TESTING FOR DIFFERENT RATES OF CONTINUOUS TRAIT EVOLUTION USING LIKELIHOOD

BRIAN C. O'MEARA,[1] CÉCILE ANÉ,[2] MICHAEL J. SANDERSON,[3,4] AND PETER C. WAINWRIGHT[3,5]
[1]*Center for Population Biology, University of California, Davis, One Shields Avenue, Davis, California 95616*
E-mail: bcomeara@ucdavis.edu
[2]*Department of Statistics, University of Wisconsin-Madison, Medical Science Center, 1300 University Avenue, Madison, Wisconsin 53706-1532*
E-mail: ane@stat.wisc.edu
[3]*Section of Evolution and Ecology, University of California, Davis, One Shields Avenue, Davis, California 95616*
[4]*E-mail: mjsanderson@ucdavis.edu*
[5]*E-mail: pcwainwright@ucdavis.edu*

Abstract.—Rates of phenotypic evolution have changed throughout the history of life, producing variation in levels of morphological, functional, and ecological diversity among groups. Testing for the presence of these rate shifts is a key component of evaluating hypotheses about what causes them. In this paper, general predictions regarding changes in phenotypic diversity as a function of evolutionary history and rates are developed, and tests are derived to evaluate rate changes. Simulations show that these tests are more powerful than existing tests using standardized contrasts. The new approaches are distributed in an application called *Brownie* and in *r8s*.

Key words.—Brownian motion, *Brownie*, comparative method, continuous characters, disparity, morphological evolution, rate.

All five extant flamingo species are long-legged filter feeders, whereas their sister group, consisting of twenty species of grebes (Van Tuinen et al. 2001; Chubb 2004; Mayr 2004), feed on prey ranging from fish and squid to minute invertebrates (Fjeldså 1983), and show a variety of body and bill shapes. Methods to test whether the difference in species number between flamingos and grebes arose by chance or reflects differences in diversification rates have been developed (Slowinski and Guyer 1989; Nee et al. 1992; Hey 1992; Harvey et al. 1994). These methods are aimed at discovering factors affecting diversification. But there are also undoubtedly factors that led to the difference in variability of ecologically important traits within these two groups. This paper is concerned with hypotheses about factors that lead to differences between groups in phenotypic and biological diversity, as opposed to species richness.

There are many potential hypotheses regarding factors that can affect the rate of evolution of phenotypic characters (which include morphological, behavioral, physiological, biochemical, and ecological traits). For example, once wings replaced legs as the primary means of locomotion in birds, newly less constrained legs may have begun to evolve new shapes more rapidly (Gatesy and Middleton 1997). The evolution of asexuality may reduce the rate of genome size evolution. The invasion of a new, competitor-free island may increase the rate of evolution of feeding structures. These hypotheses all attempt to relate a change in some aspect of the biology of the lineage with a change of the rate of evolution of a continuous character based on an idea about how evolution works. Hypotheses can also be generated from observations of patterns of diversity instead of predictions based on a mechanism. Grebes appear to have more interspecific variation in bill dimensions than flamingos: this may reflect a faster rate of bill evolution, or perhaps the grebe species have been evolving independently for more time than the flamingo species.

In this paper, we develop and implement new methods to make inferences regarding these questions. Basic results concerning character evolution on trees are presented. Our methods are illustrated using an example of genome size evolution in angiosperms.

SOME BASIC PROPERTIES OF CHARACTER EVOLUTION ON TREES

Disparity is commonly measured as variance of the states of the taxa (so higher disparity means the taxa are less similar for that particular character). The observed disparity is a function of many factors, such as the rate of phenotypic evolution, the amount of time the group has been evolving, and the relationships of the taxa. To examine any one factor, such as the evolutionary rate, a model must be used to control for the other factors, such as the phylogeny. A reasonable model to use in the case of phenotypic evolution is Brownian motion (BM). This is the standard model for continuous character evolution, used in independent contrasts (Felsenstein 1985) and estimation of ancestral states (Schluter et al. 1997). In Brownian motion, at each instant in time the state of a character can increase or decrease. The magnitude and direction of these shifts are independent of the current state of the character and have a net change of zero. Just as a bell curve is often a good fit to the distribution of a character within a population, because the state of the character for each individual is the result of the addition of many independent factors (Central Limit Theorem), the change of the state of a character as a result of the action of many displacements is often approximated well by Brownian motion. Brownian motion does not necessarily imply a process of neutral evolution, such as genetic drift (contra Butler and King 2004). Any process that creates displacements meeting the Brownian motion assumption is appropriately modeled by a Brownian motion process. For example, processes such as fluctuating di-

rectional selection (where the optimal value is allowed to change), punctuated change (long periods of stasis interrupted by abrupt change), and genetic drift all result in the same covariance structure as simple Brownian motion (Hansen and Martins 1996).

On the other hand a Brownian motion model may not be appropriate when its assumptions are violated. For example, a continuous character evolving by Brownian motion should have its chance of increasing or decreasing in state be independent of its current value. This is not the case if the character state is near its natural limits (a leg of length 0.5 mm cannot decrease by 1.0 mm, but a leg of length 6.0 mm could). Transforming the character (taking the log of leg length, for example), may be appropriate in such instances but is not always adequate. Brownian motion is also generally not the best model when there is consistent selection towards a single optimum trait value (better fit by an Ornstein-Uhlenbeck (OU) model (Hansen 1997; Butler and King 2004)).

Brownian motion also has a deficiency as a model in that it does not explicitly model a particular process. Finding that one group has a higher rate of Brownian motion than another does not reveal whether this is due to drift happening more quickly as a result of smaller population size, more changes in the location of an adaptive peak due to environmental change, or more frequent shifts in character state due to competition following speciation.

Even under simple Brownian motion models, the evolution of continuous characters on trees has some nonintuitive properties. For example, there has been uncertainty about which aspects of the phylogeny affect the expected phenotypic variation (Purvis 2004; Ricklefs 2004). Consider two sister groups with the same crown group age. The clade with the higher number of species has a higher estimated diversification rate, but this is not the case for disparity. For disparity, even with the same number of taxa in the two clades, the group with the higher phenotypic variance can actually have a lower rate of phenotypic evolution under a simple Brownian motion model (Fig. 1), depending on the distribution of node ages on the tree. Under Brownian motion, the relationship between expected disparity (expected sample variance), the tree, and the rate of phenotypic evolution (derived in the Appendix) is:

$$E(\text{disparity}) = \sigma^2 \left[\frac{1}{N}\text{tr}(\mathbf{C}) - \frac{1}{N^2}\mathbf{1}'\mathbf{C}\mathbf{1} \right] \quad (1)$$

where, following notation in Martins and Hansen (1997) and Garland and Ives (2000), N is the number of taxa, $\mathbf{C}$ is the N by N matrix in which the i-jth entry is the distance from the root node to the most recent common ancestor of taxa i and j (assuming branch lengths are proportional to time, this is the time between the root and the most recent common ancestor of i and j), σ^2 is the rate of character evolution, tr($\mathbf{C}$) is the trace of the matrix (in this case, the sum of the heights of each taxon above the root), and $\mathbf{1}$ is a column vector of ones of height N.

This equation allows numerous inferences about the relationship between the tree, rate, and phenotypic disparity. The expected amount of disparity for a clade is a function of three factors: the rate parameter (σ^2), the time to the most recent common ancestor of the clade [assuming the taxa are
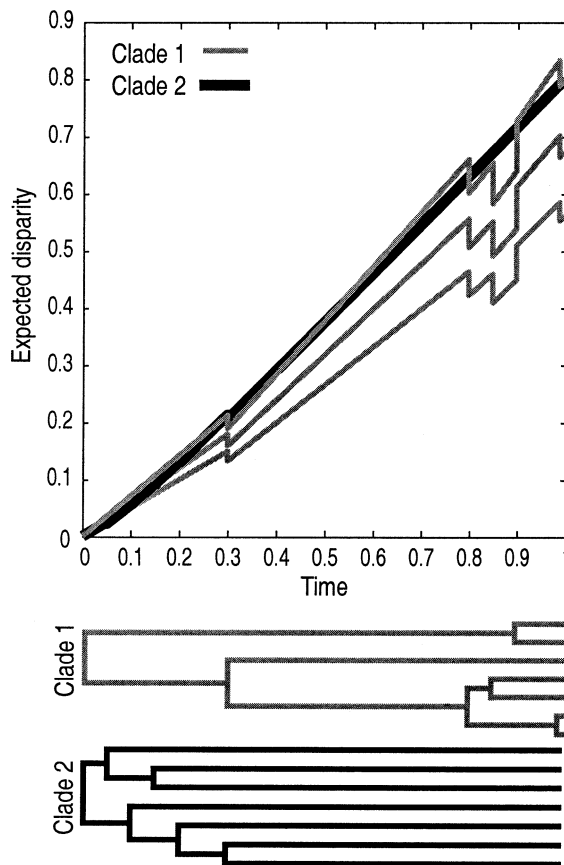

FIG. 1. Equation 1 was used to predict variance in terminal trait values within clade 1 and clade 2 through time. For clade 2, the Brownian rate parameter was always 1.0. For clade 1, the rate parameter took on values of 1.0 (the lowest line), 1.2 (the middle line), and 1.425 (the top line). Note that though the two clades have the same age and same number of taxa, clade 2 has a higher tip disparity when the rates of morphological evolution are equal and even when clade 1 has a 20% faster rate. Clade 1 has a greater amount of tip disparity when its rate is greater than 1.425 times that of clade 2.

contemporaneous] (($1/N$)tr($\mathbf{C}$)), and the average entry in the covariance matrix (($1/N^2$)$\mathbf{1}'\mathbf{C}\mathbf{1}$). Under the Brownian motion model, these covariance terms are proportional to the amount of time between the most recent common ancestor of a pair of taxa and the root. Higher rates of phenotypic evolution (σ^2) on a given tree result in more dissimilar taxa. All else being equal, older clades should have more phenotypic variation than younger clades. Clades of a given age where most speciation events have happened early (more ''starlike'' trees) should have higher disparity than clades where most speciation events occur late. In other words, as internal branches get shorter and terminal branches longer, shared covariance (the off-diagonal elements in $\mathbf{C}$) decreases, diagonal elements remain the same, and thus tip variance increases.

Patterns of change of disparity through time are also informative. The effect of speciation events on disparity is complex, with results depending on which lineages speciate when. The average time between the terminal taxa and the root (($1/N$)tr($\mathbf{C}$)) is the same on a tree terminated just before

and a tree terminated just after the instantaneous speciation event, since no time has elapsed. The average entry in the covariance matrix will generally change with a speciation event. A speciation event introduces a duplicate of an existing species, creating a species pair with the maximum possible covariance. This tends to increase $(1/N^2)\mathbf{1'C1}$ and thus decrease the expected disparity. Thus, most of the speciation events for clade 1 in Figure 1 result in an instantaneous drop in the expected disparity immediately after speciation. However, the covariance of this new taxon with the other taxa in the tree also matters. In clade 1, the early lineage sister to the rest of the lineages has below average covariance (zero covariance with every other species). Thus, when it speciates at time 0.9, a new species is created that also has below average covariance, reducing $(1/N^2)\mathbf{1'C1}$ more than covariance between the two daughter species increases $(1/N^2)\mathbf{1'C1}$. This results in an increase in predicted tip disparity immediately after that speciation event.

The story is even more complicated when extinction is considered. In the absence of extinction the time until the most recent common ancestor for the contemporaneous species increases through time. This will increase the $(1/N)\text{tr}(\mathbf{C})$ term in equation 1 and thus tend to increase expected disparity as the clade evolves (though note that speciation events may intermittently reduce expected disparity). With extinction, this need not be the case: the extinction of some taxa may decrease the time to the most recent common ancestor for the clade. For example, if, at time 0.9 on Figure 1 for clade 1, the lineage that speciated went extinct instead, the most recent common ancestor of the remaining taxa would occur at time 0.3, not time 0. Groups that diversify quickly (high ratio of speciation to extinction rates) and then reach an equilibrium species diversity would be expected to show an increase in disparity until it reached a plateau. Under more complex speciation or extinction models, disparity may even decrease through time, even when the rate of evolution does not change. Figure 2 shows a disparity-through-time plot for a diversification model in which a clade starts with one species, increases to an equilibrium value, and then fluctuates in species number near this value. Throughout the simulation, the rate of morphological evolution (under the simple Brownian motion model) remains constant. Initially, disparity increases but then it also plateaus. Halfway through the simulation, the magnitudes of the speciation and extinction rates are both increased (in other words, the turnover rate is increased). Although the equilibrium number of species in the clade and the rate of morphological evolution remain the same, disparity decreases to a new, lower plateau. This occurs because the higher turnover rate results in a lower time to most recent common ancestor of the surviving species at any one point in time, so there is less time over which to evolve disparity.

These examples (also see simulations by Foote 1996) demonstrate that phenotypic disparity and rate of phenotypic evolution are distinct characteristics, and a phylogeny is needed to link the two. For example, the pattern of disparity increasing to a plateau has often been observed in fossil taxa (reviewed in Foote 1997 and Zelditch et al. 2004) but inferences about rate have been drawn without phylogenetic context,
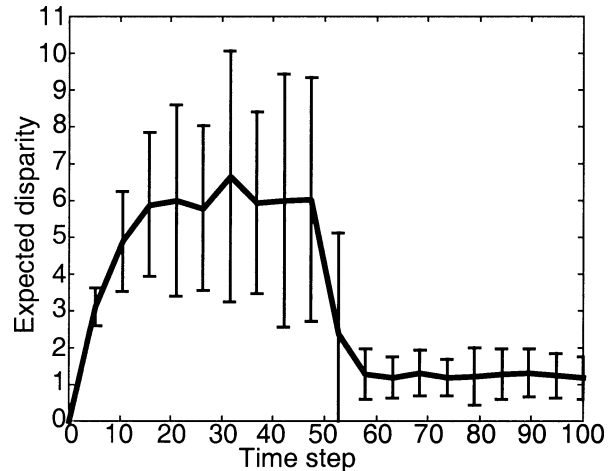

FIG. 2. Variance through time plot for 100 simulations of Brownian character evolution on trees generated under a birth-death model (error bars are one standard deviation over the simulations). In the first half of the total time, the clade is constrained to have a speciation rate of 1.0 and extinction rate of 0.0 when there are fewer than 10 taxa; a speciation rate of 1.0 and an extinction rate of 1.0 when there are 10 or 11 taxa; and a speciation rate of 0.0 and extinction rate of 1.0 when there are more than 11 taxa. For the second half of the simulation, the same model is used but with rates five times those in the first time period. This model has the effect of keeping the number of coeval taxa between 9 and 12 once at least 9 taxa evolve from the initial taxon. Notice the initial increase in variance, then a constant plateau of variance through time. The increase in speciation and extinction rates halfway through the simulation result in shorter species coalescence times and consequently less variance in tip values.

which is sometimes unavailable for fossil taxa. The considerations described above suggest the risks of doing so.

METHODS

To compare rates of evolution, one must estimate rate in two or more groups and then have a method to determine whether the observed differences are meaningful. There are several methods available to estimate phenotypic rates of evolution, such as generalized least squares (Martins 1994), standardized contrasts (Garland 1992), or Lynch's estimator (Lynch 1990). We have implemented a likelihood estimator of rate and developed new tests to compare the rates of phenotypic evolution on two or more parts of a tree.

Phenotypic characters are assumed to evolve according to Brownian motion. The maximum-likelihood (ML) estimator of the rate of phenotypic evolution under Brownian motion is

$$\hat{\sigma}^2 = \frac{[\mathbf{X} - \text{E}(\mathbf{X})]'\mathbf{C}^{-1}[\mathbf{X} - \text{E}(\mathbf{X})]}{N} \qquad (2)$$

where $\mathbf{X}$ is the column vector of tip observations (observed states of the terminal taxa) and $\text{E}(\mathbf{X})$ is the vector of estimated expected tip values. This ML estimator is very similar to the unbiased estimator of Garland and Ives (2000), equation B3, but with N instead of $(N - 1)$. Because, under Brownian motion, the expected net change from the ancestral state of the clade is zero, $\text{E}(\mathbf{X})$ is just a vector of the estimated state at the root, $\hat{a}$. From Blomberg et al. (2003), equation A16,

TABLE 1. Precision, bias, and mean squared error of the rate estimator (eq. 2). Tree shapes used were a star phylogeny (unresolved bush), a pectinate tree (a comb) with all internal branches of the same length, and a balanced tree with all branches of the same length. All trees had a root to tip length of 1. For each tree, 1000 datasets were simulated under a Brownian motion model with a rate parameter of 1.0 and then analyzed using Brownie 1.0. Precision is the variance of the rate estimates (lower is better), bias is the mean observed rate estimate minus the true parameter value (lower is better), and mean squared error (also known as accuracy) is the precision plus the square of the bias.

	Precision			Bias			Mean squared error		
Taxa	Star	Pectinate	Balanced	Star	Pectinate	Balanced	Star	Pectinate	Balanced
4	0.345	0.394	0.366	−0.257	−0.221	−0.241	0.411	0.443	0.424
8	0.252	0.215	0.205	−0.081	−0.132	−0.150	0.259	0.233	0.227
16	0.109	0.106	0.118	−0.083	−0.067	−0.071	0.116	0.111	0.123
32	0.061	0.062	0.061	−0.035	−0.040	−0.026	0.063	0.064	0.062
64	0.030	0.031	0.030	−0.014	−0.011	−0.019	0.030	0.031	0.031
128	0.016	0.015	0.016	−0.013	−0.016	−0.005	0.016	0.015	0.016

the maximum-likelihood estimate of $\hat{a} = (\mathbf{1}'\mathbf{C}^{-1}\mathbf{1})^{-1}$ $(\mathbf{1}'\mathbf{C}^{-1}\mathbf{X})$. Pagel (1998) outlines a similar approach to estimating rate and ancestral state. This estimator (eq. 2) assumes trait evolution by Brownian motion and that the topology, branch lengths, and trait values are known exactly.

Table 1 provides information on the bias, precision, and mean squared error of the rate estimator. As is common with likelihood estimators, the rate estimator is biased, in this case returning estimates that are too low. Thus, for a four-taxon tree (or subtree), the estimated rate is about 25% lower than the true rate, but this bias drops as the number of taxa increases. The relative error shows similar behavior. This suggests caution when comparing rate estimates for groups of different sizes, where at least one of the groups is small: all else being equal, the smaller group will tend to have a somewhat lower rate estimator with greater uncertainty.

Once rates are estimated, it becomes possible to infer whether they are meaningfully different. We have developed two different approaches. The ''noncensored'' approach assigns each branch (or different portions of a branch) a rate parameter (σ^2). In the simplest case of testing for a single change in rate, branches would be assigned either of two rate parameters. In the approach implemented here, the stem branch of a clade (the branch connecting the most recent common ancestor of the clade to the rest of the tree) was assigned the same rate parameter as the clade's crown group branches. This decision was arbitrary: the stem branch could have been assigned the ancestral rate parameter, or the point at which the rate parameter changed could be optimized using an additional parameter. The likelihood of a model in which the rate parameters are allowed to vary may be compared to the likelihood of the model in which the rate parameters are constrained to be equal. Parameters are estimated by numerical solution of the appropriately modified likelihood function. This function is the multivariate normal distribution, and the log likelihood of the multivariate normal observations given the estimated rate parameter, observed states of the taxa, tree, and inferred ancestral state is

$$\log(\mathbf{L}) = \log\left[\frac{\exp\left\{-\frac{1}{2}[\mathbf{X} - E(\mathbf{X})]'(\mathbf{V})^{-1}[\mathbf{X} - E(\mathbf{X})]\right\}}{\sqrt{(2\pi)^N \times \det(\mathbf{V})}}\right]. \quad (3)$$

In the case of a single rate parameter, the variance-covariance matrix $\mathbf{V} = \sigma^2\mathbf{C}$. For multiple rate parameters, each element $\mathbf{V}_{i,j}$ is the sum of the branch lengths between the root and the most recent common ancestor of taxon i and taxon j, with the length of each branch segment being the duration of that segment multiplied by the rate parameter for that segment. For example, in Figure 3, the variance of taxon 1 ($\mathbf{V}_{11}$) is the amount of time spent in rate parameter σ_A^2 (10 my + 30 my = 40 my) times σ_A^2 plus the amount of time spent in rate parameter σ_B^2 (10 my) times σ_B^2. Thus, $\mathbf{V}_{11} = 40 \sigma_A^2 + 10 \sigma_B^2$. The covariance of taxon 1 and taxon 2, element $\mathbf{V}_{12}$, is the length of shared branches spent in rate parameter σ_A^2 (30 my) times σ_A^2 plus the amount of time spent in rate parameter σ_B^2 (10 my) times σ_B^2. Thus, $\mathbf{V}_{12} = 30 \sigma_A^2 + 10 \sigma_B^2$. Also note that $\mathbf{V}_{12} = \mathbf{V}_{21}$. The likeliest ancestral state for the whole tree is $\hat{a} = (\mathbf{1}'\mathbf{V}^{-1}\mathbf{1})^{-1} (\mathbf{1}'\mathbf{V}^{-1}\mathbf{X})$. If there is one rate parameter assigned to the tree, the ancestral state estimate is independent of the value of this parameter, but the parameter values (specifically, the ratio of parameter values) do affect this estimate if there are multiple rate parameters. To calculate the likelihood of a model that fixes one rate parameter across the entire tree, all the rate parameters are set to be equal, while the multiparameter model score is calculated by allowing the rates to vary.

A ''censored'' approach, with deletion of the relevant branch and modification of the variance-covariance matrices, allows analytical solutions and makes no assumptions about where or how the rate change occurs on the branch other than that the change happened somewhere on that branch. In this approach, branches on which rate parameters change are deleted from the tree, yielding two or more subtrees. As in the noncensored approach, the rate parameters can be set to equal each other or allowed to vary. Note that regardless of whether the rate parameters are set equal or allowed to vary, the ancestral states for the subtrees are estimated independently, unlike the noncensored case.

Deleting one or more branches and treating the resulting subtrees as independent is admittedly somewhat counterintuitive. The states at the beginning and end of the deleted branch certainly are correlated. The noncensored approach assumes that this correlation can be known and is a simple function of the duration of the branch and the rate parameter. In contrast, the censored approach chooses to remain agnostic about this correlation, and instead estimates an additional parameter for the state at the end of this branch given in-

Non-censored approach
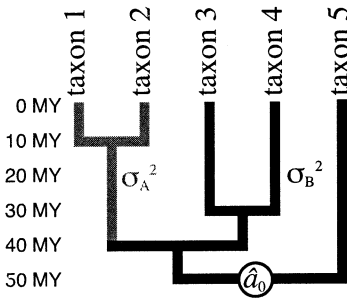
	taxon 1	taxon 2	taxon 3	taxon 4	taxon 5	$E(\mathbf{X})$
taxon 1	$40\sigma_A^2+10\sigma_B^2$	$30\sigma_A^2+10\sigma_B^2$	$10\sigma_B^2$	$10\sigma_B^2$	0	$\hat{a}_0$
taxon 2	$30\sigma_A^2+10\sigma_B^2$	$40\sigma_A^2+10\sigma_B^2$	$10\sigma_B^2$	$10\sigma_B^2$	0	$\hat{a}_0$
taxon 3	$10\sigma_B^2$	$10\sigma_B^2$	$50\sigma_B^2$	$20\sigma_B^2$	0	$\hat{a}_0$
taxon 4	$10\sigma_B^2$	$10\sigma_B^2$	$20\sigma_B^2$	$50\sigma_B^2$	0	$\hat{a}_0$
taxon 5	0	0	0	0	$50\sigma_B^2$	$\hat{a}_0$

Censored approach
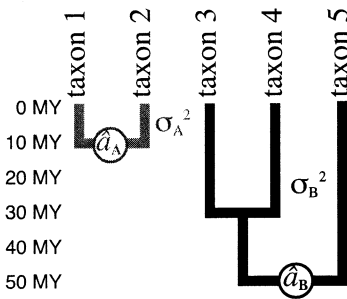
	taxon 1	taxon 2	taxon 3	taxon 4	taxon 5	$E(\mathbf{X})$
taxon 1	$10\sigma_A^2$	0	0	0	0	$\hat{a}_A$
taxon 2	0	$10\sigma_A^2$	0	0	0	$\hat{a}_A$
taxon 3	0	0	$50\sigma_B^2$	$20\sigma_B^2$	0	$\hat{a}_B$
taxon 4	0	0	$20\sigma_B^2$	$50\sigma_B^2$	0	$\hat{a}_B$
taxon 5	0	0	0	0	$50\sigma_B^2$	$\hat{a}_B$

FIG. 3. Example showing the details of calculation for the noncensored and censored tests. The noncensored test uses the entire variance-covariance matrix shown with equation 3 to calculate the likelihood. One ancestral state is estimated for the entire tree. The optimal rate parameters under the unconstrained model must generally be estimated numerically (different rate parameter values are tried until the likelihood of the model is maximized). In contrast, the censored test analyses each subtree separately (indicated by the bold lines in the matrix and ancestral state vector). Each subtree has just one Brownian motion rate parameter, the value of which can be calculated analytically using equation 2 (the ancestral state for each subtree is also calculated analytically). In the simple example here, both the censored and noncensored tests would compare the likelihood of a model where σ_A^2 and σ_B^2 were set equal to each other with a model where the rates were free to vary.

formation from only the subtree. This comes at a cost (an additional nuisance parameter), but this does not appreciably affect the power of the test (see Figure 4). The benefit is that no assumptions are made about the process occurring on the deleted branch: the result is independent of whether the change in rate was instantaneous or a gradual shift, whether the rate change occurred early or late on that branch, and whether the state of the character changed in an unusual manner along that branch.

We use two statistical approaches to testing for different rates of evolution. In a conventional hypothesis-testing likelihood ratio approach, one wants to accept or reject the null hypothesis that the rate of evolution is the same everywhere on the tree. The test statistic $2((-\log(L_{\text{simpler model}})) - (-\log(L_{\text{complex model}})))$ has approximately a chi-squared distribution with the degree of freedom equal to the difference in the number of free rate parameters in the two models (Casella and Berger 2002). A significant likelihood ratio test rejects the null model. However, because the chi-square approximation is nonconservative with small numbers of taxa, a parametric test, involving simulating data on the tree under the null model and comparing the empirical likelihood ratio with a distribution of likelihood ratios under the null model, has also been implemented and should generally be preferred. The likelihood-ratio test approach has a few disadvantages. It is limited to comparison of nested models. Testing multiple models is also somewhat difficult using this approach.

An alternative to hypothesis testing is a model selection procedure that seeks a model that well approximates the information in the data (Burnham and Anderson 2002, p. 136). The distance between the true model and the approximating model is approximated by the Akaike Information Criterion (AIC) value (Akaike 1973), or the second-order information criterion (AIC$_c$, Sugiura 1978); the best model is the one with the lowest value. The AIC$_c$ is recommended (Burnham and Anderson 2002) when the sample size (in this case, the number of taxa) is less than forty times the number of estimated parameters.

A critical measure of performance for a test is its statistical power (related to type II error) at a given level (type I error). We compared the power of the parametric test, chi-square test, and Garland's test (1992). Power (probability of correctly accepting the non-null hypothesis) is not a relevant question for model selection criteria, so the AIC and AIC$_c$ were not used for this comparison. Each tree consisted of two identical clades as sister groups. Three shapes for the clades were used: pectinate with all internal branches of the same length; balanced with all branches in a clade of the same length; or (near) star phylogenies, with internal branches of trivial length. For each of the combinations of the following parameters and for each tree shape, 100 simulations were run in the new Matlab package *Brownie* 1.0 (see Program Note below). The rate parameter doubled in value at the base of the stem of the second clade. Five hundred da-
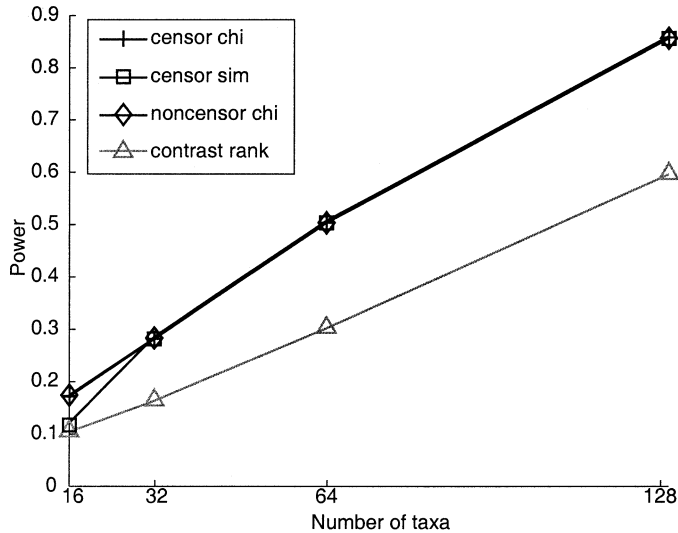
FIG. 4. Power of tests to compare rates of evolution: data were simulated and analyzed as described in the text. For each set of simulation settings and for each test, the proportion of *P*-values below an alpha level of 0.05 was measured to calculate the power of the test. The average power of each test across all topologies of a given number of taxa was calculated by averaging these values and graphed. Similar results are obtained when power for different tree shapes and stem lengths are not pooled. ''Censor chi'' is the power for the censored test described here, with *P*-values calculated based on a chi-square distribution. ''Censor sim'' is the power for the censored test with *P*-values based on parametric bootstrapping. ''Noncensor chi'' is the power for the noncensored test with *P*-values determined based on the chi-square distribution. ''Contrast rank test'' represents the power under a test comparing the standardized contrasts of the two clades using the Mann-Whitney *U* test. Note the greater power of the tests developed here.

TABLE 2. Levels (type I error) of the tests implemented here. Five hundred simulations were done for each set of parameters used to generate Figure 4, but simulated under a null model. ''Censor chi'' is the power for the censored test described here, with *P*-values calculated based on a chi-square distribution. ''Censor sim'' is the power for the censored test with *P*-values based on parametric bootstrapping. ''Noncensor chi'' is the power for the noncensored test with *P*-values determined based on the chi-square distribution. ''Contrast rank test'' represents the power under tests comparing the standardized contrasts of the two clades using the Mann-Whitney *U* test as described in Garland (1992).

ntax	Censor chi	Censor sim	Noncensor chi	Contrast rank
16	0.064	0.035	0.067	0.053
32	0.061	0.055	0.061	0.050
64	0.045	0.044	0.045	0.039
128	0.061	0.058	0.061	0.048

power of the test. The average power of each test across all topologies of a given number of taxa was calculated by averaging these values and graphed in Figure 4. The power of the tests rose approximately linearly with the number of taxa. The most powerful approaches were the censored test with a chi square distribution and the noncensored test with a chi square distribution. The parametric bootstrapping test under the censored model was more powerful than the Garland rank test at low numbers of taxa and rose to be nearly as powerful as the chi-square tests at greater numbers of taxa.

A difference in apparent power can be due to incorrect levels: a nonconservative test (one with elevated type I error) may appear more powerful than a test with appropriate level. To test this, the procedure above was repeated but with 500 datasets generated under the null model. Table 2 indicates that all the tests have approximately the appropriate Type I error (desired level of 0.05), though the chi-square tests tend to have the highest level, which may explain their higher apparent power.

Incorporating uncertainty in topology and branch lengths when using comparative methods has become increasingly important to researchers. For the approaches developed here, this can be accomplished by performing analyses on a set of trees, which can be trees generated by a bootstrap search, post-burn-in Bayesian trees (also see Huelsenbeck and Rannala 2003), or a credible set of trees. Tree weights can be used to calculate weighted averages of parameter estimates or other results across the set of trees. Note that these trees should have sensible branch lengths (typically branch lengths proportional to time, but in some cases branch lengths proportional to number of generations or estimated number of speciation events). Although branch lengths are often invented for use in comparative methods when topologies alone are available, such a technique in the present case is inadvisable because transformation of informative branch lengths underlies the model.

Our approach assumes that characters evolve according to a Brownian motion process within each section of the tree. This assumption is testable using existing techniques, although these should be modified to work on the independent subtrees (censored case) or use the tree with branch lengths multiplied by the optimal rate parameter values under the multirate model (noncensored case). For example, Pagel

tasets were simulated under the null model for each combination of simulation parameters and *P*-values were determined under the censored parametric bootstrapping approach; this null distribution was used for each of the 100 datasets. Each of the 100 datasets was used in *Brownie* to calculate *P*-values under the chi-square and parametric bootstrapping methods.

Each unique tree was saved as a distance matrix in a Nexus batch file by *Brownie*. PAUP version 4b10 (Swofford 2003) was used to save trees in formats suitable for loading into *r8s* 1.70 (for the noncensored test) and *econtrast* (Felsenstein 2003, ported to command line version as part of the EMBASSY package of EMBOSS (Rice et al. 2000)) for the Garland tests. The 100 datasets for each set of simulation parameters were saved in formats appropriate to *r8s* and *econtrast* using Matlab and Perl scripts. For *econtrast*, the tree and dataset were split to correspond to the two clades to enable contrasts to be computed independently for each clade (this does not change the contrast values in a sister group comparison). Output from *r8s* and *econtrast* was processed using Perl scripts; Garland's (1992) tests of absolute values of standardized contrasts were implemented in Matlab and *P*-values for the *econtrast* output generated using these functions. *P*-values were calculated using the chi-square distribution for the noncensored test results. For each set of simulation settings and for each test, the proportion of *P*-values below an alpha level of 0.05 was measured to estimate the

(1994, 1998, 1999) provides a set of parameters (κ, δ, and λ) to transform trees to meet the Brownian motion assumption; values not significantly different from 1 do not reject the Brownian motion assumption. Butler and King's (2004) approach may also be used to compare the models developed in this paper with an Ornstein-Uhlenbeck process. *Brownie* 2.1, still in development, will include the ability to do this comparison and do other comparisons with OU and BM models. As with other comparative methods that assume Brownian motion, traits may need to be appropriately transformed. For example, the methods here assume that an increase by a particular amount is equally likely regardless of the initial trait value. A trait such as leg length might be better fit by taking the log of length rather than the untransformed value (a leg of length 2 cm and one of length 100 cm may have an equal chance of lengthening by 10%, but not of lengthening by 3 cm).

Example

To illustrate the method with a worked example, we analyze the evolution of genome size in angiosperms. The evolution of genome size in angiosperms has been the subject of several papers (Soltis et al. 2003; Knight et al. 2005; Leitch et al 2005), which typically focus on ancestral states, not rates of evolution. These analyses have suggested that many angiosperm groups had small ancestral genome sizes. Large genome sizes have evolved a few times in monocots and in the Santalales (a clade within eudicots) but remain small in other groups. A natural hypothesis is whether this higher range of values in these two clades reflects higher rates of genome size evolution.

Angiosperm genome sizes (1C values, the amount of DNA amount in the unreplicated gametic nucleus, in picograms), were downloaded from the Plant C-values Genome Database, Release 4.0 (Bennett and Leitch 2005; Zonneveld et al. 2005). To test for different rates of evolution, an estimate of the tree with informative branch lengths (generally, proportional to time) is needed. The Soltis et al. (2000) dataset was used for this purpose. Taxa (genera) not present in the Genome Database were excluded. This left 219 taxa, of which seven were outgroups (gymnosperms). Sites for which the alignment was judged too uncertain were excluded, as was the phylogenetically problematic taxon *Ceratophyllum*, which has historically been a ''wild-card'' taxon in many analyses. A neighbor-joining search was performed in PAUP 4.0b10 (Swofford 2003). The resulting tree was used to optimize parameter values for an HKY + gamma likelihood model. These model settings were used in 150 likelihood bootstrap replicates (each starting from a neighbor-joining tree and limited to 4000 rearrangements to speed the search). The bootstrap searches enforced constraints on the monophyly of angiosperms, monocots, eudicots, and Santalales; the monophyly of these groups is uncontroversial and wellsupported in this dataset, at least using parsimony (Soltis et al. 2000). Ensuring that all the resulting trees have these groups makes later analysis easier. These bootstrap trees were brought into *r8s* (Sanderson 1997, 2002), outgroups deleted, three age constraints assigned (for angiosperms, monocots, and eudicots, based on the FOUR123 F+ estimates from Magallón

and Sanderson 2005), and converted to ultrametric trees using the Langley-Fitch method with the truncated Newton algorithm. *R8s* omits bootstrap weights from tree output, so these were hand entered based on the weights of the input trees (if N trees are found in a single bootstrap replicate, each receives a weight of $1/N$—in the likelihood bootstrap done here, N was no larger than three). Nonbootstrap tree searches were done using the same model as in the bootstrap search, using parsimony, neighbor-joining, and random taxon addition to get starting trees. The best tree from these searches was also converted to a chronogram using *r8s*.

The average entry in the Plant Genome Database for each taxon was used as its state. Two sister taxa, *Poncirus* and *Citrus*, had zero length terminal branches but different states. This causes an error in the likelihood calculations (because the observations are assumed to be without error, this change of state over zero time implies infinite rate) and thus one of these taxa was excluded and the other given the mean trait value of the two. Genome size is positively skewed (most plants have small values, with a few large outliers). After some initial data exploration, the log of genome size was used in place of raw genome size. Although previous analyses have not transformed genome size in this way, this is more appropriate than using untransformed values when estimating rate or state. Genome size may evolve through many processes, ranging from genome duplication to insertion or deletion of single bases in noncoding regions. For most of these processes, change will be better represented as a proportion of genome size than as a change of raw size. For example, a gain or loss of 10% is expected to be equally likely in a genome of size 7 pg as in a genome of 40 pg, but this is not true for the gain or loss of 5 pg over some arbitrary time period. Under the Brownian motion model used in this approach (and in approaches for estimating ancestral states of continuous characters), data should be transformed so a gain or loss by a given amount is equally likely regardless of starting state, which is accomplished by taking the log of genome size (this has the effect of improving the fit of the data given the model by approximately 300 log likelihood units). Log-transforming data is often appropriate when the character is constrained to be nonnegative, as in this case.

There are four groups to which we could reasonably consider assigning different rate parameters to test our hypotheses: Santalales, eudicots excluding Santalales, monocots, and angiosperms excluding monocots and eudicots. We could also choose to exclude some of these groups (e.g., test for different rates between Santalales and other eudicots only, excluding monocots and other angiosperms). There are 50 possible models under these conditions, although only 35 in which there are at least two rate categories on the tree. If zero represents taxon set exclusion from the model, and taxon sets assigned to the same rate parameter have the same number, we can represent the inclusion and rate parameter of Santalales/other eudicots/monocots/other angiosperms using a four digits. For example, a model that excluded Santalales (0), included other eudicots (rate parameter 1), included monocots with their own rate parameter (2), and set the remaining angiosperm rate to be equal to the eudicot rate (1), could be represented as 0121. Some of these models may be nested (for example, 1110, 1120, 1210, and 1220 are all

TABLE 3. Rate comparisons using the hypothesis-testing approach. Model symbols are as described in the text [S = Santalales, E = eudicots excluding Santalales, M = monocots, and A = angiosperms excluding monocots and eudicots]. Parametric bootstrapping tests used 1000 replicates for the optimal tree, and 200 replicates per bootstrap tree. The first P-value in each cell is that for the optimal tree; the second is from the weighted average of the bootstrap trees.

| Null model | Other model | Chi-square | Parametric bootstrap | |
Sema	Sema	P-value	P-value	Conclusion
1100	1200	0.0007 (0.0007)	0.04 (0.04)	Marginal
1110	1230	0.0006 (0.0009)	0.04 (0.03)	Marginal
1111	1234	0.0008 (0.0004)	0.07 (0.04)	Marginal
1232	1234	0.0018 (0.0015)	0.06 (0.05)	Marginal

nested in 1230). Others may have the same set of taxa but not be nested within each other (0122 and 0121) and thus only be comparable using AIC. Some may not even include the same taxa (1200 and 0121) and therefore are not easily comparable. We chose to analyze 1100 versus 1200, 1110 versus 1230, and 1111 versus 1234, and 1232 versus 1234 using a hypothesis-testing approach, and evaluate those models, plus 1120, 1220, 1212, 1123, 1222, and 1121 using a model selection approach. All these tests were performed in *Brownie* 2.0 using the censored test. Results, which are based on the maximum likelihood tree and the weighted average across the transformed bootstrap trees, are summarized in Tables 3 and 4. Santalales had the highest rate of evolution by far (rate of 0.098 on the ML tree, weighted average across bootstrap trees of 0.095, range across bootstrap trees of 0.077–0.18), followed by monocots (0.017, 0.016, 0.14–0.026), remaining eudicots (0.010, 0.009, 0.008–0.027) and remaining angiosperms (0.009, 0.007, 0.005–0.009). Likelihood ratio tests (Table 3) showed strong disagreement between the chi-square P-value and the parametric bootstrap p value, with the latter often on the border of significance while the former always indicated that the more complex model should be chosen. This may be due to the small number of samples (two) in the Santalales clade, which tends to make the chi-square test nonconservative. Analysis of the AIC and AIC$_c$ results suggests that Santalales have a different rate of genome size evolution from most other flowering plants, that there is some evidence to suggest that monocots also have a higher rate (but not that their rate is necessarily different from that of Santalales), and that remaining eudicots and remaining angiosperms do not differ strongly in rate.

DISCUSSION

Comparison of the Censored and Noncensored Approaches

Each of the approaches developed here will be useful in different contexts. They are both appropriate when investigating how a change in state in one discrete character affects the rate of evolution of another character, but have different benefits and costs. If the placement of this change along a branch or the mechanism of rate change (instantaneous change or gradual change from one rate to the other) is uncertain, the censored approach removes the need to make assumptions about this process: the branch on which the rate change is postulated to occur is removed from the analysis. A disadvantage of this approach is that the groups being examined, although they may be polyphyletic, must consist of paraphyletic groups of at least two taxa (assigning a rate to only internal branches, e.g., is not possible). In the noncensored approach, painting rate parameters on different branches (or parts of different branches) can be done, at the cost of making assumptions about the manner of the rate change. The noncensored approach can allow more flexible tests, such as testing whether the rate of character evolution in the period after a mass extinction is higher than the rate before the mass extinction.

Comparison with Other Methods

Several methods have been advanced to examine rates of continuous character evolution, including Lynch (1990), Garland (1992), Foote (1993), Martins (1994), Pagel (1994,

TABLE 4. Rate comparisons using Akaike Information Criterion. Values are only comparable when the included taxa are identical. Note that, when all the groups are included, the most complex model was not the best model. The first value in each cell is that for the optimal tree; the second is from the weighted average of the bootstrap trees. K is the number of free parameters (rate(s) and ancestral states) for each model.

Model SEMA	−lnL	AIC	AIC$_c$	K	Conclusion
1100	172.839 (172.328)	9.5 (10.2)	9.3 (10.1)	3	Different rates for Santalales and other eudicots
1200	167.106 (166.232)	0 (0)	0 (0)	4	
1110	270.414 (270.105)	10.7 (11.5)	10.5 (11.3)	4	Best model has Santalales, other eudicots, and monocots having
1120	268.790 (268.434)	9.5 (10.2)	9.3 (10.1)	5	different rates, (1230) but a model with eudicots and mono-
1220	265.938 (265.346)	3.8 (4.0)	3.6 (3.9)	5	cots having the same rate (1220) is not much worse.
1230	263.057 (262.338)	0 (0)	0 (0)	6	
1111	303.444 (303.943)	12.4 (13.9)	12.2 (13.6)	5	The most complex model (1234) is not the best (although it is
1121	301.267 (301.553)	10.1 (11.1)	9.9 (11.0)	6	close): a model where eudicots and other angiosperms (ex
1222	298.690 (298.818)	4.9 (5.6)	4.8 (5.5)	6	cluding monocots) are forced to have the same rate (1232) is
1212	297.849 (297.776)	3.2 (3.6)	3.1 (3.4)	6	better. Restricting this model further to have Santales and
1123	300.808 (300.533)	11.1 (11.1)	11.1 (11.1)	7	monocots have the same rate (1212) is not much worse.
1232	295.234 (294.995)	0 (0)	0 (0)	7	
1234	295.075 (294.437)	1.7 (0.9)	1.8 (1.0)	8	

1998, 1999), and Butler and King (2004). Lynch's is the only method that scales interspecific variation as a function of intraspecific variation. The model derives expectations from neutral theory to link inter- and intraspecific variation, and uses this theory to estimate whether there has been stabilizing or diversifying selection on morphology for pairs of species (also see Turelli et al. (1988) and Lande (1976, 1979) for related approaches). Because Lynch's method works for pairs of species, the number of phylogenetically independent comparisons is approximately half the number of taxa (although the paper actually does all pairwise comparisons). More importantly, the rate estimates depend on knowing the amount of intraspecific variation and having a model to link this to interspecific variation. Intraspecific variation may be unknown for many potential applications of this method, and the link between intra- and interspecific variation is unclear under many models. For example, if species mean phenotypes are tracking moving adaptive peaks, the rate of movement of these peaks strongly affects the rate at which interspecific variation evolves but does not necessarily have a large effect on the intraspecific variation.

Garland (1992) suggests comparing the absolute values of standardized contrasts in two clades using a *t*-test or a Mann-Whitney *U* test (also known as the Wilcoxon rank sum test) to test for significant difference in rate of evolution in two clades. As shown in Figure 4, the approach developed here is more powerful. A *t*-test would be inappropriate for the simulated data, but if performed, is still less powerful than the methods here. A likelihood, explicitly model-based approach is also potentially easier to extend than the standardized contrasts approach, as model parameters can be added in the future to make the model more realistic or to test for non-Brownian processes, such as limits on character state value, attraction to a mean value, or even repulsion in state from co-occurring taxa. One current disadvantage of our approach, although it is shared with the standardized contrast approach, is the assumption that trait values are errorless observations, ignoring measurement error, variability within the tip taxa, environmental effects on the tip values, etc. With the standardized contrast approach, the rate estimate can change with different rooting of the tree, but that is not the case in the approach developed here. For example, a four taxon unrooted tree [(A,B),(C,D)] with terminal branches of length 1, internal branch of length 2, and terminal values for the four taxa of 60, 50, 10, and 5 has an average standardized contrast magnitude of 12.7 if midpoint rooted and 14.4 if the root is placed halfway along the branch leading to taxon A, whereas our method recovers the same rate in both cases, regardless of rooting. Garland's approach is most naturally used for comparing rates in different clades, although it could likely be used with trees split into subtrees as in our censored approach to allow comparison of a clade with a paraphyletic group.

Foote, in several papers (i.e., Foote 1993, 1994, 1995), uses disparity through time plots to make inferences about changing rates (or ''step sizes'') of morphological evolution. These methods are widely used in paleobiology (e.g., Wesley-Hunt 2005), and have highlighted the interaction between extinction risk and morphology, the empirical pattern of disparity through time, and the effects of morphological con-

straints. Although the methods are informed by simulations, the use of disparity alone, even with information on number of taxa, is insufficient to adequately estimate rate. As demonstrated above (e.g., eq. 1) and in other simulations (i.e., Foote 1996), a tree is needed to link disparity to rate. For example, in our Figure 2 species diversity remained roughly constant across the midpoint of the simulation, morphological rate remained constant, but morphological diversity dramatically decreased due to increased species turnover rate alone. Recovering a resolved tree with informative branch lengths for extinct organisms may be difficult, but using at least some portions of a tree (such as ancestor-descendant pairs) is necessary to make inferences about rate from observations about disparity.

Martins (1994) developed a least squares method to estimate the value of the Brownian motion rate parameter. She discusses comparing rates in different clades, but does not propose a test to determine whether the rate difference is meaningful (although likelihood ratio tests are discussed elsewhere in the paper, in choosing between a Brownian motion model and an Ornstein-Uhlenbeck model). Martins cautions that least squares estimates of the Brownian rate parameter may be nearly undefined if all branches are of equal length due to the use of regression to optimize this parameter; the approach developed here does not share this problem (as long as none of the variance-covariance matrices are singular).

Perhaps the most commonly used approaches for examining the evolution of single continuous characters are contained in Pagel's *Continuous* (Pagel 1994, 1998, 1999). Pagel's approaches are similar to those developed here: parameters are used to transform a tree in various ways and the fit of the characters to this stretched tree is evaluated using likelihood. However, the approaches developed by Pagel use the transformation parameters (κ, δ, and λ) across the entire tree, instead of mapping them to different parts of the tree, as we do (i.e., a different rate parameter for clade A and clade B).

Our paper's approach is most similar to that developed by Butler and King (2004). We can imagine a general model (see Hansen 1997) in which each branch of a tree can have its own optimal trait value, Brownian rate parameter, and Ornstein-Uhlenbeck ''rubber band'' parameter (which describes the strength of the ''pull'' towards a particular optimal value). Note that this most general model is far too over-parameterized. The Butler and King approach restricts the Brownian rate parameter to one (optimized) value across the entire tree while allowing the other parameters to vary, while our noncensored model sets the Ornstein-Uhlenbeck rubber band parameter to zero while allowing the rate parameter to vary. The optimal trait value parameter has less importance when the rubber band parameter is set to zero, but can be considered to be the estimated state at the root of the tree (noncensored model) or at the root of each subtree (censored model). Our censored approach is distinct from their approach in allowing us to ignore the process of change on a particular branch, rather than assuming an instantaneous change from one set of parameter values to another, but the censored approach may easily be adopted for their model, as well. Eventually, we hope to implement the general model described above, allowing tests to vary both the Brownian

rate parameter and the OU rubber band parameter. Currently, as the models implemented in Butler and King's *OUCH* software and *Brownie* are generally nonnested, a comparison of their AIC values may be appropriate. In general, the Butler and King model (multiple optima, multiple attraction parameters, one Brownian rate parameter) may be more appropriate in the case of a few distinct, rarely moving adaptive peaks in phenotype space that organisms transition between infrequently. Our model (multiple rate parameters) may be more appropriate in cases in which the adaptive peaks move across phenotype space as a result of several factors or even under a model of punctuated phenotypic change with many jumps.

[Note that, after acceptance of this manuscript, we received notice of an in press paper by Thomas et al. 2006 which uses a method substantially similar to the noncensored method with likelihood ratio tests described above.]

Extensions

The approaches described here may be extended in several ways. Currently, the methods allow rate parameters to be assigned to monophyletic and paraphyletic subtrees (censored and noncensored approaches) or to individually selected branches (noncensored approach) and then suitability of the model evaluated. All these methods currently require assignment of just one rate per branch. There are several questions, however, that would be better addressed by allowing different portions of a branch to be assigned to different rate parameters. One may want to assign rates to intervals of time rather than to parts of a tree. A typical question might be whether placental mammals had a higher rate of morphological evolution in the Eocene, as they expanded into niches previously occupied by nonavian dinosaurs, than later, once many of those niches were filled. This branch subdivision may be even finer. If the hypothesis is that one particular discrete trait affects the rate of evolution of some continuous trait (for example, whether the evolution of animal-dispersed seeds affects the rate of evolution of plant height), one could use an approach such as Nielsen (2002) to stochastically map the discrete character on the tree and then test to see whether the portions of the tree with one state for the discrete character have a different rate of evolution for the continuous character than portions of the tree to which the other discrete state has been mapped. Another extension, currently in development, would be to try to examine whether the rate of evolution of a continuous character is correlated with the state of a different continuous character, rather than a discrete character. Rates of more than one character could also be examined. Huelsenbeck and Rannala (2003), for example, use the Brownian motion model for multiple characters (their eqs. 5 and 6) to examine character correlation. Lynch (1991) also develops this model, including error terms. In much the same way that the one character model was modified in this paper to test for differences in rate, the multiple character model could be modified to see whether rates or correlations changed as expected under some hypothesis (e.g., whether the correlation between body mass and brain size in mammals is also a function of whether or not the organisms live in complex social groups).

Like many models in comparative biology, the approaches above assume that trait values are known without error. The methods may be modified in at least two ways to deal with error. If measurement variance is known for each of the terminal taxa, the actual variance-covariance matrix will be the sum of a diagonal matrix containing these variances and the evolutionary variance-covariance matrix (created by multiplying covariances based on tree branch lengths by the relevant rate parameter(s), as above). The likeliest values of the rate parameters and the likelihood of the model may then be numerically estimated. If these measurement variances are not known, the measurement error diagonal matrix could be created to have diagonal entries of σ_e^2, which would then be estimated as part of the model. Estimating a different measurement error for each taxon using just one character per taxon would add too many parameters to the model to be practical. With either approach, problems such as *Poncirus* and *Citrus* having different states but zero branch length between them would no longer confound the model. Lynch (1991) also discusses how to deal with error under the Brownian motion model of character evolution.

In paleontology recovering well-supported trees can be difficult. Although the methods here can incorporate uncertainty in the tree by performing analyses across sets of weighted trees, there may be situations in which ancestor-descendant relationships or other fragments of the tree are all that can be estimated. Although power will be lost, it is still possible to perform phylogenetically correct rate comparisons in such cases. For example, if just ancestor-descendant pairs are known, the variance-covariance matrix would be a diagonal matrix for the descendants, with diagonal entries corresponding to the length of time separating the ancestor and descendant in each pair and the vector of expected states being the vector of states of the ancestors. Note, however, that if any of the supposed ancestors are not direct ancestors but rather share a most recent common ancestor with the supposed descendent further back in time, rate estimates will be biased upwards (the phenotypic change occurred over more time than is assumed). Fragments of the tree may be treated in a similar manner.

There may be interest in using the approaches here in an exploratory manner rather than in a hypothesis-testing or model-selection manner. This should be done cautiously. For a fully resolved tree with N taxa, there are $2N$-2 branches, thus there is insufficient information for estimating rates on each branch using one character (about twice as many parameters as data points). Even allowing different rates for each pair of branches, as is done by using standardized contrasts, results in N-1 parameters (including the state at the root) being estimated for N data points, and so, at best, these are very uncertain estimates. Other approaches might be to split the tree at an internal branch and test for different rates in the two resulting subtrees using the censored test, repeating this for all branches. This might be useful for suggesting key branches on which rate changes may occur. As the tests on different branches have different expected power and are not independent, simply comparing results may be inappropriate, although parametric simulation may be useful.

PROGRAM NOTE

The new tests and simulations discussed above, except for the noncensored approach (implemented in *r8s* version 1.70,

http://ginger.ucdavis.edu/r8s/index.html), have been implemented in a set of Matlab functions, collectively called *Brownie* 1.0. They are available from http://www.brianomeara.info/brownie. They will run under Matlab 5.2 or later and require the Stats package. *Brownie* 2.0, a stand-alone Macintosh, Windows, and Linux application that reads Nexus files (including trees) and can perform rate tests, is also available at this site.

LITERATURE CITED

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. *in* B. N. Petrov and F. Csaki, eds. Second International Symposium on Information Theory. Akademiai Kiado, Budapest, Hungary.

Bennett, M. D., and I. J. Leitch. 2005. Plant DNA C-values Database. Rel 4.0 (October). http://www.rbgkew.org.uk/cval/

Blomberg, S. P., T. G. Garland, Jr.,and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:717–745.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information theoretic approach, 2d ed. Springer-Verlag, New York.

Butler, M. A. and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am. Nat. 164:683–695.

Casella, G., and R. L. Berger. 2002. Statistical inference, 2d ed. Duxbury, Pacific Grove, CA.

Chubb, A. L. 2004. New nuclear evidence for the oldest divergence among neognath birds: the phylogenetic utility of ZENK(i). Mol. Phyl. Evol. 30:140–151.

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

———. 2003. PHYLIP (phylogeny inference package). Vers. 3.5c. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, WA.

Fjeldså, J. 1983. Ecological character displacement and character release in grebes Podicipedidae. Ibis 125:463–481.

Foote, M. 1993. Discordance and concordance between morphological and taxonomic diversity. Paleobiology 19:185–204.

———. 1994. Morphological disparity in Ordovician-Devonian crinoids and the early saturation of morphological space. Paleobiology 20:320–344.

———. 1995. Morphological diversification of Paleozoic crinoids. Paleobiology 21:273–299.

———. 1996. ''Models of Morphological Diversification.'' Pp. 62–86 *in* J. Jablonski, D. H. Erwin, and J. H. Lipps, eds. Evolutionary paleobiology, Univ. of Chicago Press, Chicago, IL.

———. 1997. The evolution of morphological diversity. Annu. Rev. Ecol. Syst. 28:129–152.

Garland, T. G., Jr., 1992. Rate tests for phenotypic evolution using phylogenetically independent contrasts. Am. Nat. 140:509–519.

Garland, T. G., Jr.,and A. R. Ives. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. Am. Nat. 155:346–364.

Gatesy, S. M., and K. M. Middleton. 1997. Bipedalism, flight, and the evolution of theropod locomotor diversity. J. Verte. Paleontol. 17:308–329.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351.

Hansen, T. F., and E. P. Martins. 1996. Translating between microevolutionary processes and macroevolutionary patterns: The correlation structure of interspecific data. Evolution 50:1404–1417.

Harvey, P. H., R. M. May, and S. Nee. 1994. Phylogenies without fossils. Evolution 48:523–529.

Hey, J. 1992. Using phylogenetic trees to study speciation and extinction. Evolution 46:627–640.

Huelsenbeck, J. P., and B. Rannala. 2003. Detecting correlation between characters in a comparative analysis with uncertain phylogeny. Evolution 57:1237–1247.

Knight, C. A., N. A. Molinari, and D. A. Petrov. 2005. The large genome constraint hypothesis: evolution, ecology, and phenotype. Ann. Bot. 95:177–190.

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

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

Leitch, I. J., D. E. Soltis, P. S. Soltis, and M. D. Bennett. 2005. Evolution of DNA amounts across land plants (Embryophyta). Ann. Bot. 95:207–217.

Lynch, M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. Am. Nat. 136:727–741.

———. 1991. Methods for the analysis of comparative data in evolutionary biology. Evolution 45:1065–1080.

Magallón, S. A., and M. J. Sanderson. 2005. Angiosperm divergence times: the effect of genes, codon positions, and time constraints. Evolution 59:1653–1670.

Martins, E. P. 1994. Estimating the rate of phenotypic evolution from comparative data. Am. Nat. 144:193–209.

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.

Mayr, G. 2004. Morphological evidence for sister group relationship between flamingos (Aves: Phoenicopteridae) and grebes (Podicipedidae). Zool. J. Linn. Soc. 140:157–169.

Nee, S., A. Ø. Mooers, and P. H. Harvey. 1992. Tempo and mode of evolution revealed from molecular phylogenies. Proc. Natl. Acad. Sci. USA 89:8322–8326.

Nielsen, R. 2002. Mapping mutations on phylogenies. Syst. Biol. 51:729–739.

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

———. 1998. Inferring evolutionary processes from phylogenies. Zool. Scr. 26:331–348.

——— 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Purvis, A. 2004. How do characters evolve? Nature 432.

Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16:276–277.

Ricklefs, R. 2004. How do characters evolve? [response] Nature 432.

Sanderson, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14:1218–1231.

———. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19:101–109.

Schluter, D., T. Price, A. Ø. Mooers, and D. Ludwig. 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51:1699–1711.

Slowinski, J. B., and C. Guyer. 1989. Testing the stochasticity of patterns of organismal diversity: An improved null model. Am. Nat. 134:907–921.

Soltis, D. E., P. S. Soltis, M. W. Chase, M. E. Mort, D. C. Albach, M. Zanis, V. Savolainen, W. H. Hahn, S. B. Hoot, M. F. Fay, M. Axtell, S. M. Swensen, L. M. Prince, W. J. Kress, K. C. Nixon, and J. S. Farris. 2000. Angiosperm phylogeny inferred from 18S rDNA, rbcl, and atpB sequences. Bot. J. Linn. Soc. 133:381–461.

Soltis, D. E., P. S. Soltis, M. D. Bennett, and I. J. Leitch. 2003. Evolution of genome size in the angiosperms. Am. J. Bot. 90: 1596–1603.

Sugiura, N. 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. Comm. Stat. Theory Methods. A7:13–26.

Swofford, D. L. 2003. PAUP*: phylogenetic analysis using parsimony (*and other methods). Ver. 4.0b10. Sinauer Associates, Sunderland, MA.

Thomas, G. H., R. P. Freckleton, and T. Székely. 2006. Comparative analysis of the influence of developmental mode on phenotypic diversification rates in shorebirds. Proc. R. Soc. Lond. B. *In press.*

Turelli, M., J. H. Gillespie, and R. Lande. 1988. Rate tests for selection on quantitative characters during macroevolution and microevolution. Evolution 42:1085–1089.

Van Tuinen, M., D. B. Butvill, J. A. W. Kirsch, and S. B. Hedges. 2001. Convergence and divergence in the evolution of aquatic birds. Proc. R. Soc. Lond. B. 268:1345–1350.

Wesley-Hunt, G. D. 2005. The morphological diversification of carnivores in North America. Paleobiology 31:35–55.

Zelditch, M. L., D. L. Swiderski, H. D. Sheets, and W. L. Fink. 2004. Geometric morphometrics for biologists: a primer. Elsevier Academic Press, San Diego, CA.

Zonneveld, B. J. M., I. J. Leitch, and M. D. Bennett. 2005. First nuclear DNA amounts in more than 300 angiosperms. Ann. Bot. 96:229–244.

Corresponding Editor: T. Hansen

APPENDIX

Equation 1 is derived as follows. $\mathbf{V}$ is the variance-covariance matrix for the states of the terminal taxa; all other notation is as described in the text. By definition,

$$\mathbf{V} \equiv \mathrm{E}(\mathbf{XX}') \tag{A1}$$

if $\mathbf{X}$ is transformed so that the mean is zero.

$$\mathrm{var}(\mathbf{X}) \equiv \frac{X_1^2 + \cdots + X_N^2}{N} - \bar{X}^2 \tag{A2}$$

$$\bar{X} \equiv \frac{X_1 + \cdots + X_N}{N} \tag{A3}$$

$$= \frac{1}{N}(1 \cdots 1)\begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} \tag{A4}$$

$$= \frac{1}{N}\mathbf{1}'\mathbf{X} = \frac{1}{N}\mathbf{X}'\mathbf{1} \tag{A5}$$

since $\bar{X}^2 = \bar{X} \times \bar{X}$, we get:

$$\bar{X}^2 = \frac{1}{N}(\mathbf{1}'\mathbf{X}) \times \frac{1}{N}(\mathbf{X}'\mathbf{1}) = \frac{1}{N^2}\mathbf{1}'(\mathbf{XX}')\mathbf{1} \tag{A6}$$

$$\mathrm{E}[\mathrm{var}(\mathbf{X})] = \mathrm{E}\left(\frac{X_1^2 + \cdots + X_N^2}{N}\right) - \mathrm{E}(\bar{X}^2) \tag{A7}$$

$$= \left[\frac{\mathrm{E}(X_1^2) + \cdots + \mathrm{E}(X_N^2)}{N}\right] - \mathrm{E}\left[\frac{1}{N^2}\mathbf{1}'(\mathbf{XX}')\mathbf{1}\right] \tag{A8}$$

The first term is just the mean of the sum of the diagonal elements of $\mathbf{V}$ (variances), so

$$\mathrm{E}[\mathrm{var}(\mathbf{X})] = \frac{1}{N}\mathrm{tr}(\mathbf{V}) - \frac{1}{N^2}\mathbf{1}'\mathrm{E}(\mathbf{XX}')\mathbf{1} \tag{A9}$$

substituting A1 back in A9, we get

$$\mathrm{E}[\mathrm{var}(\mathbf{X})] = \frac{1}{N}\mathrm{tr}(\mathbf{V}) - \frac{1}{N^2}\mathbf{1}'\mathbf{V}\mathbf{1} \tag{A10}$$

which is the expected disparity under any model of character evolution for which a variance-covariance matrix may be written. In the case of Brownian motion, the variance-covariance matrix $\mathbf{V}$ is equal to $\sigma^2\mathbf{C}$, thus resulting in equation 1.