doi:10.1111/evo.12300

ANCESTRAL CHARACTER ESTIMATION UNDER THE THRESHOLD MODEL FROM QUANTITATIVE GENETICS

Liam J. Revell[1,2]

[1]*Department of Biology, University of Massachusetts Boston, Boston, Massachusetts 02125*

[2]*E-mail: liam.revell@umb.edu*

Received December 15, 2012
Accepted October 13, 2013

Evolutionary biology is a study of life's history on Earth. In researching this history, biologists are often interested in attempting to reconstruct phenotypes for the long extinct ancestors of living species. Various methods have been developed to do this on a phylogeny from the data for extant taxa. In the present article, I introduce a new approach for ancestral character estimation for discretely valued traits. This approach is based on the threshold model from evolutionary quantitative genetics. Under the threshold model, the value exhibited by an individual or species for a discrete character is determined by an underlying, unobserved continuous trait called "liability." In this new method for ancestral state reconstruction, I use Bayesian Markov chain Monte Carlo (MCMC) to sample the liabilities of ancestral and tip species, and the relative positions of two or more thresholds, from their joint posterior probability distribution. Using data simulated under the model, I find that the method has very good performance in ancestral character estimation. Use of the threshold model for ancestral state reconstruction relies on a priori specification of the order of the discrete character states along the liability axis. I test the use of a Bayesian MCMC information theoretic criterion based approach to choose among different hypothesized orderings for the discrete character. Finally, I apply the method to the evolution of feeding mode in centrarchid fishes.

KEY WORDS: **Bayesian, Centrarchidae, comparative method, deviance information criterion, interspecific data, Markov chain Monte Carlo, phylogenetic tree.**

A great deal has been learned by evolutionary biologists by studying contemporary instances of rapid evolution, through computer simulations, and by experimental research on the genetic and phenotypic changes through time of fast reproducing organisms in the laboratory (e.g., Bürger et al. 1989; Rice and Hostert 1993; Losos et al. 1997). Nonetheless, much of evolutionary biology is still fundamentally a historical discipline. Evolutionary biologists are interested in reconstructing the past, and in many instances this may involve estimating something about the features of extinct ancestral species from those of their extant descendants. This estimation of the characteristics of hypothetical ancestral species in the context of a phylogeny is a domain of phylogenetic comparative biology (Brooks and McLennan 1991; Harvey and Pagel 1991; Pagel 1994; Schluter et al. 1997; Nunn 2011).

Phylogenetic comparative biology has taken many bold methodological steps in the past decade. Most significantly, researchers have developed a wide range of new methods to estimate past evolutionary processes from a phylogeny and (in many cases) phenotypic trait data for the extant species in the tree. For instance, we now have sophisticated statistical tools for studying the pace of lineage accumulation through time (e.g., Rabosky 2006; Stadler 2011), we can fit different regimes of phenotypic evolution to different parts of a tree (e.g., Butler and King 2004; O'Meara et al. 2006; Revell and Harmon 2008; Revell and Collar 2009;

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

Eastman et al. 2011; Revell et al. 2012), and we can model and study the interaction of trait evolution and species diversification in a phylogeny (e.g., Maddison et al. 2007; Fitzjohn 2010). Over the same time interval, however, less research has been conducted on the estimation of ancestral states (but see, e.g., Huelsenbeck and Bollback 2001; Nielsen et al. 2002; Huelsenbeck et al. 2003; Pagel et al. 2004; Bokma 2008).

For discretely valued character traits—such as the presence or absence of a feature, or the number of elements in a serially repeated structure—the prevailing method of ancestral state reconstruction uses a model of evolution in which phenotypic changes accrue under a continuous-time discrete-state Markov process (Pagel 1994, 1999; Schluter et al. 1997; Cunningham et al. 1998). This is the same basic model as is used for the evolution of nucleotide sequence in model-based phylogeny inference. The typical form of this model for phenotypic trait data is a generalization of the Jukes–Cantor model (Jukes and Cantor 1969) that has been dubbed the Mk-model (Pagel 1994; Lewis 2001), although it is relatively straightforward to consider related alternative models, such as a model of character change with different forward and backward transition rates between states.

The central attribute of a discrete-state continuous-time Markov model is an instantaneous transition matrix, $\mathbf{Q}$. $\mathbf{Q}$ is a square matrix of dimensions $m \times m$ for m states of a discrete character. It has a general form as follows (for the three-state case):

$$\mathbf{Q} = \begin{bmatrix} -(\alpha+\beta) & \alpha & \beta \\ \gamma & -(\gamma+\delta) & \delta \\ \varepsilon & \zeta & -(\varepsilon+\zeta) \end{bmatrix}.$$

Here, off-diagonals are the instantaneous transition rates from an initial state (row) to a derived state (column). So, for example, the instantaneous transition rate between states 1 and 2 is α. Under this model, the waiting time to a change when a lineage is in state 1 (either to states 2 or 3) is exponentially distributed with rate parameter $\alpha + \beta$, and the expected number of substitutions from state 1 to state 2 given time t is simply $\alpha \times t$. Finally, the matrix of probabilities of change between all states given time t can be computed as $\mathbf{P}(t) = \exp(\mathbf{Q}t)$, where $\exp(\mathbf{X})$ denotes matrix exponentiation.

Under this model, the instantaneous substitution rates may differ between states, and the backward and forward rates of transition can also vary; however, the process is "memoryless," which means that the probability of changing between state 1 and 2 (for instance) does not depend on the prior states for the lineage, nor on the amount of time spent in the current state. (But see Beaulieu et al. 2013 for an interesting new "hidden-rates" model.) The continuous-time Markov chain is primarily a phenomenological model—meaning that it is not linked particularly well with any specific underlying biological process for the evolution of the character traits in our analysis. Many evolutionary biologists would probably agree that the idea that a lineage can change state instantly, with an equal and indefinite probability of reversal, is unrealistic and inconsistent with our views on how complex morphological and ecological characters arise. This model may be fully appropriate (or at least an adequate approximation) for the process of nucleotide substitution and for evolution in some types of discretely valued character traits; however, for complex or polygenic discrete organismal traits, it may be time to consider another model.

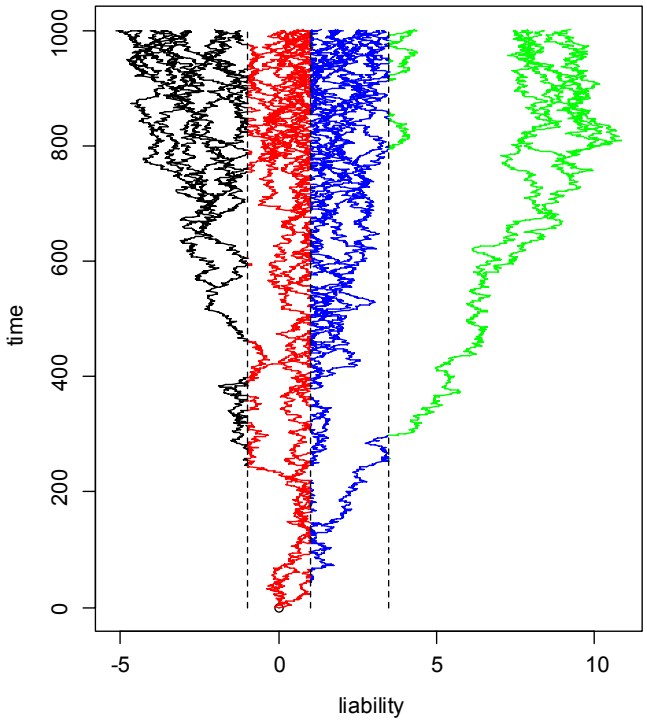

Figure 1. Simulation of evolution under the threshold model on a 40-taxon tree. The abscissa gives the (normally unobserved) liability trait. The discrete threshold character states are shown here as different colors, and the thresholds by vertical dashed lines.

The threshold model of quantitative genetics is a model in which the discrete presentation of an organismal character trait is actually based on an underlying, unobserved continuous trait called "liability" (Wright 1934; Lynch and Walsh 1998; Felsenstein 2005, 2012). The reason this is called the threshold model is because when liability (our unobserved continuous character) exceeds a fixed threshold, the character state changes value (Lynch and Walsh 1998). There may be one or multiple thresholds on our unobserved liability axis. For instance, in Figure 1 I have illustrated evolution on a 40-species phylogenetic tree under the threshold model. Here, the vertical axis gives time since the

root of the tree, color is the observed threshold character, and the abscissa gives our (normally unobserved) liability trait (Fig. 1). On this hypothetical liability axis, a change in liability from −2.0 to 0.0 will, for instance, cause the discrete character to change from state *black* to state *red*, and a further change from 0.0 to 3.0 causes a change in state from *red* to *blue* (Fig. 1). Finally, if liability exceeds 3.5, the observed threshold character changes again—this time to state *green* (Fig. 1).

Liability might be a concrete, yet unmeasured characteristic, such as blood hormone level. In this case, a threshold level of blood hormone, when exceeded, could cause, for example, the expression of a discrete character to be turned on, or vice versa. However, over evolutionary time scales, I would argue that liability may also be a very good proxy for the many and varied changes in organismal genotype and phenotype that underlie, for instance, a discrete shift in habitat use or life history. One important difference between the threshold model and nucleotide models for discrete character evolution on the tree is that although the latter can be made to be explicitly ordered (by constraining the values that can be assumed by certain elements of **Q**; e.g., Skinner and Lee 2010), the former is inherently so. Changes in liability take place on a continuous trait axis, and for a lineage initially in state *red* to change to state *green* (in my example from Fig. 1) it requires a specifically larger amount of evolution of the underlying liability trait than that required for the same lineage to evolve to state *blue* (not to mention evolution through the state *blue*). This may be realistic and advantageous under some circumstances, but could be problematic under others, for instance when the character ordering is uncertain. This will be discussed further below.

The threshold model of quantitative genetics has only recently drawn the attention of comparative biologists. Specifically, in two recent articles Felsenstein (2005, 2012) innovatively developed the threshold model as an approach for studying the evolutionary correlation between multiple discrete characters, as well as between discrete and continuously valued traits—something that had been quite difficult previously (but see Ives and Garland 2010). This article is a natural extension of Felsenstein (2005, 2012), but it nonetheless (to my knowledge) marks the first time that the threshold model has been applied specifically to try and estimate the states for a discretely valued character at ancestral nodes in a phylogenetic tree.

Methods

MATHEMATICAL AND COMPUTATIONAL DETAILS

This method for ancestral character estimation of discrete characters under the threshold model using Bayesian Markov chain Monte Carlo (MCMC) is implemented in the function `ancThresh` in the R package "phytools" (R Core Development Team 2012; Revell 2012). phytools depends extensively on the important core phylogenetics package "ape" (Paradis et al. 2004; Paradis 2012). Both phytools and ape are open-source and distributed freely from their respective project websites and via the Comprehensive R Archive Network, CRAN.

Following Felsenstein (2012), I assume Brownian motion as my model for the evolution of liability on the tree (Cavalli-Sforza and Edwards 1967; Felsenstein 1985); however, it is theoretically straightforward to extend this proposed approach to other models of continuous trait evolution, such as the Ornstein–Uhlenbeck (OU) model (Felsenstein 1988, 2012; Hansen 1997; Butler and King 2004; also see Discussion). To estimate ancestral states under the threshold model, I use Bayesian MCMC to sample liabilities (and values for the liability thresholds) from their joint posterior probability distribution conditioned on a specified sequence of trait values along the liability axis, and on a model for liability evolution.

To sample liabilities and thresholds from their posterior distribution, I use an expression for the likelihood (the probability of our data and tree given our sampled model parameters) with two parts. The first part is the probability of our sampled liabilities for the tips and nodes in our tree, given our evolutionary model. I henceforward denote the vector of sampled liabilities for tips as **x**; the vector of sampled liabilities for all internal nodes, excluding the root node, **a**; and, finally, our sampled liability at the root, a_0. Following Felsenstein (2005), the observed discrete trait values for the tips of the tree are contained in vector **y**. Because liabilities are scaleless (they are unobserved, so we can set them to any scale—as long as we adjust the positions of the thresholds proportionally), we can just fix the rate of liability evolution by Brownian motion at $\sigma^2 = 1.0$. The probability of a set of unobserved liability values under Brownian evolution with $\sigma^2 = 1.0$ is then given by the multivariate normal equation (Felsenstein 1973; Rohlf 2001):

$$P(\mathbf{x}, \mathbf{a}, a_0|\mathbf{C}) = \frac{\exp\left[-\frac{1}{2}\left([\mathbf{x}, \mathbf{a}] - a_0\mathbf{1}\right)' \mathbf{C}^{-1} \left([\mathbf{x}, \mathbf{a}] - a_0\mathbf{1}\right)\right]}{(2\pi)^{(n+i-1)/2} |\mathbf{C}|^{1/2}}. \quad (1)$$

Here, n is the number of tips in the tree, i is the number of internal nodes, and **1** is a conformable vector of 1.0s. **C** is the $(n + i - 1) \times (n + i - 1)$ matrix of variances and covariances between the n tips and $i - 1$ internal nodes (excluding the root), under Brownian evolution with rate $\sigma^2 = 1.0$. For instance, in the simplified example tree of Figure 2, the matrix **C** looks as follows:

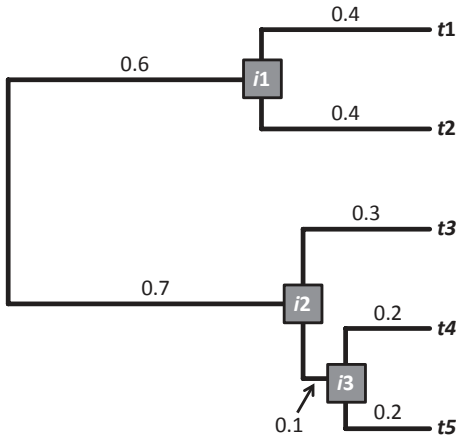

Figure 2. Example tree and branch lengths for the calculation of the among-species and node variance–covariance matrix expected under Brownian evolution with $\sigma^2 = 1.0$.

$$\mathbf{C} = \begin{array}{c} \\ t1 \\ t2 \\ t3 \\ t4 \\ t5 \\ i1 \\ i2 \\ i3 \end{array} \begin{array}{c} \begin{array}{cccccccc} t1 & t2 & t3 & t4 & t5 & i1 & i2 & i3 \end{array} \\ \begin{bmatrix} 1.0 & 0.6 & 0.0 & 0.0 & 0.0 & 0.6 & 0.0 & 0.0 \\ 0.6 & 1.0 & 0.0 & 0.0 & 0.0 & 0.6 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.7 & 0.7 & 0.0 & 0.7 & 0.7 \\ 0.0 & 0.0 & 0.7 & 1.0 & 0.8 & 0.0 & 0.7 & 0.8 \\ 0.0 & 0.0 & 0.7 & 0.8 & 1.0 & 0.0 & 0.7 & 0.8 \\ 0.6 & 0.6 & 0.0 & 0.0 & 0.0 & 0.6 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.7 & 0.7 & 0.7 & 0.0 & 0.7 & 0.7 \\ 0.0 & 0.0 & 0.7 & 0.8 & 0.8 & 0.0 & 0.7 & 0.8 \end{bmatrix} \end{array}.$$

Each i,jth element in $\mathbf{C}$ contains the height above the root for the common ancestor of species or node i and species or node j. The fact that the expected covariances between species are directly proportional to their shared history follows from the Brownian process in which variance accumulates linearly with time, and from the additivity of variances and covariances (Rohlf 2001; Revell 2008). Under an alternative model for the evolution of liability, such an OU model (Hansen 1997; Butler and King 2004), different variances and covariances would be expected.

The second part of our expression for the likelihood is very simple: it is just the probability that our sampled tip liabilities and thresholds could account for our observed interspecific phenotypic pattern for the discrete character. To compute this, we need the function $f(\mathbf{x}, \boldsymbol{\tau})$ that translates our sampled tip liabilities (in $\mathbf{x}$) and thresholds ($\boldsymbol{\tau}$) into predicted values for the discrete character. The only way that our tip liabilities are consistent with our observed data is if $f(\mathbf{x}, \boldsymbol{\tau})$ and $\mathbf{y}$ match exactly. Thus,

$$P(\mathbf{y}|\mathbf{x}, \boldsymbol{\tau}) = \begin{cases} 1 & if \quad f(\mathbf{x}, \boldsymbol{\tau}) = \mathbf{y} \\ 0 & if \quad f(\mathbf{x}, \boldsymbol{\tau}) \neq \mathbf{y} \end{cases} \tag{2}$$

gives the probability of our observed data ($\mathbf{y}$) given our sampled liabilities and thresholds.

$\mathbf{y}$ can be a vector of states, as in this example. However, if there is uncertainty about the tip value for some species, we could imagine instead that $\mathbf{y}$ is a matrix containing the probability of each tip being in each state. In this latter case, we would then evaluate equations (2) as a probability on the interval [0, 1] (rather than just 0 or 1), computed as the product of the probabilities of the trait values predicted by the liabilities sampled at each tip. This is mathematically equivalent to specifying prior probabilities on the tip states, rather than treating them as known exactly. In that case, $P(\mathbf{y}|\mathbf{x}, \boldsymbol{\tau})$ is the conditional posterior probability of our tip states, $\mathbf{y}$, given $\mathbf{x}$ and $\boldsymbol{\tau}$. This option is implemented in `ancThresh` (and used in my empirical example, see below), but, for simplicity, will not be discussed further in this article.

Note that normally only the tip liabilities ($\mathbf{x}$) are relevant here, whereas in equations (1) both tip and internal node liabilities (a_0 and $\mathbf{a}$) are used. This is because we do not have observed phenotypes for internal nodes. If we had some data for internal nodes (such as from fossils) these could theoretically be included in the computation of this probability. (Alternatively, we could include this information in our prior probabilities for internal nodes, see below.)

The likelihood of our sampled liabilities (for tip species and internal nodes; $\mathbf{x}$, a_0, and $\mathbf{a}$) and thresholds ($\boldsymbol{\tau}$), given our data at the tips of the tree ($\mathbf{y}$) and our phylogenetic tree with branch lengths ($\mathbf{C}$) thus becomes the product of the two probabilities of equations (1) and (2):

$$\begin{aligned} &l(\mathbf{x}, a_0, \mathbf{a}, \boldsymbol{\tau}|\mathbf{y}, \mathbf{C}) \\ &= \frac{\exp\left[-\frac{1}{2}\left([\mathbf{x}, \mathbf{a}] - a_0\mathbf{1}\right)' \mathbf{C}^{-1} \left([\mathbf{x}, \mathbf{a}] - a_0\mathbf{1}\right)\right]}{(2\pi)^{(n+i-1)/2} |\mathbf{C}|^{1/2}} \\ &\quad \times \begin{cases} 1 & if \quad f(\mathbf{x}, \boldsymbol{\tau}) = \mathbf{y} \\ 0 & if \quad f(\mathbf{x}, \boldsymbol{\tau}) \neq \mathbf{y} \end{cases}. \end{aligned}$$

Equivalently, the log-likelihood (L) can be written as follows:

$$\begin{aligned} L = &-\left([\mathbf{x}, \mathbf{a}] - a_0\mathbf{1}\right)' \mathbf{C}^{-1} \left([\mathbf{x}, \mathbf{a}] - a_0\mathbf{1}\right)/2 - (n + i - 1) \\ &\times \log(2\pi)/2 - \log(|\mathbf{C}|)/2 - \begin{cases} 0 & if \quad f(\mathbf{x}, \boldsymbol{\tau}) = \mathbf{y}. \\ \infty & if \quad f(\mathbf{x}, \boldsymbol{\tau}) \neq \mathbf{y} \end{cases} \end{aligned}$$

To sample $\mathbf{x}$, a_0, $\mathbf{a}$, and $\boldsymbol{\tau}$ from their joint posterior distribution using Bayesian MCMC (given some starting values), we simply update them sequentially and then accept the updated value with probability given by

$$\begin{aligned} &\min(1, \exp[L(\mathbf{x}', a_0', \mathbf{a}', \boldsymbol{\tau}') + \Pr(\mathbf{x}', a_0', \mathbf{a}', \boldsymbol{\tau}') \\ &\quad - L(\mathbf{x}, a_0, \mathbf{a}, \boldsymbol{\tau}) - \Pr(\mathbf{x}, a_0, \mathbf{a}, \boldsymbol{\tau})]), \end{aligned}$$

where $\Pr(x)$ indicates the log prior probability of x. This MCMC sampler should result in a set of values for the variables in the model obtained in proportion to their posterior probability (Metropolis et al. 1953; Yang 2006).

It seems more sensible to specify prior probabilities for ancestral values of the discrete character than for ancestral liabilities. Thus, an uninformative prior on the liabilities would be equivalent to specifying that all possible states for the discrete character (or at least all states observed among extant species) are equiprobable. Alternatively, if we have extrinsic information (for instance, from the fossil record) that is relevant to the states at internal nodes, we can supply this information via the prior probabilities for that node. During our MCMC chain, then, we use $f([a_0, \mathbf{a}], \boldsymbol{\tau})$ both to compute the prior probabilities of ancestral liability values and to translate sampled ancestral character states to the discrete character space. Our posterior probability of any discrete character value at any internal node is the relative frequency of that state in the posterior sample.

Because liabilities are scaleless, a single threshold can be fixed anywhere—here we arbitrarily fix a single threshold at the value 0.0. If we have more than two character states for our discrete trait then we must also sample the position of the thresholds between character states from their posterior probability distribution. The position of thresholds can only be interpreted relative to the rate of liability evolution (fixed at $\sigma^2 = 1.0$) and each other. This is because, for a lineage evolving under Brownian motion with initial state a_0, a_t (the state for liability after time t) will be normally distributed with mean a_0 and variance $\sigma^2 t$, in which σ^2 is the Brownian rate of liability evolution. For true thresholds $\boldsymbol{\tau} = [0, 2]$ and character states A↔B↔C, the probability of a_t resulting in the threshold character state A will be the integral of the normal density with variance $\sigma^2 t$ and mean a_0 from $-\infty$ to 0. (By the same logic, the probabilities of being in state B or C are the integrals from 0 to 2 and 2 to ∞, respectively.) The trouble is, these probabilities are exactly the same if $\sigma = 0.1$ and $\boldsymbol{\tau} = [0, 0.2]$, $\sigma = 1.0$ and $\boldsymbol{\tau} = [0, 2]$ (the generating values), or $\sigma = 10$ and $\boldsymbol{\tau} = [0, 10]$. This tells us that our discrete character data contain information about σ^2 *or* $\boldsymbol{\tau}$, not both.

The decision to fix the rate of liability evolution at $\sigma^2 = 1.0$ follows Felsenstein (2012), but the specific value at which σ^2 is fixed is arbitrary. Increasing or decreasing σ, the liabilities, and the positions of the liabilities by a constant factor k scales the likelihood in the expression above by factor $1/k$, and thus will cancel from the numerator and denominator of the posterior odds ratio during Bayesian MCMC. This means that (so long as our prior on the thresholds is uninformative or scaled to our choice of σ^2) we will sample from the same posterior distribution of states at internal nodes regardless of our choice of σ^2. The reason that the likelihood of the liabilities is influenced by the scale of σ, whereas the probability of individual character states are not (so long as the thresholds are scaled proportionately), is because the former is the product of values for a probability density function, whereas the latter is an integral computed on an interval that changes width with the position of the thresholds.

As noted by Felsenstein (2012), this changes slightly if there is within-species variation (i.e., within-species polymorphism) in the threshold character. In this case, we can choose to either fix σ^2 and estimate the within-species variation in liability relative to σ^2, or fix the within-species variation in liability and estimate σ^2; however, as with σ^2 and $\boldsymbol{\tau}$, we cannot estimate both.

SIMULATION TESTS

For all analysis of simulated data, I conducted numerical simulations under the threshold model from evolutionary quantitative genetics. I simulated 100-taxon phylogenies under a constant-rate pure-birth (Yule) process, and rescaled all trees to a total height of 1.0. To simulate under the threshold model, I first evolved a continuous character (liability) on the tree and recorded all tip and internal node states. Then I used a set of arbitrary thresholds (constant for each set of simulations) to translate simulated liabilities to discrete character states for all nodes and tips. For these simulations I used a single, constant rate of liability evolution ($\sigma^2 = 1.0$), and the following four sets of thresholds: I simulated three states (A, B, and C) for the discrete character with liability thresholds [0, 1] and [0, 2], and I simulated four states (A, B, C, and D) for the discrete trait with thresholds [0, 1, 2] and [0, 1, 4]. I specified the ancestral liability (and threshold character) at the root node of the tree by drawing a random value uniformly from the range given by the liability thresholds ± 1. I also rejected and repeated any simulation in which any character state was represented by fewer than 5% of terminal species (i.e., five species, in this case). Simulating under the threshold model is straightforward; however, because these analyses require very computationally intensive Bayesian MCMC, I performed only 20 replicate simulations for each set of conditions. To maximize comparability, I used the same 20 simulated trees for each simulation condition.

I evaluated the method using the following approach. First, for each simulated dataset and tree, I ran the Bayesian MCMC for 500,000 generations using the default conditions. These defaults included a variance on the normal proposal distribution for liabilities of 0.5, and a proposal variance of 0.05 for the position of thresholds on the liability axis. I used a uniform prior on ancestral character states, and an exponential prior with a rate parameter $\lambda = 0.01$ on the thresholds.

I sampled every 100 generation and automatically eliminated the first 100,000 generations (i.e., 1001 samples, including the initial state) as burn-in. From the post burn-in posterior sample I computed the posterior probabilities for each node in each state. I also extracted "estimates" of the ancestral character state at each node as the state with the maximum posterior probability. I compared these estimates to the generating values. In addition, I computed estimates based only on nodes in which one state had a greater than 0.5, 0.7, or 0.9 posterior probability. I then compared each of these sets of estimates to their generating values. Finally, I

estimated the relative positions of the liability thresholds (and 95% high probability intervals) from the posterior sample of thresholds.

In addition to these measures, I evaluated the accuracy of the posterior probabilities under the model by asking if, for each posterior probability x of being in state Y, that is, $PP(Y) = x$, the frequency of simulations actually in state Y was (on average) x. In other words—if we bin all nodes with a posterior probabilities of $PP(\mathrm{A}) = 0.3$, we predict that 30% of those should be in state A. To compute this measure, I used a moving window approach with a bin width of 0.025 and a step size of 0.025. This results in a set of 40 nonoverlapping bins between 0 and 1.

As this study is designed primarily as "proof-of-concept" for ancestral character estimation under the threshold model using Bayesian MCMC, I do not focus on optimizing chain convergence in my replicated simulations; however, it should be noted that standard MCMC diagnostics can be used to estimate effective sample size (ESS), convergence, and mixing of the Markov chain. I computed and will describe the results of these diagnostics for my demonstrative empirical analysis of feeding mode evolution on the tree of Centrarchidae below. It is also straightforward to run multiple chains with different or random starting conditions, and to combine the post burn-in samples of separate chains. All of these functions can be conducted easily in R using, for example, the "coda" package (Plummer et al. 2006).

In all of the above analyses, I assumed that the ordering of discrete character states along the liability axis is known a priori by the researcher (although the relative positions of the liability thresholds are not known, but sampled from their posterior probability distribution). In some evolutionary scenarios the ordering of the discrete characters may be obvious. For example, in the evolution of digit number of the manus and pes of scincid lizards, it is probably reasonable to think that digits are added or subtracted in unitary increments (Skinner and Lee 2010; also as in the classic threshold character example of polydactyly in guinea pigs, Wright 1934). However, in other circumstances for evolution under the threshold model a sequence on the liability axis might be less apparent. I propose that multiple alternative orderings be compared using an information theoretic approach. Here, I investigate the possibility of using DIC (deviance information criterion; Spiegelhalter et al. 2002), a Bayesian MCMC analog of the AIC (Akaike information criterion; Akaike 1974; Burnham and Anderson 2002), which is convenient in that it uses the likelihoods from the posterior sample and the likelihood for the mean values of the parameters from the posterior—both of which can be readily computed for this model.

To test whether using DIC could be effective in recovering the true liability threshold ordering, I reused the datasets simulated with thresholds [0, 1]. For each of the 20 simulations, I performed MCMC using each of the three unique orderings of the three character states implied by this model. For $m = 3$ character states there are only $m!/2 = 3$ possible orderings because the orders A↔B↔C and C↔B↔A are equivalent. I calculated and compared the DIC of each simulation using each of the three orderings. I then computed the fraction of orderings across simulations in which the true order had the best DIC (compared to the fraction expected by chance, i.e., 1/3), as well as the mean ΔDIC for each simulation and ordering.

COMPARISON TO OTHER METHODS

I believe that the strongest justification for use of this method comes from its improved biological realism, particularly for complex morphological or ecological characteristics; however, it's natural that many readers will be curious about how it performs in comparison to more phenomenological approaches, especially the widely used Mk or nucleotide model. To conduct this comparison, I used the following procedure. I fit a series of four different versions of the continuous-time discrete-state Markov model to the data from each simulation conducted above. These models were as follows: an equal rates model, in which the rate of change between all three or four characters are assumed to be equivalent; a symmetric model, which allows different rates of change between pairs of states, but changes between all states are theoretically possible; a one-rate, ordered model, in which changes are only permitted between adjacent character states on the liability axis, but the rate of change between all states adjacent in this ordered sequence are equal; and an ordered multirate model, in which only changes between adjacent states on the liability axis are permitted, and two or three different rates are allowed for the backward and forward transitions between adjacent states (depending on the total number of states in our dataset). I optimized each model and computed the empirical Bayesian posterior probabilities that each node was in each state (i.e., marginal ancestral state reconstruction; Yang 2006), and then applied the moving window approach described above to assess the accuracy of the posterior probabilities by comparing them to the observed frequencies across simulations.

In addition to this test, I also conducted the reciprocal analysis. I fit the three-rate ordered model to each of the four-state, three-threshold datasets with thresholds [0, 1, 4]. Using the average fitted value of the transition matrix, $\mathbf{Q}$, across all simulations, I simulated 20 datasets under the continuous-time discrete-state Markov process. I fit the threshold model to each tree and dataset using Bayesian MCMC, and compared the posterior probabilities from the MCMC to the frequencies of node in each state from the simulations. This final simulation asks how the threshold model performs when the data have evolved in an ordered sequence (as in the threshold model), but when the evolutionary process for those data more closely reflects a continuous-time discrete-state Markov chain.

Table 1. Feeding modes for the 32 centrarchid fish species used in this study.

	Feeding mode		
Species	Nonpiscivorous	Moderately piscivorous	Highly piscivorous
Acantharchus pomotis	0.00	1.00	0.00
Lepomis gibbosus	1.00	0.00	0.00
Lepomis microlophus	1.00	0.00	0.00
Lepomis punctatus	1.00	0.00	0.00
Lepomis miniatus	1.00	0.00	0.00
Lepomis auritus	1.00	0.00	0.00
Lepomis marginatus	1.00	0.00	0.00
Lepomis megalotis	1.00	0.00	0.00
Lepomis humilis	1.00	0.00	0.00
Lepomis macrochirus	1.00	0.00	0.00
Lepomis gulosus	0.00	1.00	0.00
Lepomis symmetricus	1.00	0.00	0.00
Lepomis cyanellus	0.00	1.00	0.00
Micropterus cataractae	0.00	0.00	1.00
Micropterus coosae	0.00	1.00	0.00
Micropterus notius	0.00	1.00	0.00
Micropterus treculi	0.00	0.50	0.50
Micropterus salmoides	0.00	0.00	1.00
Micropterus floridanus	0.00	0.00	1.00
Micropterus punctulatus	0.00	0.00	1.00
Micropterus dolomieu	0.00	0.00	1.00
Centrarchus macropterus	0.50	0.50	0.00
Enneacanthus chaetodon	1.00	0.00	0.00
Enneacanthus gloriosus	1.00	0.00	0.00
Enneacanthus obesus	1.00	0.00	0.00
Pomoxis annularis	0.00	1.00	0.00
Pomoxis nigromaculatus	0.00	1.00	0.00
Archoplites interruptus	0.00	1.00	0.00
Ambloplites ariommus	0.00	0.50	0.50
Ambloplites rupestris	0.00	1.00	0.00
Ambloplites cavifrons	0.00	1.00	0.00
Ambloplites constellatus	0.50	0.50	0.00

Data are from Collar et al. (2009) and D. Collar (pers. comm.). For information on the source of these data, see Collar et al. (2009). Species with uncertain feeding mode (for instance, moderate or high piscivory) were assigned equal probability of being in each plausible state.

EMPIRICAL EXAMPLE

In addition to these simulation tests, I also applied the threshold model for ancestral character estimation to an empirical dataset for the evolution of feeding mode in Centrarchidae (the sun fish). Centrarchids can be classified as completely nonpiscivorous, moderately piscivorous, or highly piscivorous. Here I use the multigene centrarchid phylogeny of Near et al. (2005) and the data for feeding mode given in Collar et al. (2009), as well as some additional states not presented in that article because data for other characters were not available (D. Collar, pers. comm.). Feeding mode is uncertain for some lineages (Collar et al. 2009; D. Collar, pers. comm.). The states for feeding mode (including ambiguity) that I used in this study are given in Table 1. For species with uncertain feeding mode, I assumed that the tip species was in each of the possible states with equal probability, which is exactly equivalent to setting a flat prior probability distribution for the possible states of that tip. In addition, I assumed the trait order *not* ↔ *moderately* ↔ *highly* for the trait "piscivorous" evolving as a threshold character. I analyzed the data and tree using the phytools function `ancThresh` under the default conditions for 1,000,000 generations, sampled every 100 generations, and rejected the first 200,000 generations (2001 samples, including the initial state) as burn-in. I tested for convergence to the posterior distribution for liabilities and the position of the thresholds using Geweke's (1992)

Table 2. Summary of results from ancestral character estimation using simulated data and trees.

Model	max(*PP*)	$PP > 0.50$	$PP > 0.70$	$PP > 0.90$	Threshold(s)		Reject
[0,1]	0.846	0.848	0.893	0.947	0.895	NA	29%
		(0.999)	(0.795)	(0.506)	(0.71, 1.10)		
[0,2]	0.907	0.907	0.943	0.959	1.439	NA	78%
		(0.999)	(0.887)	(0.674)	(1.16, 1.71)		
[0,1,2]	0.792	0.793	0.863	0.925	0.894	1.814	78%
		(0.996)	(0.722)	(0.389)	(0.70, 1.07)	(1.59, 2.05)	
[0,1,4]	0.885	0.886	0.928	0.963	0.672	2.565	>99%
		(0.994)	(0.858)	(0.676)	(0.49, 0.87)	(2.14, 3.00)	
Mean	0.858	0.859	0.907	0.949			
		(0.997)	(0.816)	(0.561)			

Columns headings are as follows: Model, the generating model (thresholds) used for simulation; max(*PP*), the mean frequency of correctly inferred nodes based on choosing a point estimate based on the maximum posterior probability; *PP* > 0.50, the mean frequency of correctly inferred nodes if only nodes in which one state had a posterior probability greater than 0.50—number in parenthesis give the mean fraction of nodes used; *PP* > 0.70, the same, but with a cutoff of 0.70; *PP* > 0.90, the same, but with a cutoff of 0.90; Threshold(s), the mean (and mean 95% HPD interval) for the position of one or more thresholds; Reject, the rejection rate of simulated datasets based on the criterion that at least 5% of tip nodes should be in each discrete character state. Rejected simulations were invariably repeated, thus the same number of simulations were used for each row of the table.

convergence diagnostic. I also computed the ESS of each posterior sample. ESS is our sample size from the posterior probability distribution that takes into account the autocorrelation of adjacent samples in the chain. I computed the posterior probability of each state at each node as the posterior frequency from the sample. I estimated the relative position of the thresholds as well as their 95% high probability density intervals (HPDs). Finally, for tips of uncertain value, I also computed their posterior probabilities of being in each state based on the model.

Results

SIMULATION TESTS

A summary of my tests of the accuracy of this new method for ancestral character estimation under various conditions is given in Table 2 and Figure 3. The average results across all simulations conditions are also given. In general, ancestral character estimation by this method can be quite accurate when the data on the tree have evolved under the assumed model. For instance, when ancestral state estimates are based on the state with the highest posterior probability, the true (i.e., simulated) ancestral state was correctly picked in more than 85% of instances averaged across all the models studied (Table 2). When only nodes in which one state was preferred with over 0.90 posterior probability are considered (over 55% of nodes, across all simulations), the correct state was selected on average nearly 95% of the time (Table 2).

In addition to this basic analysis, I also tested for the accuracy of the posterior probabilities from the Bayesian MCMC under the threshold model. Specifically, I used a moving window to ask if, for all posterior probability x of being in state Y, $PP(Y) = x$, the frequency of simulations actually in state Y was (on average) x. The results from this analysis across all four sets of simulations show strongly that posterior probabilities obtained assuming the threshold model accurately reflect our certainty in ancestral character states (Fig. 3).

Accuracy in estimating the location of the thresholds on the liability axis was mixed. In general, the estimated thresholds were downwardly biased, and in some cases this bias was quite substantial (Table 2). It seems highly possible that this is due to the rejection procedure I used during simulation. Specifically, I repeated any simulation resulting in fewer than 5% of tips in each state. For some simulation conditions, the consequence of applying this criterion to the simulations was an extremely high rate of rejection (Table 2). This type of rejection procedure will tend to favor simulations in which the realized variance of liabilities (relative to the position of the thresholds) is high. Because the evolutionary variance (Brownian rate) is fixed during estimation, this selection bias should cause the estimated thresholds to be closer to one another than the true thresholds. For simulations with more than one "estimable" threshold (i.e., a threshold neither fixed at 0 nor set to ∞), we might still anticipate that the expected value of their ratio would be unaffected. Indeed, this seems to be the case. Specifically, for thresholds [0, 1, 2] the mean ratio of thresholds 2 and 3 is 2.07 (very close to the generating value of 2.0); similarly, for thresholds [0, 1, 4] the mean ratio is 4.30 (relatively close to the generating ratio of 4.0).

I also evaluated the prospect of using an information theoretic approach to choose among possible orderings of the discrete character on the liability axis. For this analysis, I used the data simulated with thresholds [0, 1]. Given two thresholds there are three possible states for the discrete character, and in this case they were simulated with the sequence A↔B↔C. I reran the MCMC

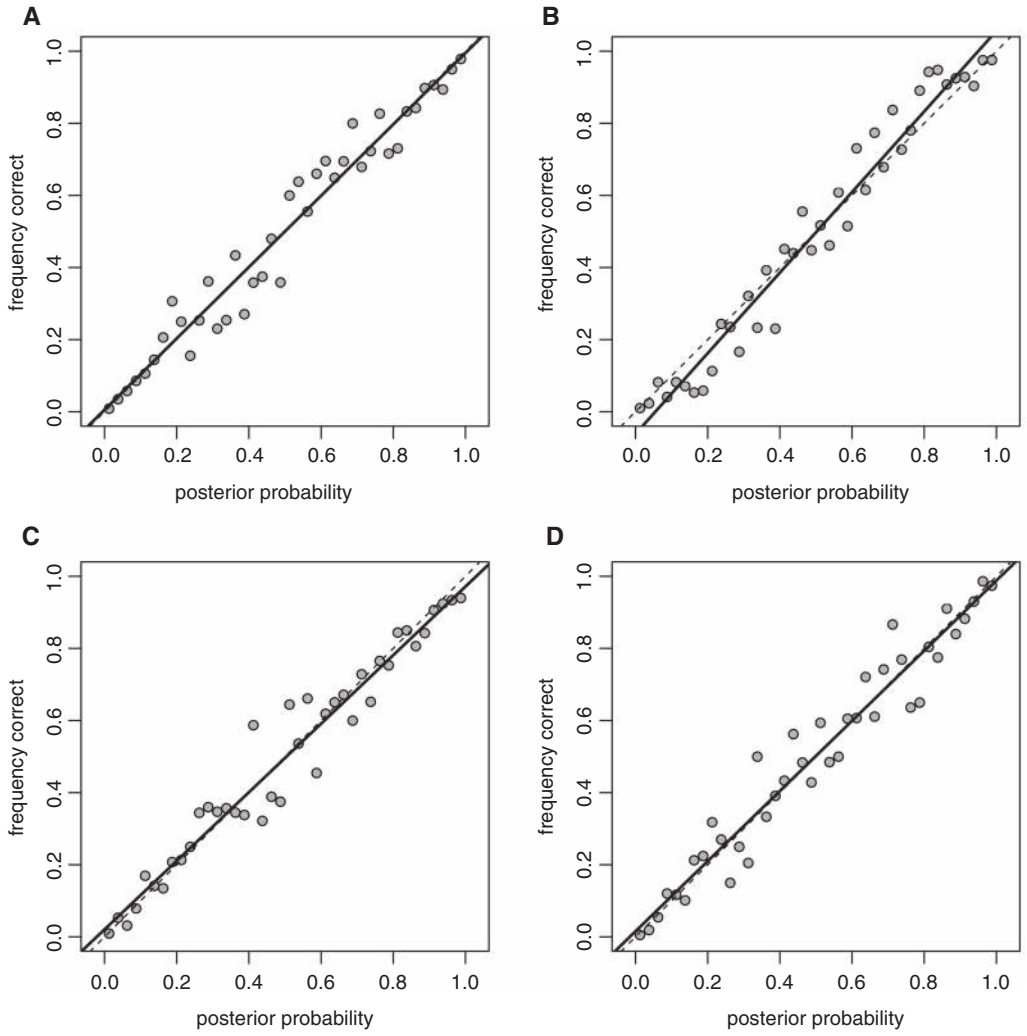

Figure 3. **Accuracy of posterior probabilities from the threshold model when data are generated under the model. Accuracy is measured as the fraction of times in which a node with posterior probability of state *Y* equal to *x* (i.e., *P P*(*Y*) = *x*) is in state *Y*. Points were obtained by aggregating close posterior probabilities and frequencies using a "moving-window" approach. The closer the points lie to the 1:1 line (dashed), the more accurate the posterior probabilities. The solid line in each plot is the ordinary least-squares regression of frequency on the posterior probabilities. Each panel shows the result from a different simulation scenario, as follows: (A) three states (A↔B↔C) with thresholds [0, 1]; (B) three states (A↔B↔C) with thresholds [0, 2]; (C) four states (A↔B↔C↔D) with thresholds [0, 1, 2]; and (D) four states (A↔B↔C↔D) with thresholds [0, 1, 4].**

for each simulated dataset, but incorrectly assumed the two alternative possible sequences of A↔C↔B and B↔A↔C. (Note that there are only $m!/2 = 3$—not $m!$—unique sequences for $m = 3$ states of the discrete character because the assumed sequences $X \leftrightarrow Y \leftrightarrow Z$ and $Z \leftrightarrow Y \leftrightarrow X$ are equivalent for all X, Y, and Z.) For each simulation, I computed the DIC (Spiegelhalter et al. 2002) and chose the ordering with the lowest DIC. DIC is analogous to AIC but is calculated from the mean deviance, $\bar{D} = -2\overline{L(\theta)}$, in the posterior sample along with the likelihood at the posterior sample parameter means, $D(\hat{\theta}) = -2L(\hat{\theta})$, as follows:

$$\begin{aligned} \text{DIC} &= p_D + \bar{D} \\ &= 2\bar{D} - D(\hat{\theta}). \end{aligned}$$

Here, $p_D = \bar{D} - D(\hat{\theta})$ is an estimate of model parameterization. This is because the greater the effective number of parameters in a model, the greater the difference between the likelihood of any sample from the posterior and the likelihood of the optimized parameters. DIC, unlike AIC, does not require an analytical or numerical solution of the likelihood function.

Table 3 shows the results of this analysis. DIC favors the generating model in 55% of simulations, which is significantly more frequently than expected by chance ($\chi^2 = 6.1$, $P = 0.047$). In addition, average ΔDIC is much lower for the generating model than for other sequences of character evolution on the liability axis (Table 3). One notable (and apparently discrepant) result from these simulations is that the log-likelihood for the mean

Table 3. Summary of results for the deviance information criterion (DIC) analysis of different assumed discrete character ordering on the liability axis.

Generating model	Assumed model	min (DIC)	$\overline{\Delta DIC}$	max[$L(\hat{\theta})$]
A↔B↔C	A↔B↔C	0.55	5.184	0.85
A↔B↔C	A↔C↔B	0.10	17.23	0.15
A↔B↔C	B↔A↔C	0.35	14.84	0.00

In all cases, the generating order was A↔B↔C. Column headings are as follows (if not obvious): min(DIC), the frequency, from 20 simulated datasets, in which each assumed ordering had the minimum DIC; $\overline{\Delta DIC}$, mean difference in DIC from minimum for each assumed ordering; and max[$L(\hat{\theta})$], the frequency, from 20 datasets, in which each ordering had the maximum likelihood at the mean parameters from the posterior sample ($\hat{\theta}$).

parameter values from the posterior sample, $L(\hat{\theta})$, is consistently higher under the generating model than when other sequences of character evolution are assumed ($\chi^2 = 24.7, P < 0.001$; Table 3), in spite of more variable results from DIC. This can only be due to higher estimated parameterization (p_D) under the generating model than under other models. Because the true number of variables in each model is the same, the reason for this is unclear; however, one possible explanation for this pattern is that it is the spurious by-product of some of the MCMC runs under the generating model failing to converge to the posterior distribution during burn-in.

In addition to these results based on DIC, my analyses suggest that getting the sequence correct has important consequences for the accuracy of ancestral character estimation under the threshold model. Table 4 shows the accuracy of ancestor state estimation under each of the three assumed models. Accuracy tends to be 5–14% higher for the correct character sequence than when an incorrect sequence is assumed. This difference may even understate the "real" accuracy improvement resulting from the correct sequence, because in many simulations, the majority of nodes are unambiguous (e.g., see Empirical Example) and would probably be estimated with quite high accuracy under any model.

COMPARISON TO OTHER METHODS

Although the results above show that the threshold model can be quite accurate and that the Bayesian posterior probabilities correctly reflect our uncertainty about ancestral states under the model, many readers are probably curious about how ancestral character estimation under the threshold model compares to conventional, phenomenological models—specifically, the Mk model. To address this, I took the results from simulations under the threshold model described above, and fit four different flavors of the continuous-time Markov chain model: an equal rates model, a symmetric model, a one-rate ordered model, and a multirate ordered model. Under each model, I computed the empirical Bayesian posterior probabilities (Yang 2006) that each node was in each state, and then I used a moving window approach to compare these posterior probabilities to the frequencies with which each node was actually in each state, across all replicate simulations.

The results from all simulation conditions were qualitatively similar. Figure 4 shows one exemplar result, taken from the set of simulations conducted using character states A↔B↔C↔D and thresholds [0, 1, 4]. Comparison to Figure 3D (the plot for these datasets obtained from Bayesian MCMC analysis under the threshold model) shows that the continuous-time Markov model does not accurately estimate the true posterior probabilities—although its performance improves as increasingly realistic features (i.e., ordered character states, multiple rates) are added to the model.

In addition to this comparison, I also conducted the converse test in which I simulated under the continuous-time Markov chain and then estimated ancestral states using the threshold model. For these simulations, I used the mean fitted value of the transition matrix **Q** from Figure 4D for simulation. The results are shown in Figure 5A. Unsurprisingly (and, to some extent, reassuringly), the generating model, in this case a continuous-time Markov chain

Table 4. Summary of the accuracy of ancestral character estimation for different assumed orderings of the discrete character along the liability axis.

Generating model	Assumed model	max(*PP*)	$PP > 0.50$	$PP > 0.70$	$PP > 0.90$	Threshold
A↔B↔C	A↔B↔C	0.846	0.848	0.893	0.947	0.895
			(0.999)	(0.795)	(0.506)	(0.71, 1.10)
A↔B↔C	A↔C↔B	0.755	0.781	0.851	0.911	0.461
			(0.920)	(0.679)	(0.416)	(0.36, 0.60)
A↔B↔C	B↔A↔C	0.713	0.734	0.807	0.898	0.678
			(0.943)	(0.714)	(0.447)	(0.55, 0.83)

Column headers are the same as in Table 2.

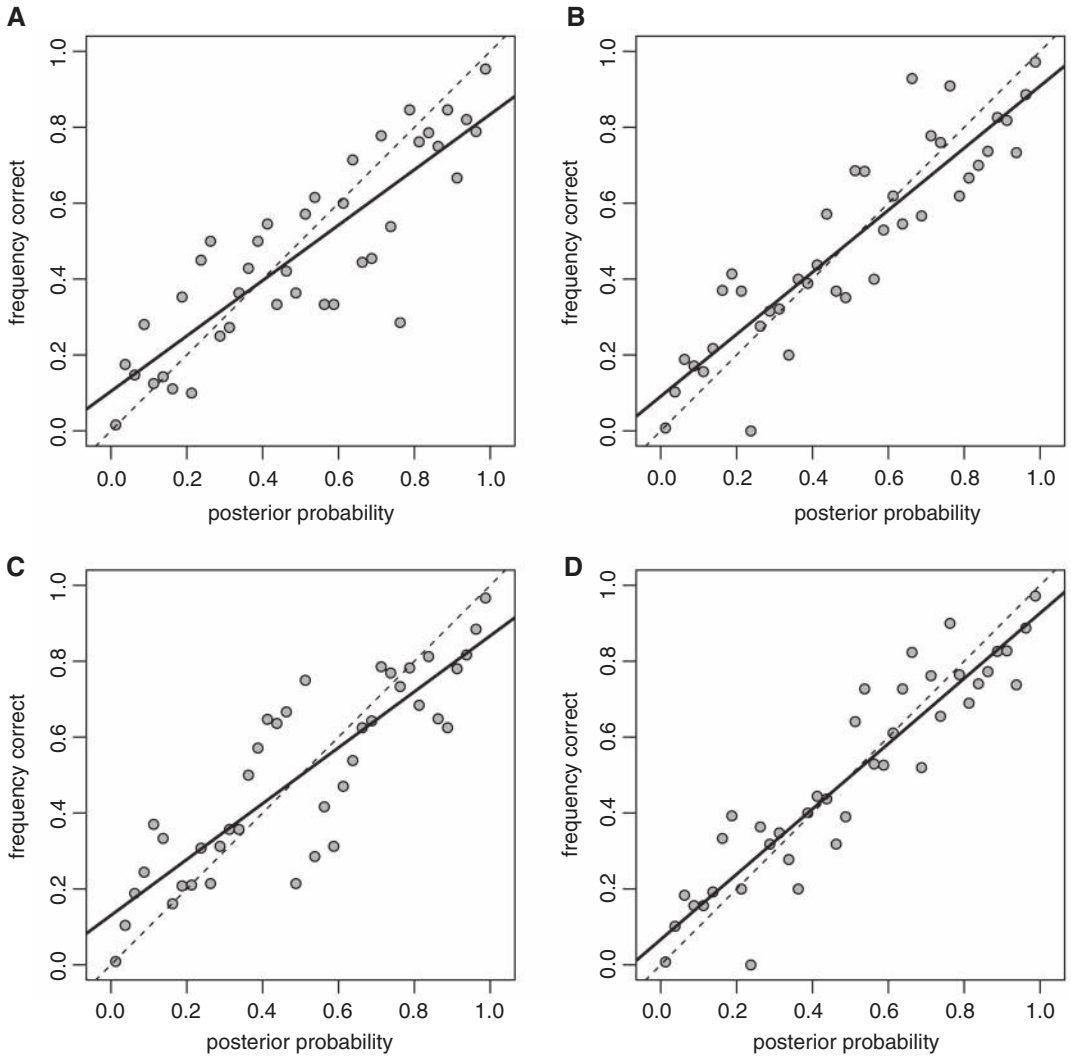

Figure 4. Accuracy of posterior probabilities from several variants of the continuous-time discrete-state Markov model (Mk-model) when data are generated under the threshold model. Data are from a simulation with four states (A↔B↔C↔D) and thresholds [0, 1, 4] (Fig. 3D). Models are as follows: (A) equal rates model, in which a single transition rate is modeled and transitions are allowed between all states; (B) symmetric model, all transitions are theoretically possible, but all forward and backward transitions are modeled with a different rate; (C) ordered, single-rate model, in which only transitions A↔B↔C↔D are permitted, and all permitted transitions have the same rate; and (D) ordered, multirate model, in which transitions A↔B↔C↔D are permitted, and each transition type can have a different rate. Interpretation of the accuracy of posterior probabilities is as in Figure 3.

model, substantially outperforms the threshold model in estimation (Fig. 5B).

EMPIRICAL EXAMPLE

In my empirical analysis of feeding mode evolution in Centrarchidae, I first computed Geweke's (1992) convergence diagnostic and ESSs from the posterior sample for all the variables (liabilities and thresholds) in the model. Geweke's convergence test and ESS calculations are implemented in the R package coda (Plummer et al. 2006). Geweke's (1992) convergence diagnostic showed MCMC convergence for all but two of 64 model variables (63 liabilities and one threshold). This test compares the first and last parts of a (post burn-in) MCMC chain to ask whether they come from the same distribution. The test statistic has a standard normal distribution under the null; thus we would expect 5%, or about three tests, to be significant at the $\alpha = 0.05$ level even if convergence had been achieved by the MCMC chain. The mean absolute value of Geweke's test statistic was 0.82 (range: −3.42, 1.90). Mean ESSs for all parameters was 282.2 (range: 32.4, 1208.5).

Figure 6A shows the results from ancestral character estimation under the threshold model for feeding mode in centrarchid fishes. Assuming this model of trait evolution, my results suggest that the ancestor of the group was most likely moderately piscivorous or nonpiscivorous. If the former, then nonpiscivory evolved

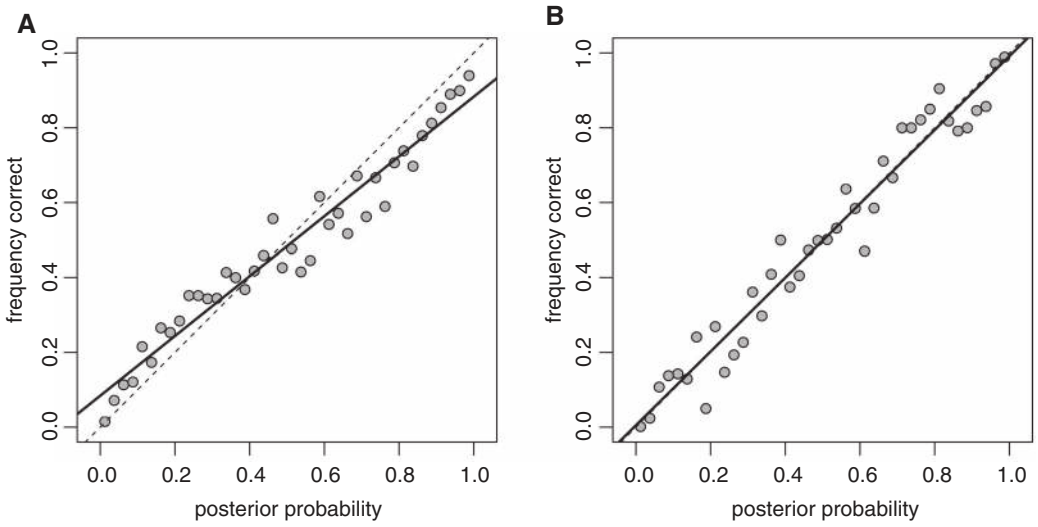

Figure 5. Accuracy of posterior probabilities from the threshold model and the Mk-model when data are generated under the Mk-model. Simulation was of a four-state character using the mean fitted transition matrix, *Q*, from the simulations of Figure 3D and the results from Figure 4D. Fitted models were as follows: (A) the threshold model; and (B) an ordered, multirate continuous-time discrete-state Markov model.

in at least two separate genera on the tree: *Lepomis* and *Enneacanthus*, both clades dominated by nonpiscivorous species. High piscivory evolved only in *Micropterus*, but this occurred either more than once or with multiple reversions to moderate piscivory.

Figure 6B gives an estimate of the posterior density for the location of the threshold between moderate piscivory and high piscivory. Because liabilities are scaleless, only the relative—not absolute—value of the positions of the thresholds are meaningful. The threshold between nonpiscivory and moderate piscivory is fixed at 0. The distance between this threshold and the mean value from the posterior sample for the threshold between moderate piscivory and high piscivory is 0.86. This is very close to the expected absolute value of a change in liability over one time unit of $\sigma\sqrt{2/\pi} = 0.80$ for a rate of liability evolution set to $\sigma^2 = 1.0$. The sum of the branch lengths in the tree is about 3.28; thus one interpretation of the position of the threshold tells us that about three to four times the evolutionary distance between the two thresholds should be traversed (on average) over the entire history of this clade.

DISCUSSION

Phylogenetic comparative biology provides some of the most important tools for making evolutionary inferences about the biological past (Felsenstein 1985, 2004; Brooks and McLellan 1991; Nunn 2011). For many years, evolutionary biologists have been fascinated by the prospect of estimating the ancestral characteristics of extinct species from the phenotypic traits of living taxa related by a phylogeny. In fact, this was one of the most popular endeavors among phylogenetic comparative biologists for over a decade. In recent years, the practice of ancestral character estimation has garnered both considerable criticism and skepticism (e.g., Cunningham et al. 1998; Omland 1999; Ekman et al. 2008; Losos 2009; Skinner and Lee 2010). Many critiques focus on the fact that ancestral state estimation can be vulnerable to very large uncertainty, imposing severe limits on the nature of evolutionary inferences from ancestral phenotype reconstructions (e.g., Losos 1999, 2009).

What has received less attention is the sensitivity of ancestral state estimates to the adequacy of our phylogenetic models (but see, e.g., Ryan and Rand 1999; Ekman et al. 2008; Goldberg and Igic 2008; Skinner and Lee 2010). Specifically, the classic "nucleotide model"—in other words, a model in which discrete character change occurs according to a discrete-state continuous-time Markov process—may be inadequate for complex morphological, behavioral, or physiological traits. For instance, in the present study I analyzed feeding mode in Centrarchidae, the sunfishes. The fish species in this study are classified into three feeding classes (nonpiscivorous, moderately piscivorous, and highly piscivorous) based on the results of considerable prior study. Because feeding mode in fishes involves extensive anatomical, behavioral, and physiological adaptation (Collar et al. 2009), it seems unrealistic to model this trait's evolution on the tree using a process in which the trait can change between any pair of states for feeding mode instantaneously and with an indefinite constant probability of reversion. I would argue that it is more realistic to assume that feeding mode evolves as a function of many small, incremental changes to underlying, unobserved, or unmeasured attributes of the species; and, furthermore, that the longer a species has exhibited a constant feeding mode, the less likely it should be to

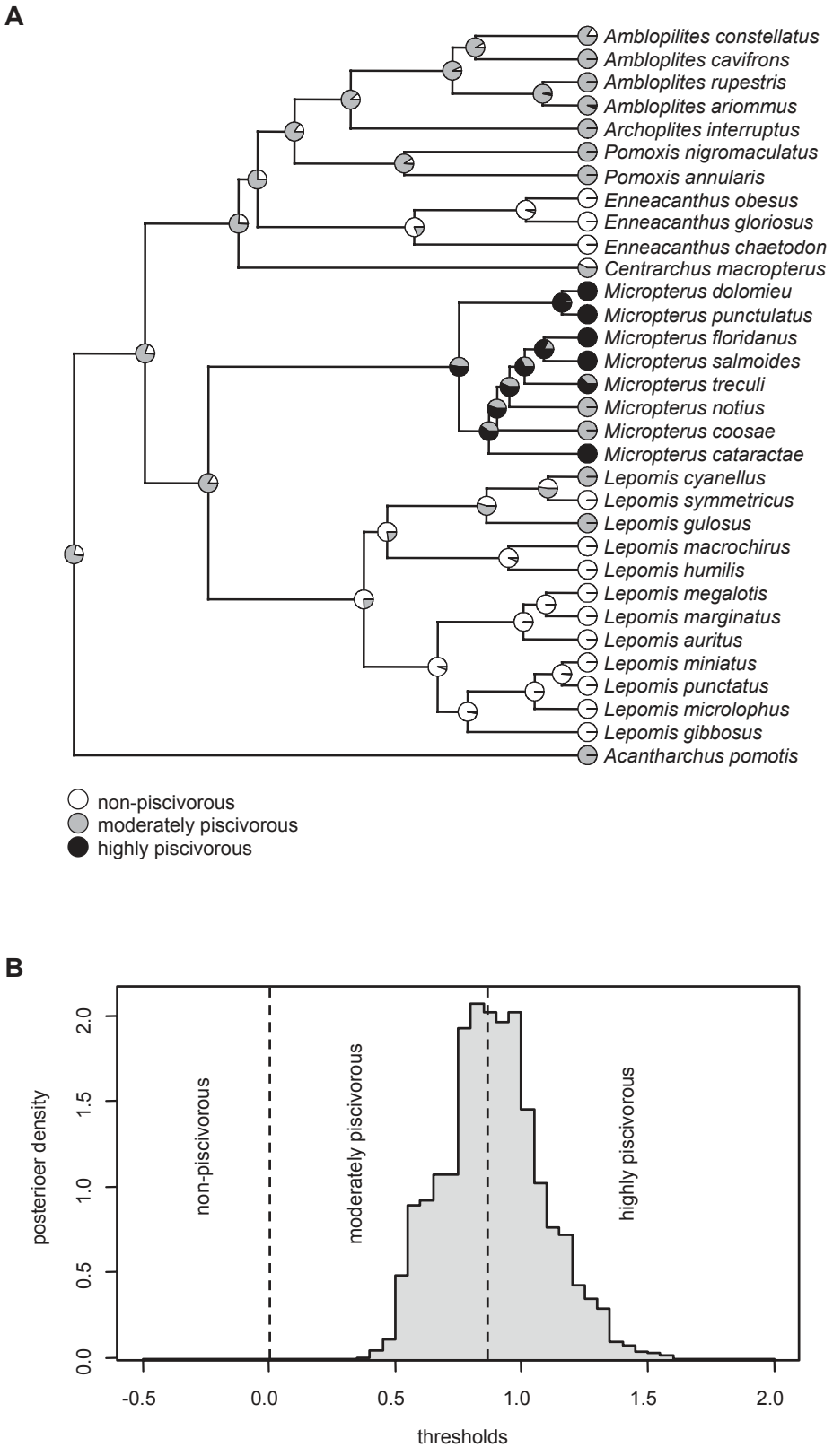

Figure 6. **(A) Posterior probabilities for the evolution of feeding mode as a threshold character in centrarchid fishes. (B) Posterior density for the position of the threshold between moderate and high piscivory.**

change again or revert to a prior state. Both of these assumptions are features of the threshold model (Felsenstein 2012).

In the present article (and implementation of this method), I allow for the possibility that trait values for terminal species are uncertain. This is accomplished by first specifying (prior) probabilities that each tip species is in each of the observed states on the tree. In the empirical component of this study, for instance, I have (in some cases of uncertain feeding mode) assigned equal prior probability that a species belonged in two different feeding mode classes. For example, I assigned equal prior probability that the species *Ambloplites ariommus* (the shadow bass) was either moderately or highly piscivorous. That is because diet data in this species consist of samples from only eight individuals in one study (Viosca 1936). Although that study showed that fish comprise a component of the diet of *A. ariommus*, the data are too few to assign the species conclusively to one or the other piscivory classes. Although not the case here, if we were completely ignorant about feeding mode for one or more species we could assign equal prior probabilities for all three feeding modes. Because uncertainty is included via the prior, we also get back posterior probabilities that our tip species are found in each state. For the centrarchid dataset, these posterior probabilities are shown at the tips of the tree in Figure 6A.

Allowing uncertainty in the states for tip species should not be confused with allowing for the possibility of within-species polymorphism for the discrete character state. As discussed by Felsenstein (2012) this is theoretically possible, but not equivalent to what has been done here. To model within-species polymorphism one would need to add an additional parameter—the within-species variability in liability (relative to the among-species Brownian rate of species mean liability evolution, here fixed at 1.0). This is because if the rate of species mean liability is high relative to the variance of liability within species, within-species polymorphism should be extremely transitory on an evolutionary time scale. Conversely, if the variance in liability is large relative to the rate of liability evolution, then polymorphism could persist over considerable evolutionary time.

In addition, within-species polymorphism contains information about the position of the species mean liability relative to the threshold. For instance (assuming symmetry of the distribution of liabilities within species) a 50:50 within-species polymorphism, known exactly, indicates that the species mean liability sits directly on a threshold. This is wholly different from a 0.50:0.50 prior on the discrete character, which merely indicates ignorance; in other words, the species phenotype is poorly known and thus we should not reject a posterior sample in which a species-mean liability was sampled from either side of the threshold.

For some character traits the ordering of the discrete character along the liability axis is obvious. For instance, in my empirical example from centrarchid fishes it seems quite likely (given the myriad of behavioral and morphological characteristics associated with feeding mode) that the ordering *not* $\leftrightarrow$ *moderately* $\leftrightarrow$ *highly* for the trait "piscivorous" is the only plausible ordering of these three character traits. Readers should keep in mind that an assumed ordering is not the same as assuming a temporal evolutionary sequence from the base of the tree forward. In fact, nothing about an assumed ordering gives us any information about, for instance, the state at the root—which must be estimated from the data and tree.

In the case of characters in which we are confident that a natural ordering exists—but for which that ordering is unknown or multiple orderings are plausible—I propose an information theoretic approach. Here, we can compute the DIC from the posterior sample. I have shown using simulation that this method shows some promise in identifying the true character order on the liability axis; however, if we have a good idea of the ordering a priori, this would probably be better. Furthermore, the number of possible orderings for a character with m states increases as $m!/2$. This means that although for three states there are three orderings, and for four states there are only 12 orderings; with as few as six states there are 360 orderings, and for 10 states there are over 1.8 million unique orderings. Given the computationally intensive nature of the Bayesian MCMC method used in this article, it will probably not be viable to consider all possible orderings for more than a few different states of a discretely valued threshold character. Finally, for many instances in which the ordering is unknown, the character may not be ordered or the ordering may change in different parts of the phylogeny. Because there is no way to compare ordered to unordered trait evolution using the threshold model (only alternative possible orderings), I would recommend caution in applying the method to datasets of this type.

Finally, in both Felsenstein (2012) and the present article we have focused on Brownian motion as a model for evolution of the liabilities on tree. Although this is a widely used model for quantitative trait evolution on phylogenies, one attribute of the model that makes it slightly unrealistic for many circumstances is that, given enough time, variance among species increases infinitely without bounds. That means that most lineages that cross a threshold have a very narrow window of time during which there is significant probability they will cross back, after which that probability falls dramatically. An alternative continuous trait evolution model with growing popularity in phylogenetic comparative biology is the OU model (Felsenstein 1988; Hansen 1997; Butler and King 2004). In this model, liability would evolve as a random walk with tendency to revert toward a central location. The OU process is most often used by comparative biologists to model stabilizing selection (Hansen 1997). Under the strictest interpretation of the Wright's (1934) threshold model, liabilities are invisible and thus cannot be under selection directly; however, OU may nonetheless be a good model for the evolution of

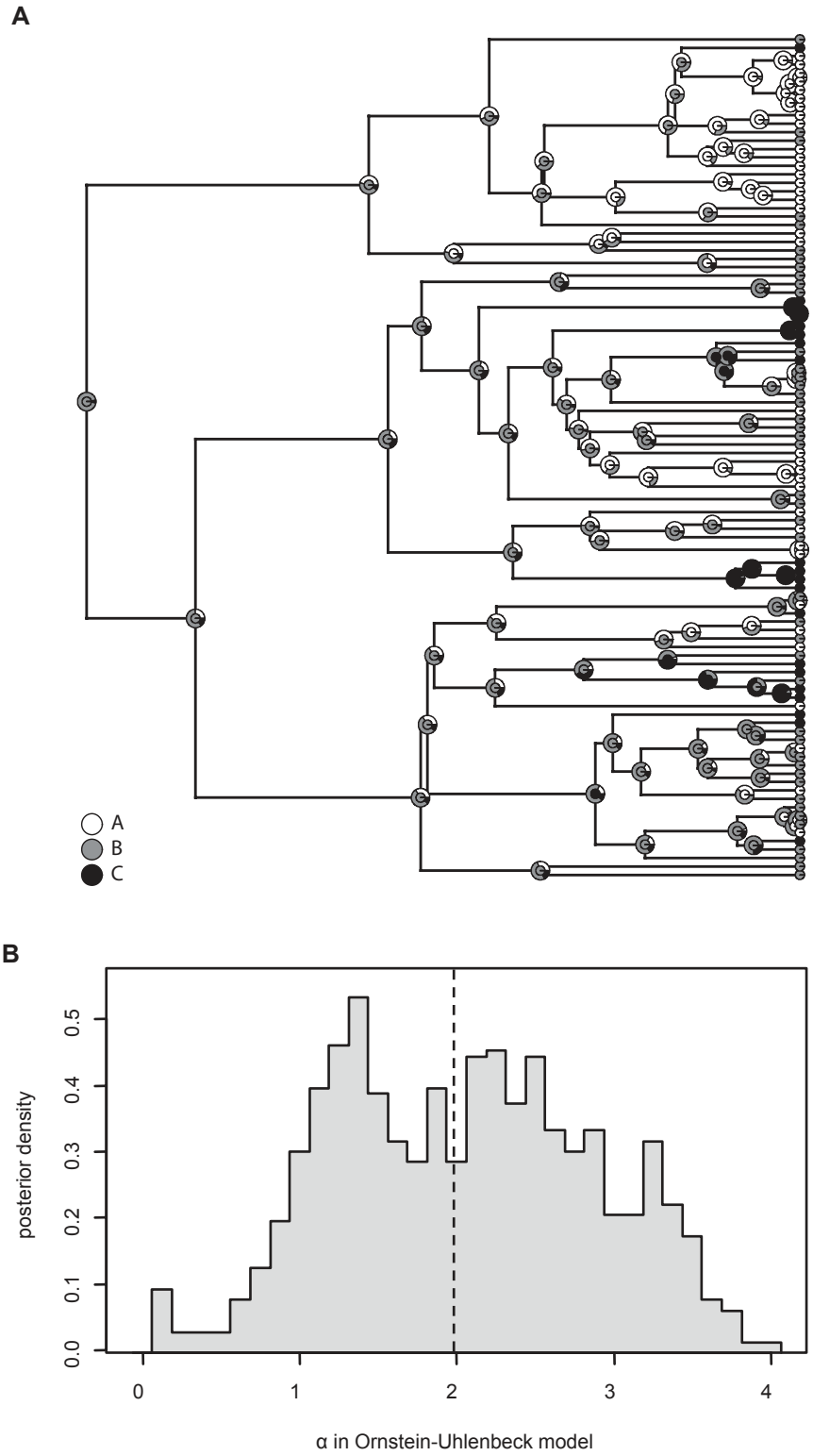

Figure 7. **(A) Posterior probabilities for evolution of a three-state simulated threshold character in which the underlying model for evolution of the liability is an Ornstein–Uhlenbeck model with $\alpha = 2.0$. The total tree length was set to 1.0. The center of each pie chart gives the true state from simulation. (B) Posterior density for α from the OU model. Although the posterior density is broad, it is very nearly centered exactly (mean = 1.98; vertical dashed line) on the generating value of 2.0. The HPD interval for α is (0.65, 3.51).**

liability on the tree if liability is under indirect selection (through correlation with unmeasured continuous traits), or if liability has natural boundaries in the values that it can assume. The phytools implementation of the method of this article now includes the OU model as an option. Figure 7A shows simulated and reconstructed ancestral states for data generated under the threshold model with OU as the underlying evolutionary process for the evolution of liability on the tree. In this simulation the total tree length is 1.0, thresholds were set to [0, 0.5] with character sequence A↔B↔C, and the simulated value of α was set to 2.0. Figure 7B shows the posterior density of α from 1,000,000 generations of Bayesian MCMC, with the first 200,000 generations excluded as burn-in. The posterior density is centered very closely on the generating value of α (posterior mean = 1.98), and, although the posterior density is quite broad, the 95% HPD interval (0.65, 3.51) does not include zero. Although this result is preliminary, it certainly suggests that it may be viable to use OU or other stochastic processes (in addition to Brownian motion) to model the evolution of liability on the branches of the tree.

On a whole I believe that this new approach for ancestral state estimation shows considerable promise. Intuitively, the threshold model seems much more realistic for discretely coded, complex, morphological, behavioral, or ecological traits than, say, the "nucleotide model"—in which the data are modeled using a discrete-state Markov process (the same process generally assumed for the evolution of DNA sequences). Furthermore, for data simulated under the threshold model, accuracy of this method was quite high. I hope to explore additional details of this method—for instance, improving the efficiency of the MCMC sampler, testing multiple models for the evolution of liability, and incorporating intraspecific polymorphism—with my future research.

ACKNOWLEDGMENTS

Thanks to D. Collar for providing data on feeding mode for Centrarchidae and for some helpful discussion of this method, and thanks to T. Near, an anonymous reviewer, and especially B. O'Meara for constructive criticism that helped to significantly improve this article over earlier versions.

DATA ARCHIVING

The doi for my data is 10.5061/dryad.4t157.

LITERATURE CITED

Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19:716–723.

Beaulieu, J. M., B. C. O'Meara, and M. J. Donoghue. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in the campanulid angiosperms. Syst. Biol. 62:725–737.

Bokma, F. 2008. Detection of "punctuated equilibrium" by Bayesian estimation of speciation and extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a molecular phylogeny. Evolution 62:2718–2726.

Brooks, D. R., and D. A. McLennan. 1991. Phylogeny, ecology, and behavior: a research program in comparative biology. University of Chicago Press, Chicago, IL.

Bürger, R., G. P. Wagner, and F. Stettinger. 1989. How much heritable variation can be maintained in finite populations by mutation-selection balance? Evolution 43:1748–1766.

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

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

Cavalli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. Am. J. Hum. Genet. 19:233–257.

Collar, D. C., B. C. O'Meara, P. C. Wainwright, and T. J. Near. 2009. Piscivory limits diversification of feeding morphology in centrarchid fishes. Evolution 63:1557–1573.

Cunningham, C. W., K. E. Omland, and T. H. Oakley. 1998. Reconstructing ancestral character states: a critical reappraisal. Trends Ecol. Evol. 13:361–366.

Eastman, J. M., M. E. Alfaro, P. Joyce, A. L. Hipp, and L. J. Harmon. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. Evolution 65:3578–3589.

Ekman, S., H. L. Andersen, and M. Wedin. 2008. The limitations of ancestral state reconstruction and the evolution of the ascus in the Lecanorales (lichenized Ascomycota). Syst. Biol. 57:141–156.

Felsenstein, J. 1973. Maximum likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. 25:471–492.

———. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

———. 1988. Phylogenies and quantitative characters. Annu. Rev. Ecol. Syst. 19:445–471.

———. 2004. Inferring phylogenies. Sinauer, Sunderland, MA.

———. 2005. Using the quantitative genetic threshold model for inferences between and within species. Philos. Trans. R. Soc. Lond. B 360:1427–1434.

———. 2012. A comparative method for both discrete and continuous characters using the threshold model. Am. Nat. 179:145–156.

Fitzjohn, R. G. 2010. Quantitative traits and diversification. Syst. Biol. 59:619–633.

Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *In* J. M. Bernado, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds. Bayesian Statistics 4. Clarendon Press, Oxford, UK.

Goldberg, E. E., and B. Igić. 2008. On phylogenetic tests of irreversible evolution. Evolution 62:2727–2741.

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

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

Huelsenbeck, J. P., and J. P. Bollback. 2001. Empirical and hierarchical Bayesian estimation of ancestral states. Syst. Biol. 50:351–366.

Huelsenbeck, J. P., R. Nielsen, and J. P. Bollback. 2003. Stochastic mapping of morphological characters. Syst. Biol. 52:131–158.

Ives, A. R., and T. Garland Jr. 2010. Phylogenetic logistic regression for binary dependent variables. Syst. Biol. 59:9–26.

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pp. 21–132 *in* M. N. Murno, ed. Mammalian protein metabolism. Vol. III. Academic Press, New York, NY.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925.

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

———. 2009. Lizards in an evolutionary tree: ecology and adaptive radiation of anoles. University of California Press, Berkeley, CA.

Losos, J. B., K. I. Warheit, and T. W. Schoener. 1997. Adaptive differentiation following experimental island colonization in *Anolis* lizards. Nature 387:70–73.

Lynch, M., and B. Walsh. 1998. Genetics and the analysis of quantitative traits. Sinauer, Sunderland, MA.

Maddison, W. P., P. E. Midford, and S. P. Otto. 2007. Estimating a binary character's effect on speciation and extinction. Syst. Biol. 56:701–710.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller. 1953. Equations of state calculations by fast computing machines. J. Chem. Phys. 21:1087–1092.

Near, T. J., D. I. Bolnick, and P. C. Wainwright. 2005. Fossil calibrations and molecular divergence time estimates in the centrarchid fishes (Teleostei: Centrarchidae). Evolution 59:1768–1782.

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

Nunn, C. L. 2011. The comparative approach in evolutionary anthropology and biology. University of Chicago Press, Chicago, IL.

O'Meara, B. C., C. Ané, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922–933.

Omland, K. E. 1999. The assumptions and challenges of ancestral state reconstructions. Syst. Biol. 48:604–611.

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.

Pagel, M. 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Syst. Biol. 48:612–622.

Pagel, M., A. Meade, and D. Barker. 2004. Bayesian estimation of ancestral character states on phylogenies. Syst. Biol. 53:673–684.

Paradis, E. 2012. Analysis of phylogenetics and evolution with R. 2nd ed. Springer, New York, NY.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: convergence diagnosis and output analysis for MCMC. R News 6:7–11.

R Core Development Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rabosky, D. L. 2006. Likelihood methods for detecting temporal shifts in diversification rates. Evolution 60:1152–1164.

Revell, L. J. 2008. On the analysis of evolutionary change along single branches in a phylogeny. Am. Nat. 172:140–147.

———. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223.

Revell, L. J., and D. C. Collar. 2009. Phylogenetic analysis of the evolutionary correlation using likelihood. Evolution 63:1090–1100.

Revell, L. J. and L. J. Harmon. 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. Evol. Ecol. Res. 10:311–321.

Revell, L. J., D. L. Mahler, P. R. Peres-Neto, and B. D. Redelings. 2012. A new method for identifying exceptional phenotypic diversification. Evolution 66:135–146.

Rice, W. R., and E. E. Hostert. 1993. Laboratory experiments on speciation: what have we learned in 40 years? Evolution 47:1637–1653.

Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. Evolution 55:2143–2160.

Ryan, M. J., and A. S. Rand. 1999. Phylogenetic influence on mating call preference in female túngara frogs, *Physalaemus pustulosus*. Anim. Behav. 57:945–956.

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

Skinner, A., and M. S. Y. Lee. 2010. Plausibility of inferred ancestral phenotypes and the evaluation of alternative models of limb evolution in scincid lizards. Biol. Lett. 6:354–358.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. B 64:583–639.

Stadler, T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proc. Natl. Acad. Sci. USA 108:6187–6192.

Viosca, P. 1936. A new rock bass from Louisiana and Mississippi. Copeia 1:37–45.

Wright, S. 1934. An analysis of variability in the number of digits in an inbred strain of guinea pigs. Genetics 19:506–536.

Yang, Z. 2006. Computational molecular evolution. Oxford Univ. Press: Oxford, UK.

Associate Editor: T. Near