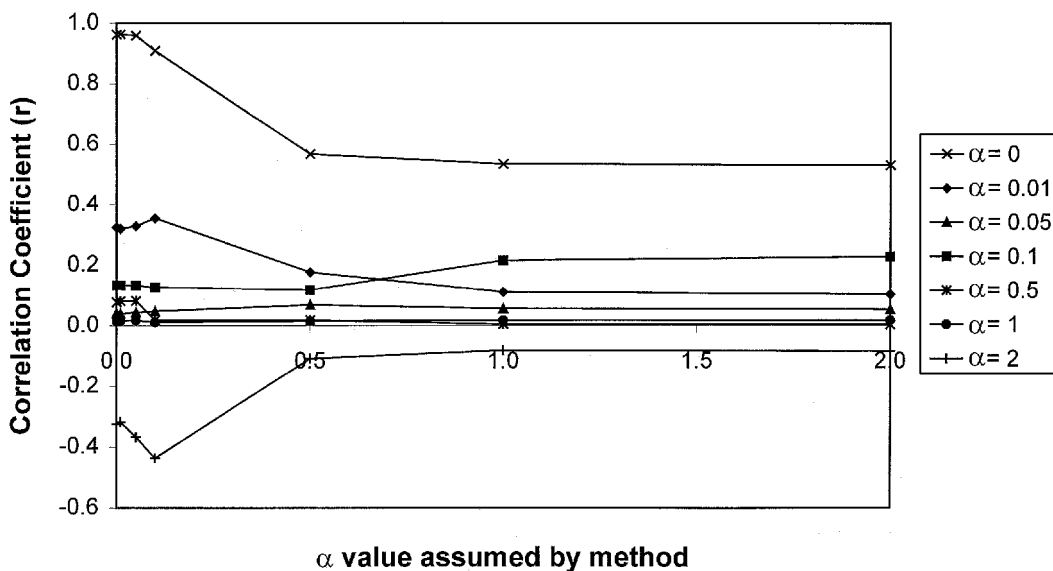
FIGURE 1. Correlation coefficients between the known state at the root of the 42-taxon phylogeny (from the simulation procedure) and the states estimated by each method. Values along the abscissa represent different versions of the GLS method. The value at 0 represents the GLS-linear method, whereas all other values refer to different values of the restraining force (α) assumed by the GLS-exponential method. Note that both the SSP and ML methods yield results identical to the GLS-linear method (method $\alpha = 0$). Each line connects values for a single run of the simulation procedure using a different microevolutionary model: $\times$'s mark Brownian motion evolution; all other lines refer to OU models, with values of α given in the box to the right.

exponential and GLS-linear on the 9-taxon phylogeny.

DISCUSSION

From one perspective, the results of this study are reassuringly positive. In the best-case scenario (Brownian motion evolution), SSP, ML, and GLS-linear often produced excellent estimates of some ancestral states, far better than has been anticipated by earlier authors (e.g., Schluter et al., 1997). Unfortunately, they also occasionally yielded exceedingly poor estimates for an ancestor, particularly those that were phylogenetically distant from the extant species data. Thus, it really is quite important to have some measure of estimation uncertainty. Unfortunately, the two methods (ML and GLS) that provide measures of the uncertainty of those states do so with only mediocre accuracy. Furthermore, the accuracy of ancestral state estimates depends critically on the type of evolution underly-

ing the characters involved. In a much less favorable scenario (i.e., with traits evolving under a strong constraint), all methods yielded very poor estimates of both ancestral states and estimation uncertainty. Use of the GLS-exponential method improved the situation very slightly with the 42-taxon phylogeny, but its use requires knowing the correct microevolutionary model, and the improvement was really quite small. Virtually no improvement was seen with data generated on the 9-taxon phylogeny, except (perhaps) in the estimates of error.

As suggested by Schluter et al. (1997), it seems quite difficult to estimate historically ancient events by using only comparative data. Long periods of time and the fluctuations of microevolutionary rates and scenarios all conspire to make ancestral state estimation seem an impossible task. Nevertheless, when phenotypic evolution was modeled as a Brownian motion process, 47 of the 49 nodes in this study were estimated quite well by SSP, ML, and

TABLE 1.   Results for the 49 nodes on the 2 phylogenies. Bias refers to the difference between the state estimated by each method and the actual ancestral state. Corr. refers to the Pearson product-moment correlation between estimates from each method and the actual ancestral state. MEB is the difference between the standard error calculated by each method and the true standard error (the squared deviation of each estimate from the true value of the ancestor) for each nodal estimate. MEB Corr. is the Pearson product-moment correlation between those two types of standard error. ML refers to Schluter et al.'s (1997) maximum likelihood method; GLS-linear is Martins and Hansen's (1997) method, assuming a linear model (unconstrained evolution); and GLS, $\alpha = 1$ is the same method assuming an exponential model (constrained evolution) with α set equal to 1. BM refers to data generated under a Brownian motion model of phenotypic evolution. OU refers to data generated under an Ornstein–Uhlenbeck model of phenotypic evolution. See text for details.

| | Range (and mean) | | | |
	Bias	Corr.	MEB	MEB Corr.
42-taxon tree				
BM				
ML	−0.71 to 0.39 (−0.07)	0.94 to 0.96 (0.96)	10.72 to 223.05 (50.58)	0.30 to 0.37 (0.33)
GLS-linear	−0.71 to 0.39 (−0.07)	0.94 to 0.96 (0.96)	10.72 to 223.04 (50.67)	0.30 to 0.36 (0.33)
GLS, $\alpha = 1$	−1.56 to 0.77 (−0.06)	0.45 to 0.56 (0.50)	194.45 to 564.66 (423.33)	0.23 to 0.50 (0.41)
OU, $\alpha = 1$				
ML	−0.09 to 0.09 (0.00)	−0.06 to 0.04 (−0.02)	−1.21 to 1.10 (0.59)	−0.17 to −0.12 (−0.15)
GLS-linear	−0.09 to 0.09 (0.00)	−0.06 to 0.04 (−0.02)	−1.20 to 1.11 (0.60)	−0.17 to −0.12 (−0.15)
GLS, $\alpha = 1$	−0.07 to 0.06 (0.00)	−0.02 to 0.04 (0.01)	−1.97 to 1.91 (−1.66)	−0.45 to −0.04 (−0.38)
9-taxon tree				
BM				
ML	−0.18 to 0.07 (−0.04)	0.50 to 0.99 (0.82)	0.27 to 31.3 (6.11)	0.94 to 1.00 (0.96)
GLS-linear	−0.18 to 0.07 (−0.04)	0.50 to 0.99 (0.82)	0.27 to 31.3 (6.11)	0.97 to 1.00 (0.98)
GLS, $\alpha = 1$	−0.18 to 0.08 (−0.04)	0.50 to 0.99 (0.82)	0.30 to 31.4 (6.34)	0.73 to 0.95 (0.82)
OU, $\alpha = 1$				
ML	−0.06 to 0.02 (−0.02)	−0.03 to 0.03 (−0.00)	0.17 to 3.28 (0.86)	0.80 to 1.00 (0.93)
GLS-linear	−0.06 to 0.02 (−0.02)	−0.03 to 0.03 (−0.00)	0.17 to 3.28 (0.86)	0.80 to 1.00 (0.93)
GLS, $\alpha = 1$	−0.06 to 0.02 (−0.02)	−0.03 to 0.03 (−0.00)	0.17 to 3.26 (0.85)	0.99 to 1.00 (0.99)

GLS-linear; i.e., bias was negligible and correlation coefficients between true and estimated ancestral states were > 0.90. Thus, in some situations, it may be reasonable to place quite a bit of confidence in our estimates of hypothetical ancestors. Furthermore, because these three methods all give essentially identical ancestral estimates, they can be used interchangeably for this purpose.

In some cases (for nodes that are deep in the tree or otherwise far from most of the extant taxa on a phylogeny), however, trait values for ancestors may be off by as much as 5 times the actual value of the ancestor. Unfortunately, both ML and GLS methods gave relatively poor estimates of the uncertainty involved in reconstructing ancestral states. Thus, it is difficult to know when to be confident and when to be worried. One suggestion would be to estimate standard errors with several possible methods (e.g., ML

and several versions of GLS-exponential), with the hopes that the combined result would bound the possibilities. Under Brownian motion evolution, standard errors were routinely underestimated by all the methods tested. Thus, it would be reasonable to examine the range of possible standard errors and consider the largest as a rough minimum estimate. Nevertheless, any specific standard error should be regarded with some degree of caution.

On another pessimistic note, certain microevolutionary scenarios may cause all three of the methods to provide exceedingly poor or even misleading estimates of both ancestral states and their standard errors. Brownian motion is commonly used as a model of phenotypic evolution under random genetic drift or under directional selection when the direction of selection fluctuates at random. It is a powerful null model and is a common assumption underlying
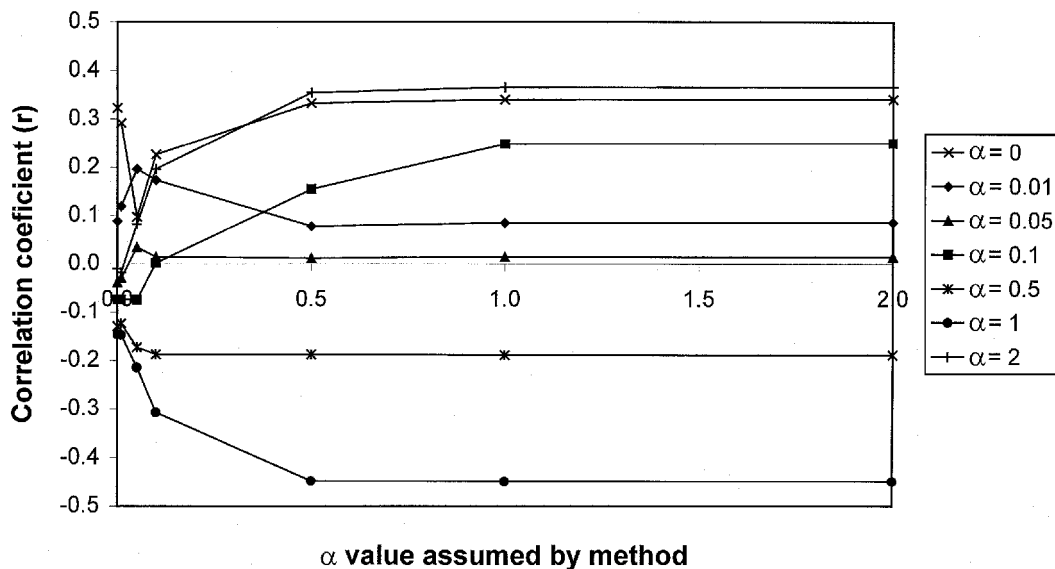
FIGURE 2.   Correlation coefficients between the known and estimated standard errors for each method for the root of the 42-taxon phylogeny. Values along the abscissa represent different versions of the GLS method. The value at 0 represents the GLS-linear method, whereas all other values refer to different values of the restraining force (α) assumed by the GLS-exponential method. Note that the ML method yields results that are virtually identical to the GLS-linear method (method $\alpha = 0$). Each line connects values for a single run of the simulation procedure using a different microevolutionary model: ×'s mark Brownian motion evolution; all other lines refer to OU models, with values of α given in the box to the right.

many comparative analyses (e.g., Felsenstein, 1985). Thus, this model is often sufficient to describe the evolution of a variety of traits (e.g., Schluter et al., 1997). Still, many of the traits used in comparative analyses are thought to have evolved under other microevolutionary models, including stabilizing selection, adaptation to fluctuating environments, and "burst-like" change. These types of microevolutionary scenarios tend to erase history from phenotypic data, replacing it with responses to an environment or changes in the environment (Hansen and Martins, 1996). Traits evolving under these types of scenario are unlikely to be well described by Brownian motion.

In the current study, I used an OU (rubber band) process to describe phenotypic evolution with a restraining force (measured as the parameter α). Not surprisingly, all of the methods have a much harder time estimating ancestral states with data generated by means of this model. Even fairly low levels of constraint (e.g., very weak

stabilizing selection) caused problems for ancestral state estimation, and strong constraints could lead the methods to give positively misleading results (i.e., negatively correlated with the true values). The rather tiny improvement of the GLS method set up to assume (correctly, in this case) that evolution was constrained under an OU model is discouraging. Future studies might consider even larger phylogenies to see whether a larger number of taxa would improve our ability to estimate the nuisance parameters in the GLS-exponential model. For now, though, it seems that stabilizing selection is similar to directional selection (which imposes a trend) in the way it quickly destroys our ability to infer phenotypic history.

Thus, although ML does not entirely solve the problem of obtaining standard errors, it improves the situation by providing some initial information (a minimum estimate) regarding the uncertainty of estimated ancestral states. Similarly, although GLS does not entirely solve the problem of alternative mi-

croevolutionary models, it does provide a way of obtaining a set of possible ancestral states under different microevolutionary scenarios. Future development of tests among these models would be quite useful, as would further simulation studies to determine whether this poor performance of standard error estimators and sensitivity to microevolutionary scenario is also true for methods used to estimate ancestral states of categorical characters. Finally, studies examining the effects of branch length transformations and within-species variability would also be useful.

REFERENCES

BUTLER, M. A. AND J. B. LOSOS. 1998. Testing for unequal amounts of evolution in a continuous character on different branches of a phylogenetic tree using linear and squared-change parsimony: An example using Lesser Antillean *Anolis* lizards. Evolution 51:1623–1646.

FARRIS, J. S. 1970. Methods for computing Wagner trees. Syst. Zool. 19:83–92.

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

FELSENSTEIN, J. 1988. Phylogenies and quantitative characters. Ann. Rev. Ecol. Syst. 192:445–471.

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

HUEY, R. B., AND A. F. BENNETT. 1987. Phylogenetic studies of coadaptation: Preferred temperatures versus optimal performance temperatures of lizards. Evolution 41:1098–1115.

LANDE, R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314–334.

LANDE, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33:402–416.

LARSEN, K. R. AND W. W. TANNER. 1974. Numeric analysis of the lizard genus *Sceloporus* with special reference to cranial osteology. Great Basin Natur. 34:1–41.

MADDISON, W. P. 1991. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. Syst. Zool. 40:304–314.

MARTINS, E. P. 1993. A comparative study of the evolution of *Sceloporus* push-up displays. Am. Nat. 142:994–1018.

MARTINS, E. P. 1998. COMPARE: Statistical analysis of comparative data. Distributed by the author (http://evolution.uoregon.edu/~COMPARE/or via anonymous ftp to evolution.uoregon.edu). Department of Biology, University of Oregon, Eugene.

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

MARTINS, E. P., AND T. F. HANSEN. 1997. Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149:646–667.

MARTINS, E. P., AND J. LAMONT. 1998. Evolution of communication and social behavior: A comparative study of *Cyclura* rock iguanas. Anim. Behav. 55:1685–1706.

MCARDLE, B., AND A. G. RODRIGO. 1994. Estimating the ancestral states of a continuous-valued character using squared-change parsimony: An analytical solution. Syst. Biol. 43:573–578.

SCHLUTER, D. 1997. ANCML. Distributed by the author (via http://www.zoology.ubc.ca/~schluter/ancml.html). Department of Zoology, University of British Columbia, Canada.

SCHLUTER, D., T. PRICE, A. Ø. MOOERS, AND D. LUDWIG. 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51:1699–1711.

SWOFFORD, D. L., AND W. P. MADDISON. 1987. Reconstructing ancestral character states under Wagner parsimony. Math. Biosci. 87:199–229.