

MODELING STABILIZING SELECTION: EXPANDING THE ORNSTEIN–UHLENBECK MODEL OF ADAPTIVE EVOLUTION

Jeremy M. Beaulieu,^{1,2} Dwueng-Chwuan Jhwueng,^{3,4} Carl Boettiger,⁵ and Brian C. O’Meara⁶

¹Department of Ecology and Evolutionary Biology, Yale University, P.O. Box 208106, New Haven, Connecticut 06520–8106

²E-mail: jeremy.beaulieu@yale.edu

³National Institute for Mathematical and Biological Synthesis, 1534 White Ave, University of Tennessee, Knoxville, Tennessee, 37996–1527

⁴Department of Statistics, Feng-Chia University, Taichung, Taiwan 40724, R.O.C.

⁵Center for Population Biology, University of California, Davis, 1 Shields Avenue, Davis, California, 95616

⁶Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee, 37996–1610

Received September 2, 2011

Accepted February 6, 2012

Comparative methods used to study patterns of evolutionary change in a continuous trait on a phylogeny range from Brownian motion processes to models where the trait is assumed to evolve according to an Ornstein–Uhlenbeck (OU) process. Although these models have proved useful in a variety of contexts, they still do not cover all the scenarios biologists want to examine. For models based on the OU process, model complexity is restricted in current implementations by assuming that the rate of stochastic motion and the strength of selection do not vary among selective regimes. Here, we expand the OU model of adaptive evolution to include models that variously relax the assumption of a constant rate and strength of selection. In its most general form, the methods described here can assign each selective regime a separate trait optimum, a rate of stochastic motion parameter, and a parameter for the strength of selection. We use simulations to show that our models can detect meaningful differences in the evolutionary process, especially with larger sample sizes. We also illustrate our method using an empirical example of genome size evolution within a large flowering plant clade.

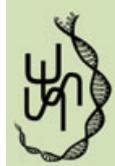
KEY WORDS: Brownian motion, comparative method, continuous characters Hansen model, Ornstein–Uhlenbeck.

Single-rate Brownian motion works reasonably well as a model for evolution of traits. It models drift, drift-mutation balance, and even stabilizing selection toward a moving optimum (Hansen and Martins 1996). However, a single parameter model can certainly not explain the evolution of traits across all life. There have been extensions to the model, such as a single Ornstein–Uhlenbeck (OU) process that has a constant pull toward an optimum value, a multiple mean OU process with different possible means for different groups (Hansen 1997; Butler and King 2004), and multiple rate Brownian motion processes allowing different rates of evolution on different branches (O’Meara et al. 2006; Thomas et al. 2006). These models, while useful, still do not cover all the scenarios biologists want to examine. For example, existing models with a value toward which species are being pulled all have a fixed

strength of pull over the entire history of the group. It is possible to allow the rate of stochastic motion to vary, or the value of the attractor to vary, but not for both to vary. Such restrictions on model complexity may make sense when phylogenies are limited to a few dozen taxa. However, in an era where phylogenies can have over 55,000 taxa (Smith et al. 2011), we may be so bold as to attempt to fit models that vary both rates and means of the evolutionary process. This article develops and implements such models.

Generalizing the Hansen Model

Hansen (1997) described a model where quantitative characters are assumed to evolve according to an OU process. The Hansen model, as it has become known, expresses the amount of change



in a quantitative trait along each branch in a phylogeny and is given by the stochastic differential equation:

$$dX_i(t) = \alpha [\beta_i(t) - X_i(t)] dt + \sigma dB_i(t). \tag{1}$$

Equation (1) describes the amount of change in a quantitative trait (X_i) during an increment of time, t , when it is assumed that along each branch there is an optimum trait value ($\beta_i(t)$) that identifies a selective regime acting on a lineage over the course of its history. The evolution of the trait toward this optimum is governed by a constant, α , describing the strength of selection. The terms $dB_i(t)$ are random variables denoting the increments of a Brownian motion process and are assumed to be normally distributed with an expectation of zero and a variance equal to $\sigma^2 dt$. Thus, σ^2 is a constant describing the rate of stochastic evolution away from the optimum.

Butler and King (2004) implemented the Hansen model in a likelihood framework. Rather than assuming an optimum for every branch in the tree, Butler and King (2004) assumed that only a small number of distinct selective regimes have operated on a quantitative trait, with each being defined by a single optimum θ_k . An assumption of current implementations of the Hansen model is that both α and σ^2 are constants, and do not vary among selective regimes. Here, we broaden the implementation of Butler and King (2004) to not only allow α and σ^2 to vary across selective regimes, but also to allow smaller parts of branches to be assigned different models. For consistency, we rely on the same terminology and notation as in Butler and King (2004). For example, the times at which changes in selective regimes or speciation events take place are referred to as “epochs.” The i th lineage is divided into $\kappa(i)$ epochs, and thus $t_{i,\gamma}$ can refer to the beginning of an epoch assigned model γ . The term $\beta_{i,\gamma}$ is the optimum for lineage i with model γ . Note we rely on the same notation as Butler and King (2004), but using subscripts instead of superscripts (e.g., $t_{i,\gamma} = t_i^\gamma$).

MULTIPLE VARIANCE PARAMETERS

We begin by broadening the Hansen model to allow the stochastic motion parameter, σ^2 , to vary across selective regimes. Because Brownian motion is nondirectional, the expected values of this model will remain the same as in Hansen (1997) and Butler and King (2004), although here we allow a given edge to be subdivided into more units for a greater number of optima per edge. Thus, following Butler and King equations (A2) and (A3) we start with the moments of equation (1):

$$E[X_i(T) | X_i(0) = \theta_0] = \theta_0 e^{-\alpha T} + \int_0^T \alpha e^{-\alpha t} \beta_i(T - t) dt \tag{2}$$

$$\begin{aligned} V_{ij} &= Cov[X_i(T), X_j(T) | X_i(0)] \\ &= X_j(0) = \theta_0 = \int_0^T \sigma^2 e^{-2\alpha t} \rho_{ij}(T - t) dt. \end{aligned} \tag{3}$$

The correlation ρ_{ij} is based on the assumption that taxa evolve independently after divergence ($\rho_{ij} = 0$) and are the same before diverging ($\rho_{ij} = 1$; see Appendix). Butler and King (2004) convert the continuous trait optimum $\beta_i(t)$ in equation (2) to a series of piecewise-constant selection regimes,

$$E[X_i(T) | X_i(0) = \theta_0] = \theta_0 e^{-\alpha T} + \sum_{\gamma=1}^{\kappa(i)} e^{-\alpha T} (e^{\alpha t_{i,\gamma}} - e^{\alpha t_{i,\gamma-1}}) \beta_{i,\gamma}, \tag{4}$$

where $\kappa(i)$ is the index of the last epoch of lineage i . To allow σ^2 to vary across selective regimes, we only have to modify equation (3) to allow σ^2 to be a function of time. In the Appendix, we derive the following formula for calculating the V_{ij} th element of \mathbf{V} describing the covariance between species i and j :

$$V_{ij} = \frac{e^{-2\alpha T}}{2\alpha} \sum_{\gamma=1}^{\kappa(i,j)} (e^{2\alpha s_{ij,\gamma}} - e^{2\alpha s_{ij,\gamma-1}}) \sigma_\gamma^2. \tag{5}$$

For $t > 0$, the path between the root and the most recent common ancestor (mrca) of the i th and j th lineage is divided into $\kappa(i,j)$ epochs, $[0, s_{ij,1}]$, $[s_{ij,1}, s_{ij,2}]$, ..., $[s_{ij,\kappa(i,j)-1}, s_{ij,\kappa(i,j)}]$, each assigned model γ . When α approaches zero, the covariance among species converges to the covariance obtained from a model that assumes one Brownian motion rate parameter per regime (i.e., $V_{ij} = \sigma_\gamma^2 s_{ij}$). Similarly, when $\sigma_\gamma^2 = \sigma_{\gamma-1}^2$, the elements of V_{ij} reduces to $V_{ij} = \sigma^2 s_{ij}$.

MULTIPLE ATTRACTION PARAMETERS

Allowing α to vary among selective regimes alters the way the expected values are calculated. The Hansen model assumes that the expectation is a weighted average of θ , where more recent regimes tend to have a greater effect on expected values than more ancient regimes (although this is influenced by the amount of time spent in each). This effect is stronger with greater values of α . With multiple attraction parameters, the calculation of expected values will be even more complex. For example, if a lineage spends the first half of its time in one regime with a high α , and the rest of the time in a regime with a very weak α , the expected trait value may be closer to the mean from the first regime than that of the second.

Similarly, varying α will complicate the calculation of the covariance among species. Under the Hansen model, the covariance equals the variance of the common ancestor multiplied by an exponential decay with separation time (Hansen 1997). However, when α varies the rate of decay will become even more dependent on when any two species shared a particular selective regime. For example, if a lineage spends the first half of its time in one regime

with a high α , and the rest of the time prior to speciation in a regime with a very weak α , the covariance between species i and j will decay much slower than if the more recent regime had the higher value of α .

For both α and σ^2 to vary, we can solve equation (1) so that these variables are now a function of time (see Appendix):

$$E[X_i(T)|X_i(0) = \theta_0] = \theta_0 e^{-\int_0^T \alpha_i(t) dt} + e^{-\int_0^T \alpha_i(t) dt} \left(\int_0^T \alpha_i(t) \beta_i(t) e^{\int_0^t \alpha_i(x) dx} dt \right) \quad (6)$$

$$V_{ij} = Cov[X_i(T), X_j(T)|X_i(0) = X_j(0) = \theta_0] = E \left[\left(e^{-\int_0^T \alpha_i(t) + \alpha_j(t) dt} \right) \left(\int_0^T \sigma_i(t) e^{\int_0^t \alpha_i(x) dx} dB_i(t) \right) \times \left(\int_0^T \sigma_j(t) e^{\int_0^t \alpha_j(x) dx} dB_j(t) \right) \right] \quad (7)$$

Similar to equation (2) described above, we can convert the continuous trait optimum $\beta_i(s)$ in equation (6) as well as $\alpha_i(s)$ into a series of piecewise-constant selection regimes,

$$E[X_i(T)||X_i(0) = \theta_0] = \theta_0 e^{-\sum_{\gamma=1}^{\kappa(i)} \alpha_{i,\gamma}(s_{i,\gamma} - s_{i,\gamma-1})} + \left(e^{-\sum_{\gamma=1}^{\kappa(i)} \alpha_{i,\gamma}(s_{i,\gamma} - s_{i,\gamma-1})} \right) \sum_{\gamma=1}^{\kappa(i)} \beta_{i,\gamma} (e^{\alpha_{i,\gamma} s_{i,\gamma}} - e^{\alpha_{i,\gamma} s_{i,\gamma-1}}) \quad (8)$$

For $t > 0$, the i th lineage is divided into $\kappa(i)$ epochs, $[0, s_{i,1}]$, $[s_{i,1}, s_{i,2}]$, \dots , $[s_{i,\kappa(i)-1}, s_{i,\kappa(i)}]$. However, as written, the model assumes one optimum, $\beta_{i,\gamma}$, and one attraction parameter, $\alpha_{i,\gamma}$, for each branch in a phylogeny. To reduce the number of parameters in the model, Butler and King (2004) assumed that only a small number (r) of distinct selective regimes have operated on any given phylogeny, and reduced the number of $\beta_{i,\gamma}$ by substituting in the θ_k ($k = 1, \dots, r$) corresponding to the selective regime operating on each branch. Under this assumption, each branch optimum $\beta_{i,\gamma}$ only depends on the values of $\theta_1, \theta_2, \dots, \theta_r$:

$$\beta_{i,\gamma} = \sum_{k=1}^r \beta_{ik,\gamma} \theta_k \quad (9)$$

The indicator variable, $\beta_{ik,\gamma}$, reflects the mapping of these selective regimes on the phylogeny. When $\beta_{ik,\gamma} = 1$, $\beta_{i,\gamma}$ is the optimum on lineage i with model γ , and model γ has optimum θ_k , such that for any i , $\beta_{i,\gamma} = \theta_\gamma$. We follow these same assumptions for reducing the number of $\alpha_{i,\gamma}$ in the model. A single $\tilde{\alpha}_k$, $k = 1, \dots, r$ defines the strength of selection operating on each selection regime, and we replace each $\alpha_{i,\gamma}$ with $\tilde{\alpha}_k$ corresponding to the selective regime operating on that branch. Thus, the strength of selection, $\alpha_{i,\gamma}$, operating on each branch

depends on $\alpha_1, \alpha_2, \dots, \alpha_r$:

$$\alpha_{i,\gamma} = \sum_{k=1}^r \beta_{ik,\gamma} \tilde{\alpha}_k \quad (10)$$

The expected mean trait values at the end of each evolutionary lineage are calculated as a weighted sum of each optima and the ancestral state. Taking into account equations (9) and (10), we can express equation (2) using matrix notation as:

$$E[\mathbf{X}(T)|\mathbf{X}(0) = \theta_0] = \mathbf{W}\theta, \quad (11)$$

where the vector $\theta = (\theta_0, \theta_1, \dots, \theta_r)'$ and \mathbf{W} is the matrix of weights with entries

$$W_{i0} = e^{-\sum_{\gamma=1}^{\kappa(i)} \left(\sum_{k=1}^r \beta_{ik,\gamma} \tilde{\alpha}_k \right) (s_{i,\gamma} - s_{i,\gamma-1})},$$

$$W_{ik} = \left(e^{-\sum_{\gamma=1}^{\kappa(i)} \left(\sum_{k=1}^r \beta_{ik,\gamma} \tilde{\alpha}_k \right) (s_{i,\gamma} - s_{i,\gamma-1})} \right) \times \sum_{\gamma=1}^{\kappa(i)} \beta_{ik,\gamma} \left(e^{\left(\sum_{k=1}^r \beta_{ik,\gamma} \tilde{\alpha}_k \right) s_{i,\gamma}} - e^{\left(\sum_{k=1}^r \beta_{ik,\gamma} \tilde{\alpha}_k \right) s_{i,\gamma-1}} \right), \quad (12)$$

for $i = 1, \dots, N$ and $k = 1, \dots, r$. Because \mathbf{W} reflects the expected weights for each species, each row entry in \mathbf{W} is divided by the sum of its row to ensure that the weights for each species sum to 1.

Finally, we allow both α and σ^2 to be a function of time when computing the covariance of i and j and define V_{ij} th element of \mathbf{V} as:

$$V_{ij} = e^{-\left(\sum_{\gamma=1}^{\kappa(i)} \alpha_{i,\gamma}(s_{i,\gamma} - s_{i,\gamma-1}) + \sum_{\gamma=1}^{\kappa(j)} \alpha_{j,\gamma}(s_{j,\gamma} - s_{j,\gamma-1}) \right)} \times \left(\sum_{\gamma=1}^{\kappa(i,j)} \sigma_\gamma^2 \frac{e^{2\alpha_\gamma s_{ij,\gamma}} - e^{2\alpha_\gamma s_{ij,\gamma-1}}}{2\alpha_\gamma} \right) \quad (13)$$

We note that equation (13) reduces to V_{ij} in the simple case of $\alpha_k = \alpha_{k'}$ and $\sigma_\gamma^2 = \sigma_{\gamma-1}^2$ (see Appendix).

GENERALIZED LEAST-SQUARES ESTIMATION

Estimates of the vector θ can be solved using the generalized least-squares (GLS) estimator

$$\hat{\theta} = (\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1}\mathbf{W}'\mathbf{V}^{-1}\mathbf{x}, \quad (14)$$

where \mathbf{x} is the vector of species values, \mathbf{W} is the matrix of weights, and \mathbf{V} is the scaled variance–covariance matrix computed for each species pair. The GLS estimates of $\hat{\theta}$ are conditional on the maximum likelihood estimates of the vector $\alpha = (\hat{\alpha}_1, \dots, \hat{\alpha}_r)$ and the vector $\sigma = (\hat{\sigma}_1^2, \dots, \hat{\sigma}_r^2)$ with α and σ entering into \mathbf{V} through equation (13) and α entering into \mathbf{W} through equation (12). The log likelihood of the vectors α , σ , and θ given the data is

evaluated by the function,

$$\log(L) = \log \left[\frac{1}{\sqrt{(2\pi)^N \det(\mathbf{V})}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mathbf{W}\theta)' (\mathbf{V}^{-1}) (\mathbf{x} - \mathbf{W}\theta) \right) \right] \quad (15)$$

which is solved using a nonlinear optimization routine. This likelihood is used in this article when comparing models using the Akaike Information Criterion (AIC; Akaike 1974), but it could also be used in likelihood ratio tests or even a Bayesian context.

Current implementations of the Hansen model employ a parametric bootstrap procedure to estimate the confidence intervals surrounding each parameter, which can be computationally intensive. An alternative to this approach is to calculate the confidence intervals directly. Here, we first calculate the approximate standard errors associated with the estimated values of the vectors α , σ , and θ by computing two separate variance–covariance matrices. The first is the estimated variance–covariance matrix of $\hat{\theta}$ and is computed as $\mathbf{S}_{\hat{\theta}} = (\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1}$. The square roots of the diagonals of this matrix are the standard errors of $\hat{\theta}$. The second is the variance–covariance matrix of the estimated values of α and σ , which is computed as the inverse of the Hessian matrix; the approximate standard errors for α and σ are the square roots of the diagonals of this matrix. The Hessian is a matrix of second-order derivatives obtained by evaluating how changes in parameter values influence the maximum of the log-likelihood function. If changes in the value of a parameter results in sharp changes in the slope around the maximum of the log-likelihood function, the second-order derivative will be large, the standard error will be small, and the parameter estimate is considered stable. On the other hand, if the second-order derivative is nearly zero, then the change in the slope around the maximum is also nearly zero, indicating the parameter value can be moved in any direction without greatly affecting the log-likelihood value. In such situations, the standard error of the parameter will be large. The approximate upper and lower 95% confidence interval for all parameters can then be computed by multiplying each approximate standard error by the critical value in the *t*-distribution where the cumulative probability is equal to 0.975 (i.e., $t(0.975, \infty) = 1.96$).

Examples

SIMULATIONS

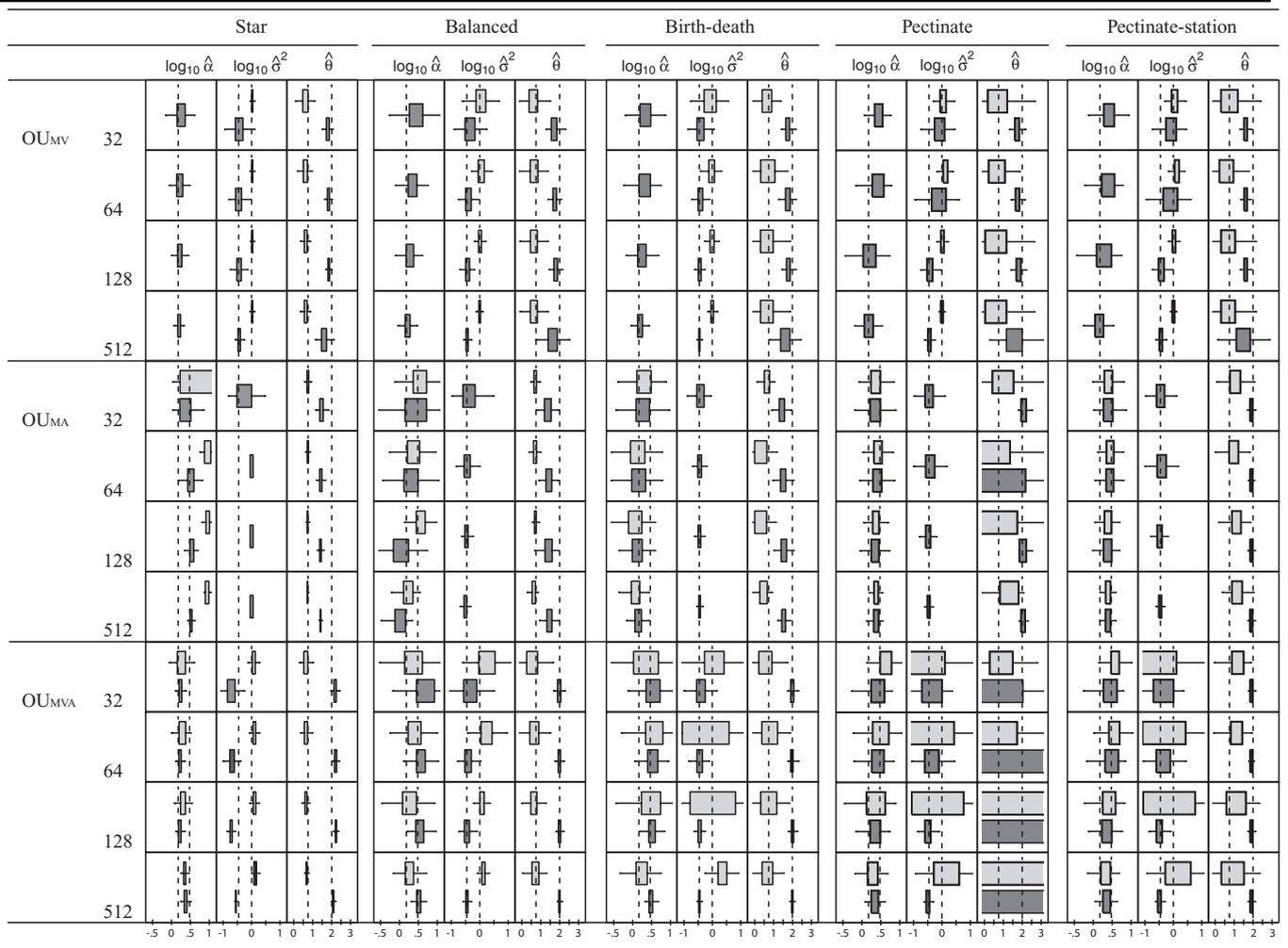
We evaluated the performance of our method by applying it to datasets created by simulating multiple-mean OU models that variously relax the assumptions of a constant α and σ^2 . The first model allowed θ and σ^2 to vary among regimes, while keeping α constant. In other words, the simulation tests an OU model with different state means and multiple variance parameters, and we refer to this model hereafter as the OU_{MV} model. The second allowed θ and α to vary among selective regimes, while keeping

σ^2 constant (referred to hereafter as the OU_{MA} model). The third model allowed θ , α , and σ^2 to vary among the selective regimes (referred to hereafter as the OU_{MVA} model).

To account for biases resulting from tree shape, datasets were simulated on a star tree (unresolved), completely balanced tree, a pectinate tree (a comb), and a random tree generated under the birth–death process (birth = 0.4, death = 0.2). For each tree shape, we also varied taxon sampling by generating trees comprised of 32, 64, 128, and 512 taxa and we scaled the root to tip length to be one in all trees. In the star trees, we divided the number of species equally between two selective regimes. For all other tree shapes, we assumed that all lineages began in the same selective regime and a transition to a second selective regime occurred only once and at some point along a branch leading to a subset of species. For the balanced tree, we assumed the transition occurred along the branch leading to one of the two subclades arising from the initial divergence. For the pectinate tree, we assumed the transition occurred along the branch leading to a subclade that contained exactly half the total diversity in the tree. Finally, in the birth–death tree, we randomly assigned the transition to occur along a branch leading to a subclade that contained roughly a quarter of the total diversity contained in the tree.

Historically, approaches such as this have been used to estimate parameter values or to compare models (e.g., Butler and King 2004; Davis et al. 2007; Harmon et al. 2008, 2010; Pinto et al. 2008; Collar et al. 2009; Smith and Beaulieu 2009; Beaulieu et al. 2010; Edwards and Smith 2010). Although we favor analyses under the former approach (see Discussion), we investigated the performance of the models under both approaches. We tested the fit, bias, and precision of each model under different tree shapes and taxon sampling strategies by estimating the distance between the approximating model and the true model from which the data were generated. Data were simulated under each of the three OU models described above. Each dataset was evaluated under the generating model as well as a simple OU model that assumed a single optimum for all species (referred to hereafter as the OU1 model), and an OU model that assumed different state means and a single α and σ^2 acting on all selective regimes (referred to hereafter as OU_M model, and which is equivalent to the model of Butler and King [2004]). For each combination of model, tree shape, and taxon sampling, 1000 datasets were simulated with known parameters for α , θ , and σ^2 and we assumed an OU process with a trend, where the respective means for each selective regime ($\theta_1 = 2.0$; $\theta_2 = 0.75$) were set to be larger than the starting value ($\theta_0 = 0.25$). We also assumed that a shift from Regime 1 to Regime 2 resulted in either the relaxation of selection ($\alpha_1 = 3.0$; $\alpha_2 = 1.5$), an increase in the rate of stochastic motion ($\sigma_1^2 = 0.35$; $\sigma_2^2 = 1.0$), or both, depending on the generating model. The AIC weight (w_i), which represents the relative likelihood that model *i* is the best model given a set of models (Burnham and Anderson

Table 1. Boxplots depicting the bias and precision of the estimates of α , σ^2 , and θ . Data were simulated on a star tree (unresolved), a completely balanced tree, a completely pectinate tree, and a random tree generated under the birth–death process. For each tree, all lineages were assumed to have begun in the same selective regime (dark gray) and transitioned to a second selective regime (light gray) only once, which always resulted in either the relaxation of selection ($\alpha_1 = 3.0$; $\alpha_2 = 1.5$; OU_{MA}), an increase in the rate of stochastic motion ($\sigma_1^2 = 0.35$; $\sigma_2^2 = 1.0$; OU_{MV}), or both (OU_{MVA}), depending on the generating model. In all cases, the respective means for each selective regime ($\theta_1 = 2.0$; $\theta_2 = 0.75$) were set to be larger than the starting value ($\theta_0 = 0.25$). When estimating parameters for datasets simulated under both OU_{MA} and OU_{MVA} on the pectinate trees (Pectinate), the models almost returned values of θ that were nonsensical. However, when θ_0 was assumed to be distributed according to a stationary distribution and dropped from the model (Pectinate-station), this no longer occurred in any of the simulated datasets and θ_1 and θ_2 can be reasonably estimated from the data. All simulations were carried out in the new R package *OUwie*. Note that some of the plots under the pectinate tree are truncated by the need to have a consistent but useful scale across the majority of the simulations.



2002), was calculated for all the models. All simulations were carried out in the new R package *OUwie*.

Table 1 provides information on the bias and precision of the estimated parameter values. With the exception of the pectinate trees, there is a general bias for OU_{MV} , OU_{MA} , and OU_{MVA} to underestimate θ , in this case, returning values that are too low. The bias cannot be attributed to any particular θ , as θ_1 , and θ_2 all exhibited a pattern of generally being underestimated (Table 1). In contrast, both α and σ^2 were consistently overestimated for

all three models, regardless of whether they were allowed to vary between the two selective regimes (Table 1). However, there was a general trend in all three models for the biases in these parameters to decrease as sample size increased (Table 1).

The influence of sample size on the bias and precision of α was particularly pronounced, which tended to have a negative effect on the bias and precision of σ^2 in the OU_{MVA} model (Table 1). Under both OU_{MA} and the OU_{MVA} model, the 32 taxa tree in the birth–death tree set had estimated values of α that were about

twice, on average, the true values (Table 1). The bias dropped as sample size increased, but even with a tree of 512 taxa the bias was never less than 1.10 times the true value. However, when simulating datasets under the OU_M model (results not shown), the bias with a 512 taxa tree was larger and exceeded 1.25 times the true value. A key question for many empiricists using these methods is not “what is the value of α in each regime” but “is the value of α in Regime 1 much bigger than the value in Regime 2.” Looking at cases where the model returned an α for Regime 1 that was at least 10% higher than the α returned for Regime 2, roughly 19% of the datasets simulated on the 512 taxa tree were below this threshold (with only 12% of all datasets inferring a higher α for Regime 2). However, for the 128 taxa tree, nearly 50% of the simulated datasets returned an α for Regime 1 that was less than 10% higher than α for Regime 2. Taken together, these results suggest that even for moderately sized datasets (e.g., ~150 taxa) it may be difficult to confidently infer meaningful differences in α among selective regimes. The model parameters for these datasets will tend to return larger values of α with greater uncertainty.

In the pectinate trees, when simulating datasets under both OU_{MA} and OU_{MVA} , the models returned values of θ that were nonsensical. Interestingly, when we reran the simulations across the pectinate trees with θ_0 dropped from the model, this behavior no longer occurred in any of the simulated datasets (Table 1). Dropping θ_0 from the model assumes that the starting value is distributed according to the stationary distribution of the OU process, where the conditional distribution of $X(T)$ given $X(0)$ in equation (8) are assumed to have converged on the same distribution. This would not fit a biological scenario involving a trend away from an ancestral state, but it does fit a scenario of a stationary evolutionary process: the same distribution of optima, selection, and rates occur in the past as in the present. Under this assumption, the general bias was for OU_{MA} and OU_{MVA} to overestimate θ associated with Regime 2 (Table 1). This suggests caution when comparing values of θ among selective regimes when the tree is highly imbalanced—they will tend to be unreliable or, at least, very imprecise and associated with very large standard errors. In such instances, it may be helpful to drop θ_0 from the model entirely.

We note that under some circumstances it may be impossible to estimate θ_0 accurately. Under the OU process the influence of the starting state, θ_0 , on the expected values exponentially decreases at a rate that is determined by α . For larger values of α , a trait is generally free from the constraints of the starting state and will be much closer to the optimal value for the selective regime it is currently in. As a result, the weighting factor in W representing the influence of the root will decrease and estimates of θ_0 should approach zero. Thus, for large values of α (i.e., $\alpha > 2$), the estimates of θ_0 will always be small, and the standard error surrounding θ_0 is likely to be positively misleading (i.e., false precision).

From a model comparison perspective, the results from comparisons of model fit for the birth–death trees are shown in Figure 1 and the results for the balanced and pectinate trees can be found in Figures S1, S2. Overall, we found that regardless of tree shape there was a general trend for the mean AIC weight for the generating model (i.e., OU_{MV} , OU_{MA} , or OU_{MVA}) to increase with increasing sample size (Fig. 1). Under both the OU_{MV} and OU_{MVA} models, the 32, 64, and 128 tip trees had datasets that favored either the OU1 or the OU_M models. However, the OU1 model was not favored very often (<2% in all cases), and no tree size had a large proportion of datasets that favored the OU_M model. When the number of taxa was increased to 512, there were never any simulated datasets that found support for either OU1 or the OU_M model.

Interestingly, when the generating model was OU_{MA} , there were no tree sizes that had a greater proportion of datasets that correctly favored the OU_{MA} model, except in the case of the balanced trees. When simulating datasets on the balanced trees, both the 128 and 512 tip trees had a mean AIC weight that was >0.50 , and in both cases the number of datasets that correctly favored the OU_{MA} model was nearly 50%. However, it is worth pointing out that the clades that comprise Regime 2 in the balanced trees contained more species and tended to span more time than those of the birth–death trees. This suggests that when shifts occur more recently among selective regimes with unequal sample sizes, greater uncertainty in model choice is introduced and it is difficult to correctly favor the OU_{MA} model even when it is the better model.

THE EVOLUTION OF PLANT GENOME SIZE

Flowering plants exhibit a growth form dependent distribution in genome size, with “woody” species (i.e., trees/shrubs) characterized by small genome sizes with lower overall variance compared to herbaceous species (Ohri 2005; Beaulieu et al. 2008). When viewed in the context of microevolutionary processes, such as selection and drift, the growth form dependent distribution of genome size could be explained, in part, by consequences associated with life history. Woody angiosperms take many years to reach reproductive maturity (Verdú 2002), which has been hypothesized to be linked to their generally slower rates of molecular evolution across several loci (Gaut et al. 1992, 1996, 1997; Smith and Donoghue 2008). This implies that the pattern may also manifest at the level of the whole genome, with the longer generation times of woody angiosperms providing fewer opportunities for random insertion/deletions to occur per unit time. Indeed, woody lineages have accumulated changes in genome size at rates that are an order of magnitude slower than related herbaceous lineages (Beaulieu et al. 2010). On the other hand, angiosperm trees are reported to have large effective population sizes (Petit and Hampe 2006), which would make selection more efficient at removing

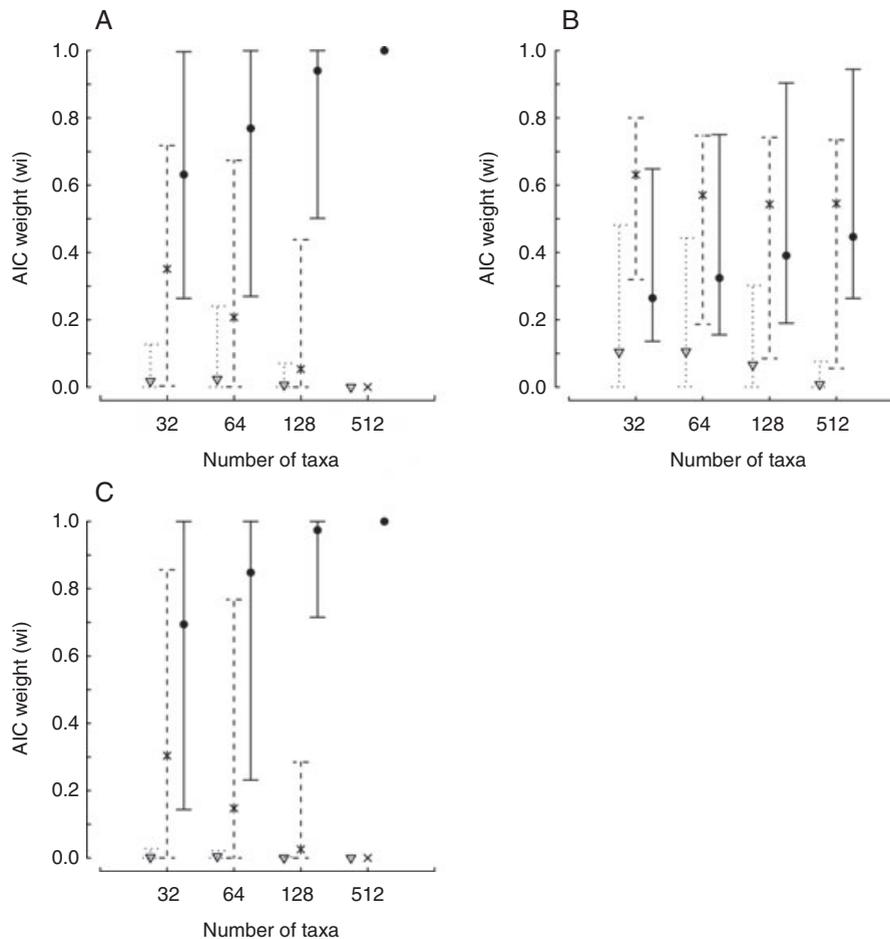


Figure 1. Comparisons of model fit based on AIC weights (w_i) for birth–death trees comprised of 32, 64, 128, and 512 taxa. Data were simulated under each of the three OU models that allowed the rate of stochastic motion (σ^2) to vary (OU_{MV}, A), or the strength of selection (α) to vary (OU_{MA}, B), for both to vary (OU_{MVA}, C). When their fit (•) was compared against a simple OU model that assumed a single optimum for all species (OU1, ▽) and an OU model that assumed different state means and a single α and σ^2 acting all selective regimes (OU_M, ◐), there was a general trend for the mean AIC weight to increase with increasing sample size.

deleterious mutations and excess DNA (Lynch 2007). Stronger selection in woody species would also be consistent with the suggestion that large increases in DNA content negatively affect woody growth and physiology (Stebbins 1938; Beaulieu et al. 2008). Here, we illustrate our method by testing for an asymmetry in the strength of selection and stochastic motion acting on genome size due to growth form.

We focus our analysis on the Monocotyledonae (monocots; Cantino et al. 2007), a major branch of the angiosperms. The monocots are a large clade of mainly herbaceous angiosperms that also contain a few clades of predominately woody species, including the palms (Arecales; APG III 2009). However, it is worth noting that unlike other woody angiosperms, “woody” monocots do not produce true “wood” (i.e., secondary xylem). Here, we generally define the tree/shrub or “woody” category as simply large plants with long generation times (e.g., Smith and Donoghue

2008). Genome size estimates were taken from the Plant DNA C-values database (release 5.0; Bennett and Leitch 2010) for monocot species where both the 1C amount (the amount of DNA in the unreplicated gametic nucleus) and the ploidy level were known. Genome size was \log_{10} -transformed prior to all analyses to ensure the data minimally conformed to Brownian motion evolution (O’Meara et al. 2006; Oliver et al. 2007).

The procedure used to construct a large sequence based phylogeny follow the methods described by Smith et al. (2009) and implemented in the program PHLAWD. We specified *atpB*, *matK*, *ndhF*, *rbcL*, and *trnL-F* as our genes of interest and we limited our search of GenBank to only return sequences for taxa represented in our genome size dataset. Our final combined sequence matrix contained 590 species and 7213 aligned sites. We conducted 100 maximum likelihood analyses on this matrix using the standard RAxML search algorithm under the GTR+CAT

approximation of rate heterogeneity partitioned for each gene (Stamatakis 2006). The final 100 topologies were rooted with Acorales (sensu Chase et al. 2006) and scored under GTR+ Γ to estimate molecular branch lengths and to identify the tree with highest likelihood score. We assigned minimum age constraints based on 14 described fossils with unequivocal affinities to clades nested within the monocots (see Supporting Information). The maximum-likelihood (ML) tree was then converted to ultrametric using the semiparametric penalized likelihood method developed by Sanderson (2003) and implemented in r8s. The cross-validation procedure was used to find the optimal value of the smoothing parameter (λ). To accommodate our large tree, we randomly pruned our ML tree to 60 taxa 10 times, and the consistent best estimate of λ was used to smooth the rates across the entire ML tree. The final tree is available at TreeBASE (<http://www.treebase.org>), accession number 12409.

Slow growing, tall, and “woody” genera have been described in several different monocot families, including Areaceae, Asparagaceae (e.g., *Dasyllirion*, *Dracaena*, *Nolina*, *Xanthorrhoea*), Bromeliaceae (e.g., *Puya*), Dasypogonaceae (*Dasypogon*, *Kingia*), Pandanaceae (*Pandanus*), Strelitziaceae (e.g., *Ravenala*), and the woody bamboo genera in the tribe Bambuseae of Poaceae (e.g., *Phyllostachys*, *Semiarundinaria*, *Sasa*). However, due to the absence of genome size and/or sequence data for many of these genera the effect of growth form analyses were restricted to comparisons between (1) *Dasypogon* (Dasypogonaceae; 1 species) and the “woody” palms (Areaceae; 36 species), and (2) the remaining species that were scored as herbaceous (553 species). We reconstructed the likeliest growth form at all internal nodes of our dated ML tree using a likelihood-based ancestral state reconstruction model (Pagel 1999) that assumed equal transitions rates between character states. Given these internal estimates, the simplest explanation is a single transition from herbaceous to the woody habit reconstructed along the branch leading to the split between *Dasypogon* and the Arecales (Fig. 2A).

We fit six different models of genome size evolution. The two simplest models were a single-rate Brownian motion model (BM1) and an OU model with one optimum for all species of monocots (OU1). We fit a multiple rate Brownian motion model that assigned a separate rate for each character state (BM_{RC}). In this case, a separate rate was assigned to woody and herbaceous lineages. The OU_M model assumed two optima for each growth form while keeping both α and σ^2 constant. Finally, we assessed the fit of the OU_{MV}, OU_{MA}, and OU_{MVA} models that assumed two optima but varied α and σ^2 between woody and herbaceous lineages. In all cases, we dropped θ_0 from the model and assumed that the starting value was distributed according to the stationary distribution of the OU process (see above). All tests were carried out in *OUwie* using the “noncensored” approach of O’Meara

Table 2. The fit of alternative models of genome size evolution in monocots. The best model, based on Δ AIC and Akaike weights, was the OU_{MVA}, which estimated a separate θ , α , and σ^2 for woody and herbaceous monocot lineages.

Model	-lnL	AIC	Δ AIC	w_i
BM1	-227.6	459.1	156.2	<0.01
BMS	-203.0	412.0	109.1	<0.01
OU1	-160.0	326.1	23.2	<0.01
OU _M	-159.2	326.5	23.6	<0.01
OU _{MV}	-147.3	304.6	1.7	0.290
OU _{MA}	-159.2	328.5	25.6	<0.01
OU _{MVA}	-145.3	302.9	0.0	0.678

Table 3. Parameter estimates and their associated 95% confidence interval (CI) for the OU_{MVA} model, the model that best fit the genome size data. Each CI was obtained by multiplying each approximate standard error by the critical value in the t-distribution where the cumulative probability is equal to 0.975 (i.e., $t(0.975, \infty) = 1.96$).

	Herb Estimate	95% CI	Woody Estimate	95% CI
α	3.85	± 0.955	<0.001	$\pm <0.01$
σ^2	2.51	± 0.376	0.531	± 0.281
θ	0.618	± 0.143	<0.001	$\pm \infty$

et al. (2006) that assumes that the placement of a state change along an internal branch is known.

Results, which are based on parameters estimated on the maximum likelihood tree, are summarized in Tables 2 and 3. The best model, based on Δ AIC and Akaike weights (Table 2), was one indicating that both the strength of selection and the rate of stochastic motion in genome size differ between the two selective regimes. The optimum value for woody plants was less (the back-transformed mean = 1.0 pg; Fig. 2B) than the optimum inferred for herbaceous lineages (the back-transformed mean = 4.15 pg; Fig. 2B). However, the optimum value for woody lineages was not identifiable (Fig. 2B), which may be explained by very weak selection ($\alpha_w < 0.001$). As α approaches 0, a selective regime will behave more consistently with a Brownian motion process and the likelihood profile for θ will be flat, making it difficult to estimate this parameter (Butler and King 2004). Changes in genome size in woody lineages are, therefore, driven more by stochastic changes than by adaptive movement toward an optimal value. However, woody lineages had a rate of stochastic motion that was nearly five times slower than the rate of herbaceous lineages (Table 3), suggesting that these stochastic changes do not happen very often. Finally, we note that the width of the 95% confidence interval calculated from the approximate standard errors for all parameters was strongly correlated with the width of confidence

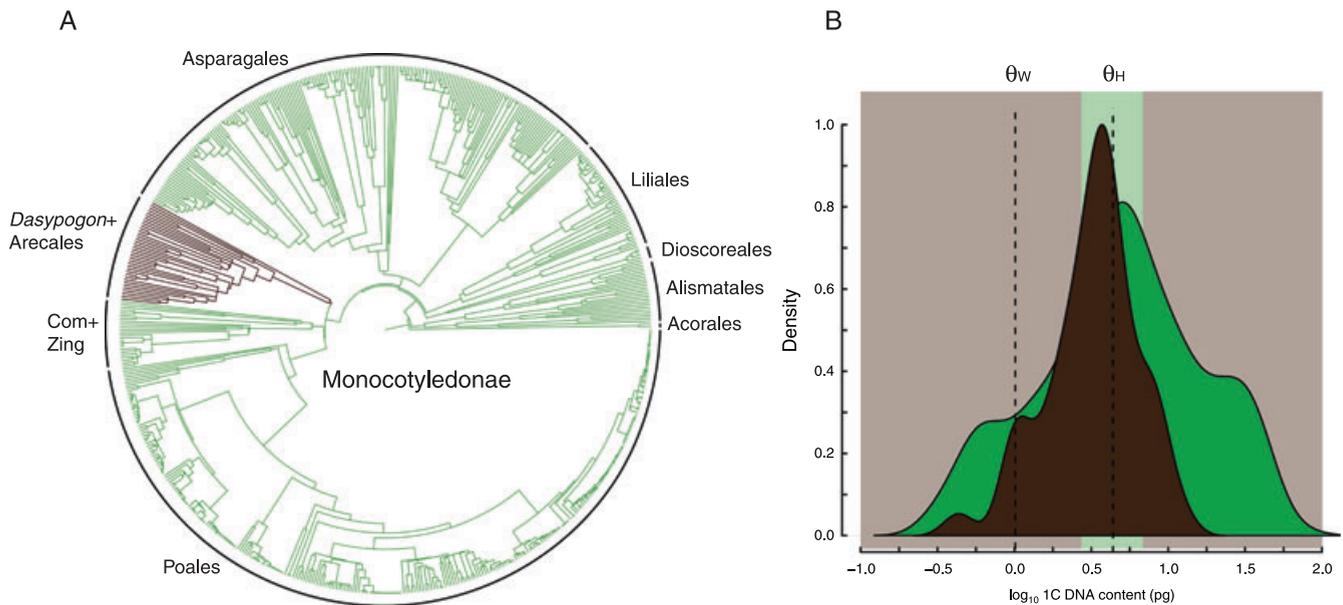


Figure 2. (A) Time-calibrated phylogeny of Monocotyledonae (monocots). The phylogeny is taken from a maximum likelihood analysis of 590 species based on combined analysis *atpB*, *matK*, *ndhF*, *rbcL*, and *trnL-F*. The major clades of monocots are labeled, and estimates of the likeliest growth form state (woody = brown; herbaceous = green) across all branches in the tree. Com + Zing represents the combined clade of Commelinales and Zingiberales. (B) The distributions of 1C DNA content among growth form, with the optimum value for woody plants (θ_W) estimated to be larger than the optimum inferred for herbaceous lineages (θ_H). However, the optimum value for woody lineages was not identifiable, which may be explained by very weak selection ($\alpha < 0.001$) operating within this regime.

intervals estimated from 100 replicates of a parametric bootstrap ($r = 0.984$, $P = 0.016$). However, while coverage probability (the proportion of times a true value lies within a specified confidence interval) is unknown for the particular parameter values in our data, simulations where we specify the true value resulted in a coverage probability of approximately 75%, rather than the ideal 95%, for both parametric bootstrap and approximate standard errors. This suggests that while approximate standard errors are an efficient replacement for parametric bootstrapping, both underestimate somewhat the true uncertainty in parameter values.

Confidence intervals are usually estimated for individual parameters. However, it is possible for the confidence intervals for pairs of parameters to be much wider than the univariate confidence intervals if the parameters are correlated with one another. For example, for OU models a decrease in σ^2 has a similar effect as an increase in α : less variation at the tips. There are differences, so the parameters are identifiable (greater α tends to erase phylogenetic signal, whereas lower σ^2 does not). For practical problems, it is possible to have a ridge of nearly equal likelihood where if one changes just one parameter, one moves off the ridge, but where changing both allows one to slide along the ridge. Univariate confidence intervals would not capture this effect. We investigated this by creating contour maps of the likelihood surface for pairs of parameters in our best-fitting model, letting the other parameters vary to find their maximum likelihood estimates given the values of our focal parameters. Results are shown in Figure 3. For

this empirical dataset, parameters appear distinguishable. However, this is an issue that users of this method should consider: even if univariate confidence intervals suggest precise estimates, it is worth considering the effect of varying pairs or potentially correlated parameters.

Taken together, these results are consistent with the view that life history alone can impose constraints to the evolution of genome size. Although genome size in monocots is generally under selective constraints, there are additional constraints that likely reflect the influence of generation time. The longer generation times that characterize the “woody” palms (Gaut et al. 1992) have provided fewer opportunities for changes in genome size to occur per unit of time.

Discussion

The Hansen model assumes that the strength of selection and the rate of stochastic motion (e.g., genetic drift, environmental fluctuations, random mutations, etc.) do not vary among selective regimes. This has placed much of the emphasis on identifying and quantifying trait optima and not enough on investigating other characteristics of the evolutionary process, such as rate or strength of attraction to these optima. The methods developed here relax these assumptions, by allowing the strength of selection and stochastic motion parameters to vary among selective regimes. In its most complex form, our modification of the

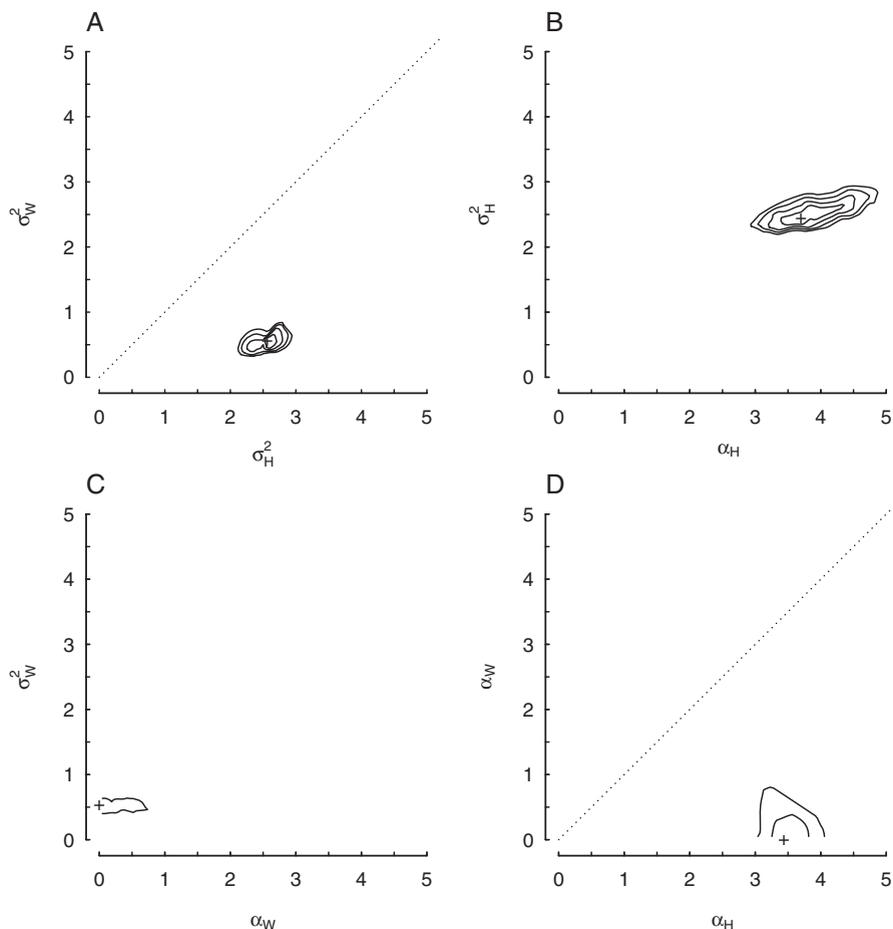


Figure 3. Contour plots of the likelihood surface. Each surface is constrained such that the parameters on the axes are fixed but the other parameters are free to find their MLE given a pair of fixed parameter values. A “+” indicates the global maximum likelihood estimate for the parameter values. Contour lines are drawn at points 0.5, 1.0, 1.5, and 2 InL units away from the optimum. For plots of the same kind of parameter, a dotted 1:1 line is also shown. Results indicate that the likelihood surfaces are peaked enough that parameter values are readily distinguishable in this case.

Hansen model can assign each branch a trait optimum, a rate of stochastic motion parameter, and a parameter for the strength of selection.

In general, however, our method will be most appropriate when investigating simpler models. As with existing approaches (Butler and King 2004; O’Meara et al. 2006), our initial implementation requires a priori assignment of model regimes to branches. However, there is no intrinsic need for this given the model. One could adopt this to an MCMC approach, like that of Revell et al. (2011) and Eastman et al. (2011), to estimate the best division of the tree into different model regimes. As with Huelsenbeck et al. (2004) and Pagel and Meade (2006), a reversible jump MCMC approach could also be used to get the posterior probabilities of various models, ranging from BM1 to OU_{MVA} , in a single analysis.

A set of models, like those developed here, can be used in several ways to address biological questions. The approach we typically advocate is to select one or more models using AIC, re-

turn maximum likelihood parameter estimates from these models (using model averaging if appropriate), and then focus discussion on the parameter estimates and their biological significance. For example, we found that “woody” and herbaceous monocot species appear to be moving toward different optimum values at different rates (Table 3), and these values and rates are different enough to be biologically relevant. Parameter estimates could be inferred using a Bayesian approach instead, which would include posterior probabilities for the values at the cost of using prior information. However, these models can be used in other ways as well. Some researchers focus on model comparison, asking whether model X is better than model Y. By “better” we mean a model that loses less information about reality under AIC (Burnham and Anderson 2002), a model better supported by the data under Bayes Factors (Kass and Raftery 1995), a model with greater posterior probability using reversible jump MCMC (Green 1995), or the simplest model that is not rejected in favor of a more complex one under likelihood ratio tests.

As biologists, we find the exercise of being primarily concerned about which model is best rather than about what the parameter estimates are under the best model(s), less compelling. Biological reality is undoubtedly more complex, especially across many species, than any single evolutionary model we may examine. Fit of a model, especially when compared with a more complex model, is based on both how well the model parameters match the generating processes in the real world but also how much data are available to use in fitting. A model for human height where the dependent variable of offspring height is related to maternal height is probably a better fit to the data than a model that relates day of the week of birth for good biological reasons. A model incorporating maternal height, maternal nutrition, offspring nutrition, history of offspring disease, and the full genomic sequence of the offspring will undoubtedly be a better fit than either model, but only once there are enough height data. We would rather investigate, say, the magnitude of the heritable effect of maternal height on offspring height rather than whether offspring disease history is included or excluded from the best model. In any case, there are many ways to do good science, and we have endeavored in this article to evaluate how well these new methods will perform in whatever framework they are used in, whether parameter estimation or model comparison.

We have described the models here as reflecting drift, selection, and other evolutionary processes. However, we note a common misunderstanding with this sort of model: although the overall parameters resemble those of microevolutionary models (for example, for selection) they are actually describing the pattern of evolutionary change. As shown by Hansen and Martins (1996), even a simple model like single rate Brownian motion is consistent with neutral genetic drift, selection toward a moving optimum, drift-mutation balance, and other evolutionary processes. Thus, one can investigate whether parameter estimates are consistent with predictions (i.e., a hypothesis that after a mass extinction, the rate of evolution is increased due to diminished population sizes and more drift), but it is difficult to work the other way and use information about the best model or parameter estimates to draw conclusions about evolutionary processes.

In regards to parameter estimation, our simulations suggest that accurate parameter estimates will often require larger sample sizes, especially when the underlying hypothesis involves testing for differences in the strength of selection. The issue of sample size need not only apply to the total size of a given tree. Accurate parameter estimates may be difficult even for a large tree if it is divided up among many selective regimes each comprised of relatively few species and branches. That is not to say that our methods should not be applied to smaller datasets or selective regimes comprised of few species. We only advise that conclusions be made cautiously. As our simulations suggest, there will be circumstances in which meaningful patterns can be extracted

from smaller sample sizes. These will be instances where the signal in the data is particularly strong. However, one should bear in mind that parameter estimates (especially for α) are generally overestimated with smaller datasets (i.e., <128 species). In this regard, it will be prudent to interpret results in light of the standard error or confidence interval associated with each of the parameters. At times, it may be more meaningful to make qualitative statements about parameter differences.

It would appear then that the models described here will naturally compliment the very large, comprehensive trees that are currently being generated by mining data from DNA sequence repositories (e.g., McMahon and Sanderson 2006; Sanderson et al. 2008; Smith et al. 2009, 2011; Thomson and Shaffer 2010). Applying our methods to these large datasets hold the promise of greater insights into evolution across broader phylogenetic and temporal scales and can point to a number of patterns that could not previously have been recognized and quantified. The methods described here can begin to address important comparative questions and hypotheses underlying major morphological radiations across the tree of life. For example, is the immense morphological diversity observed within flowering plants consistent with higher rates of evolution due to their generally faster generation times (Stebbins 1981; Bond 1989)? Is the low morphological diversity observed within marsupials related to constraints posed by their mode of reproduction (Lillegraven 1975; Sears 2004)? Do biogeographic movements result in tracking ancestral optima or the evolution of new climatic tolerances (Donoghue 2008)? We believe that studies along these lines will open up new avenues of research.

PROGRAM NOTE

The methods and simulations described above are implemented in the R package *OUwie* (pronounced “au-wi”), available through CRAN. We have also made *OUwie* available in the Discovery Environment at <http://preview.iplantcollaborative.org>. The Discovery Environment is a web-based integrated platform for data exploration and scientific discovery <http://www.iplantcollaborative.org/discover/discovery-environment>.

As input *OUwie* requires a phylogeny with branch lengths and internal node labels denoting the selective regimes at ancestral nodes, and a trait file that contains information for each species regarding the current selective regime and the values of a quantitative trait. The user can specify a series of models ranging from those based on Brownian motion processes available in Brownie (O’Meara et al. 2006), to the OU models available in Butler and King’s (2004) program *ouch* and the ones described here. *OUwie* uses a bounded subplex routine for minimizing the likelihood function to find the optimal parameter estimates. The user is also provided a series of model diagnostics indicating whether the likelihood search returned stable and reliable parameter estimates.

ACKNOWLEDGMENTS

Thanks are due to Jason Shapiro, Alex Dornburg, Jeff Oliver, Beth Forrestel, Barb Banbury, and Matt Pennell for useful feedback on the method implementation and for helpful suggestions on the manuscript. Support for JMB and BCO has been provided by the iPTOL program within the NSF-funded iPlant Collaborative (<http://www.iplantcollaborative.org/>). Support for DCJ has been provided by the National Institute for Mathematical and Biological Synthesis, an Institute sponsored by the National Science Foundation, the U.S. Department of Homeland Security, and the U.S. Department of Agriculture through NSF Award #EF-0832858, with additional support from The University of Tennessee, Knoxville.

LITERATURE CITED

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19:716–723.
- Angiosperm Phylogeny Group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161:105–121.
- Beaulieu, J. M., I. J. Leitch, S. Patel, A. Pendharkar, and C. A. Knight. 2008. Genome size is a strong predictor of cell size and stomatal density in angiosperms. *New Phytol.* 179:975–986.
- Beaulieu, J. M., S. A. Smith, and I. J. Leitch. 2010. On the tempo of genome size evolution in angiosperms. *J. Bot.* doi:10.1155/2010/989152
- Bennett, M. D., and I. J. Leitch. 2010. Plant DNA C-values database (release 5.0, Dec. 2010). Available at <http://kew.org/genomesize/homepage>
- Bond, W. J. 1989. The tortoise and the hare: ecology of angiosperm dominance and gymnosperm persistence. *Biol. J. Linn. Soc.* 36:227–249.
- Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. New York, Springer-Verlag.
- Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.* 164:683–695.
- Cantino, P. D., J. A. Doyle, S. W. Graham, W. S. Judd, R. G. Olmstead, D. E. Soltis, P. S. Soltis, and M. J. Donoghue. 2007. Towards a phylogenetic nomenclature of Tracheophyta. *Taxon* 56:822–846.
- Chase, M. W., M. F. Fay, D. S. Devey, O. Maurin, N. Rønsted, T. J. Davies, Y. Pillon, G. Petersen, O. Seberg, M. N. Tamura, et al. 2006. Multigene analyses of monocot relationships: a summary. *Aliso* 22:63–75.
- 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.
- Davis, C. C., M. Latvis, D. L. Nickrent, K. J. Wurdack, and D. A. Baum. 2007. Floral gigantism in Rafflesiaceae. *Science* 315:1812.
- Donoghue, M. J. 2008. A phylogenetic perspective on the distribution of plant species. *Proc. Natl. Acad. Sci. USA* 105:11549–11555.
- Eastman, J. M., M. E. Alfaro, P. Joyce, A. L. Hipp, and L. J. Harmon. 2011. AUTEUR: a novel comparative method for modeling shifts in the rate of character evolution on trees. *Evolution* 65:3578–3589.
- Edwards, E. J., and S. A. Smith. 2010. Phylogenetic analyses reveal the shady history of C4 grasses. *Proc. Natl. Acad. Sci. USA* 6:2532–2537.
- Gaut, B. S., S. V. Muse, W. D. Clark, and M. T. Clegg. 1992. Relative rates of nucleotide substitution at the rbcL locus of monocotyledonous plants. *J. Mol. Evol.* 35:292–303.
- Gaut, B. S., B. R. Morton, B. C. McCraig, and M. T. Clegg. 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL. *Proc. Natl. Acad. Sci. USA* 93:10274–10279.
- Gaut, B. S., L. G. Clark, J. F. Wendel, and S. V. Muse. 1997. Comparisons of the molecular evolutionary process at rbcL and ndhF in the grass family (Poaceae). *Mol. Biol. Evol.* 14:769–777.
- Green, P. J. 1995. Reversible jump Markov chain monte carlo computation and Bayesian model determination. *Biometrika* 82:711–732.
- 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 process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* 50:1404–1417.
- Harmon, L. J., J. Melville, A. Larson, and J. B. Losos. 2008. The role of geography and ecological opportunity in the diversification of day geckos (Phelsuma). *Syst. Biol.* 57:562–573.
- Harmon, L. J., J. B. Losos, T. J. Davies, R. G. Gillespie, J. L. Gittleman, W. B. Jennings, K. H. Kozak, M. A. McPeck, F. Moreno-Roark, T. J. Near, et al. 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64:2385–2396.
- Huelsenbeck, J. P., B. Larget, and M. E. Alfaro. 2004. Bayesian phylogenetic model selection using reversible jump Markov Chain Monte Carlo. *Mol. Biol. Evol.* 21:1123–1133.
- Kass, R. E., and A. E. Raftery. 1995. Bayes factors. *J. Am. Stat. Assoc.* 90:773–795.
- Lillegraven, J. A. 1975. Biological considerations of the marsupial-placental dichotomy. *Evolution* 29:707–722.
- Lynch, M. 2007. The origins of genome architecture. Sinauer Assoc., Inc., Sunderland.
- McMahon, M. M., and M. J. Sanderson. 2006. Phylogenetic supermatrix analysis of GenBank sequences from 2228 papilionoid legumes. *Syst. Biol.* 55:818–836.
- Oliver, M. J., D. Petrov, D. D. Ackerly, P. Falkowski, and O. M. Schofield. 2007. The mode and tempo of genome size evolution in eukaryotes. *Genome Res.* 17:594–601.
- O'Meara, B. C., C. Ane, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution. *Evolution* 60:922–933.
- Ohri, D. 2005. Climate and growth form: the consequences for genome size in plants. *Plant Biol.* 7:449–458.
- Pagel, M. 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst. Biol.* 48:612–622.
- Pagel, M., and A. Meade. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov Chain Monte Carlo. *Am. Nat.* 167:808–825.
- Petit, R. J., and A. Hampe. 2006. Some evolutionary consequences of being a tree. *Annu. Rev. Ecol. Evol. Syst.* 37:187–214.
- Pinto, G., D. L. Mahler, L. J. Harmon, and J. B. Losos. 2008. Testing the island effect in adaptive radiation: rates and patterns of morphological diversification in Caribbean and mainland Anolis lizards. *Proc. R. Soc. B* 275:2749–2757.
- Revell, L. J., L. D. Mahler, P. R. Peres-Neto, and B. D. Redelings. 2011. A new phylogenetic method for identifying exceptional phenotypic diversification. *Evolution* 66:135–146.
- Sanderson, M. J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19:301–302.
- Sanderson, M. J., D. Boss, D. Chen, K. A. Cranston, and A. Wehe. 2008. The PhyLoTA browser: processing GenBank for molecular phylogenetics research. *Syst. Biol.* 57:335–346.
- Sears, K. E. 2004. Constraints on the morphological evolution of marsupial shoulder girdles. *Evolution* 58:2353–2370.

Smith, S. A., and J. M. Beaulieu. 2009. Life history influences rates of climatic niche evolution in flowering plants. *Proc. R. Soc. Lond. B* 276:4345–4352.

Smith, S. A., and M. J. Donoghue. 2008. Rates of molecular evolution are linked to life history in flowering plants. *Science* 322:86–89.

Smith, S. A., J. M. Beaulieu, and M. J. Donoghue. 2009. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. *BMC Evol. Biol.* 9:1–12.

Smith, S. A., J. M. Beaulieu, A. Stamatakis, and M. J. Donoghue. 2011. Understanding angiosperm diversification using small and large phylogenetic trees. *Am. J. Bot.* 98:1–12.

Stamatakis, A. 2006. Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.

Stebbins, G. L. 1938. Cytological characteristics associated with different growth habits in dicotyledons. *Am. J. Bot.* 25:189–198.

———. 1981. Why are there so many species of flowering plants? *Bioscience* 31:573–577.

Thomas, G. H., R. P. Freckleton, and T. Szekeley. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. *Proc. R. Soc. Lond. B* 273:1619–1624.

Thomson, R. C., and H. B. Shaffer. 2010. Sparse supermatrices for phylogenetic inference: taxonomy, alignment, rogue taxa, and the phylogeny of living turtles. *Syst. Biol.* 59:42–58.

Verdú, M. 2002. Age at maturity and diversification in woody angiosperms. *Evolution* 56:1352–1361.

Associate Editor: L. Kubatko

Appendix The Generalized Hansen Model

Let $X_i(t)$ be the trait at time t . We assume that $X_i(t)$ satisfies the solution of the stochastic differential equation

$$dX_i(t) = \alpha_i(t)(\beta_i(t) - X_i(t))dt + \sigma_i(t)dB(t),$$

where $dB(t)$ denotes the white noise process, $\alpha_i(t)$ measures the rate of adaption, $\beta_i(t)$ is the primary optimum of X_i , and $\sigma_i(t)$ measures size of stochastic perturbation at time t .

Multiplying both sides by the integral factor $e^{\int_0^t \alpha_i(s)ds}$ and applying the chain rule and the fundamental theorem of calculus yields the differential equation

$$dX_i(t)e^{\int_0^t \alpha_i(s)ds} = \alpha_i(t)\beta_i(t)e^{\int_0^t \alpha_i(s)ds}dt + \sigma_i(t)e^{\int_0^t \alpha_i(s)ds}dB_i(t).$$

Given the initial condition $X_i(0) = \theta_0$, the solution to this equation is

$$X_i(t) = \left(\int_0^t \alpha_i(s)\beta_i(s)e^{\int_0^s \alpha_i(x)dx}ds \right) e^{-\int_0^t \alpha_i(s)ds} + \left(\int_0^t \sigma_i(s)e^{\int_0^s \alpha_i(x)dx}dB(s) \right) e^{-\int_0^t \alpha_i(s)ds} + \theta_0 e^{-\int_0^t \alpha_i(s)ds}.$$

Because this defines a Gaussian process, the first moment of $X_i(t)$ is completely specified. In particular, we have

$$E(X_i(T)|X_i(0) = \theta_0) = \theta_0 e^{-\int_0^T \alpha_i(t)dt} + e^{-\int_0^T \alpha_i(t)dt} \left(\int_0^T \alpha_i(t)\beta_i(t)e^{\int_0^t \alpha_i(x)dx}dt \right),$$

and the covariation between species i and species j is

$$\begin{aligned} Cov[X_i(T), X_j(T)|X_i(0) = X_j(0) = \theta_0] &= E[X_i(T) \cdot X_j(T)|X_i(0) = X_j(0) = \theta_0] \\ &\quad - E[X_i(T)|X_i(0) = \theta_0]E[X_j(T)|X_j(0) = \theta_0]. \\ &= E\left((e^{-\int_0^T \alpha_i(t) + \alpha_j(t)dt}) \left(\int_0^T \sigma_i(t)e^{\int_0^t \alpha_i(x)dx}dB_i(t) \right) \right. \\ &\quad \left. \times \left(\int_0^T \sigma_j(t)e^{\int_0^t \alpha_j(x)dx}dB_j(t) \right) \right). \end{aligned}$$

COVARIANCE STRUCTURE

The differential equation defining the covariation between species i and j assumes that both α_i and σ_i are a function of time. However, this equation also includes the much simpler cases such as when neither is allowed to vary (OU_M), or when only σ_i is allowed to vary (OU_{MV}), or when both are allowed to vary (OU_{MVA}). We show how the covariance between species i and j can be computed under these special cases. In each case, for $t > 0$ the path between root and the most recent common ancestor (mrca) of the i th and j th (mrca(i, j)) lineage is divided into $\kappa(ij)$ epochs, $[0, s_{ij,1}]$, $[s_{ij,1}, s_{ij,2}]$, \dots , $[s_{ij,\kappa(ij)-1}, s_{ij,\kappa(ij)}]$, that represent a change in selective regime. The parameters are then assumed as constant values in the γ th epoch, that is, $\alpha_i(t) = \alpha_j(t) = \alpha_\gamma, \beta_i(t) = \beta_j(t) = \beta_\gamma$ and $\sigma_i(t) = \sigma_j(t) = \sigma_\gamma$ for $\gamma = 1, 2, \dots, \kappa(ij)$. We also apply the common assumption that taxa evolve independently after diverging at time $s_{ij,\kappa(ij)}$. The path between the taxa i and mrca(i, j) is divided into $\kappa(i) - \kappa(ij)$ epochs $[s_{ij,\kappa(ij)}, s_{ij,\kappa(ij)+1}]$, \dots , $[s_{ij,\kappa(i)-1}, s_{ij,\kappa(i)}]$ with the constant value of parameters in the γ th epoch, that is, $\alpha_i(t) = \alpha_{i,\gamma}, \beta_i(t) = \beta_{i,\gamma}$ and $\sigma_i(t) = \sigma_{i,\gamma}$, $\gamma = \kappa(ij) + 1, \dots, \kappa(i)$.

Because the generalized Hansen model includes several special cases, we derived the covariance matrix between two species as:

Special case 1, OU_M : $\alpha_i(s) = \alpha_j(s) = \alpha; \sigma_i(s) = \sigma_j(s) = \sigma$

$$\begin{aligned} Cov[X_i(T), X_j(T)|X_i(0) = X_j(0) = \theta_0] &= E \left(e^{-2\alpha T} \left(\int_0^T \sigma e^{\alpha t} dB_i(t) \right) \left(\int_0^T \sigma e^{\alpha t} dB_j(t) \right) \right). \end{aligned}$$

Following Butler and King (2004), $dB_i(s)$ and $dB_j(s)$ denote the increments of two standard Brownian motions with covariation,

$$Cov[dB_i(t), dB_j(t)] = \rho_{ij}(t)dt.$$

Given the assumption of $\rho_{ij}(t)$, the covariation between species i and j can be computed as

$$Cov[X_i(T), X_j(T)|X_i(0) = X_j(0) = \theta_0] = e^{-2\alpha T} \int_0^{s_{ij}} \sigma^2 e^{2\alpha t} dt,$$

which is equivalent to $\int_0^T \sigma^2 e^{-2\alpha t} \rho_{ij}(T-t) dt$ and the integral can be calculated as,

$$\begin{aligned} Cov[X_i(T), X_j(T)|X_i(0) = X_j(0) = \theta_0] \\ = \frac{\sigma^2}{2\alpha} e^{-2\alpha(T-s_{ij})} (1 - e^{-2\alpha s_{ij}}). \end{aligned} \tag{A1}$$

General case, OUMVA: $\alpha_{i,\gamma}(t) = \alpha_{i,\gamma}$, $\alpha_{j,\gamma}(t) = \alpha_{j,\gamma}$, $\sigma_{i,\gamma}(t) = \sigma_{i,\gamma}$, $\sigma_{j,\gamma}(t) = \sigma_{j,\gamma}$, and $s_{ij} \leq t \leq T$; $\alpha_{i,\gamma}(t) = \alpha_{j,\gamma}(t) = \alpha_\gamma$, $\sigma_{i,\gamma}(t) = \sigma_{j,\gamma}(t) = \sigma_\gamma$, and $0 \leq t \leq s_{ij}$; $\gamma = 1, 2, \dots, \kappa(ij), \dots, \kappa(i)$ or $\gamma = 1, 2, \dots, \kappa(ij), \dots, \kappa(j)$.

$$Cov[X_i(T), X_j(T)|X_i(0) = X_j(0) = \theta_0]$$

$$\begin{aligned} = E \left[\left(e^{-\int_0^T \alpha_i(t) + \alpha_j(t) dt} \right) \left(\int_0^T \sigma_i(t) e^{\int_0^t \alpha_i(x) dx} dB_i(t) \right) \right. \\ \left. \times \left(\int_0^T \sigma_j(t) e^{\int_0^t \alpha_j(x) dx} dB_j(t) \right) \right] \end{aligned} \tag{A2.1}$$

$$\begin{aligned} = E \left[\left(e^{-\left(\sum_{\gamma=1}^{\kappa(i)} \alpha_{i,\gamma}(s_{i,\gamma}-s_{i,\gamma-1}) + \sum_{\gamma=1}^{\kappa(j)} \alpha_{j,\gamma}(s_{j,\gamma}-s_{j,\gamma-1}) \right)} \right) \right. \\ \times \left(\sum_{\gamma=1}^{\kappa(i)} \int_{s_{i,\gamma-1}}^{s_{i,\gamma}} \sigma_{i,\gamma} e^{\int_0^t \alpha_{i,\gamma} dx} dB_{i,\gamma}(t) \right) \\ \left. \times \left(\sum_{\gamma=1}^{\kappa(j)} \int_{s_{j,\gamma-1}}^{s_{j,\gamma}} \sigma_{j,\gamma} e^{\int_0^t \alpha_{j,\gamma} dx} dB_{j,\gamma}(t) \right) \right] \end{aligned} \tag{A2.2}$$

$$\begin{aligned} = \left(e^{-\left(\sum_{\gamma=1}^{\kappa(i)} \alpha_{i,\gamma}(s_{i,\gamma}-s_{i,\gamma-1}) + \sum_{\gamma=1}^{\kappa(j)} \alpha_{j,\gamma}(s_{j,\gamma}-s_{j,\gamma-1}) \right)} \right) \\ \times E \left[\left(\sum_{\gamma=1}^{\kappa(ij)} \int_{s_{ij,\gamma-1}}^{s_{ij,\gamma}} \sigma_\gamma e^{\alpha_\gamma t} dB_{i,\gamma}(t) \right) \right. \\ \left. + \sum_{\gamma=\kappa(ij)+1}^{\kappa(i)} \int_{s_{i,\gamma-1}}^{s_{i,\gamma}} \sigma_{i,\gamma} e^{\alpha_{i,\gamma} t} dB_{i,\gamma}(t) \right) \cdot \\ \left(\sum_{\gamma=1}^{\kappa(ij)} \int_{s_{ij,\gamma-1}}^{s_{ij,\gamma}} \sigma_\gamma e^{\alpha_\gamma t} dB_{j,\gamma}(t) \right) \\ \left. + \sum_{\gamma=\kappa(ij)+1}^{\kappa(j)} \int_{s_{j,\gamma-1}}^{s_{j,\gamma}} \sigma_{j,\gamma} e^{\alpha_{j,\gamma} t} dB_{j,\gamma}(t) \right) \end{aligned} \tag{A2.3}$$

$$\begin{aligned} = e^{-\left(\sum_{\gamma=1}^{\kappa(i)} \alpha_{i,\gamma}(s_{i,\gamma}-s_{i,\gamma-1}) + \sum_{\gamma=1}^{\kappa(j)} \alpha_{j,\gamma}(s_{j,\gamma}-s_{j,\gamma-1}) \right)} \\ \left(\sum_{\gamma=1}^{\kappa(ij)} \int_{s_{ij,\gamma-1}}^{s_{ij,\gamma}} \sigma_\gamma^2 e^{\int_0^t 2\alpha_\gamma dx} dt \right) \end{aligned} \tag{A2.4}$$

$$\begin{aligned} = e^{-\left(\sum_{\gamma=1}^{\kappa(i)} \alpha_{i,\gamma}(s_{i,\gamma}-s_{i,\gamma-1}) + \sum_{\gamma=1}^{\kappa(j)} \alpha_{j,\gamma}(s_{j,\gamma}-s_{j,\gamma-1}) \right)} \\ \left(\sum_{\gamma=1}^{\kappa(ij)} \int_{s_{ij,\gamma-1}}^{s_{ij,\gamma}} \sigma_\gamma^2 e^{2\alpha_\gamma t} dt \right) \end{aligned} \tag{A2.5}$$

$$\begin{aligned} = e^{-\left(\sum_{\gamma=1}^{\kappa(i)} \alpha_{i,\gamma}(s_{i,\gamma}-s_{i,\gamma-1}) + \sum_{\gamma=1}^{\kappa(j)} \alpha_{j,\gamma}(s_{j,\gamma}-s_{j,\gamma-1}) \right)} \\ \left(\sum_{\gamma=1}^{\kappa(ij)} \sigma_\gamma^2 \frac{e^{2\alpha s_{ij,\gamma}} - e^{2\alpha s_{ij,\gamma-1}}}{2\alpha_\gamma} \right) \end{aligned} \tag{A2.6}$$

where from equation (A2.3) to equation (A2.4), the term

$$\begin{aligned} E \left[\left(\sum_{\gamma=1}^{\kappa(ij)} \int_{s_{ij,\gamma-1}}^{s_{ij,\gamma}} \sigma_\gamma e^{\alpha_\gamma t} dB_{i,\gamma}(t) \right) \right. \\ \left. + \sum_{\gamma=\kappa(ij)+1}^{\kappa(i)} \int_{s_{i,\gamma-1}}^{s_{i,\gamma}} \sigma_{i,\gamma} e^{\alpha_{i,\gamma} t} dB_{i,\gamma}(t) \right) \cdot \\ \left(\sum_{\gamma=1}^{\kappa(ij)} \int_{s_{ij,\gamma-1}}^{s_{ij,\gamma}} \sigma_\gamma e^{\alpha_\gamma t} dB_{j,\gamma}(t) \right) \\ \left. + \sum_{\gamma=\kappa(ij)+1}^{\kappa(j)} \int_{s_{j,\gamma-1}}^{s_{j,\gamma}} \sigma_{j,\gamma} e^{\alpha_{j,\gamma} t} dB_{j,\gamma}(t) \right) \end{aligned}$$

is expanded into four different terms. The Itô Isometry is applied to the term

$$\begin{aligned} E \left[\left(\sum_{\gamma=1}^{\kappa(ij)} \int_{s_{ij,\gamma-1}}^{s_{ij,\gamma}} \sigma_\gamma e^{\alpha_\gamma t} dB_{i,\gamma}(t) \right) \cdot \left(\sum_{\gamma=1}^{\kappa(ij)} \int_{s_{ij,\gamma-1}}^{s_{ij,\gamma}} \sigma_\gamma e^{\alpha_\gamma t} dB_{j,\gamma}(t) \right) \right] \\ = \sum_{\gamma=1}^{\kappa(ij)} \int_{s_{ij,\gamma-1}}^{s_{ij,\gamma}} \sigma_\gamma^2 e^{2\alpha_\gamma t} dt \end{aligned}$$

as the integrand functions are the same. For the other three terms, as species i and species j are on different regimes, the independent increment property of Brownian motion and the property of the Itô integral apply, so we have

$$\begin{aligned} E \left[\left(\sum_{\gamma=1}^{\kappa(ij)} \int_{s_{ij,\gamma-1}}^{s_{ij,\gamma}} \sigma_\gamma e^{\int_0^t \alpha_\gamma dx} dB_{i,\gamma}(t) \right) \right. \\ \left. \times \left(\sum_{\gamma=\kappa(ij)+1}^{\kappa(j)} \int_{s_{j,\gamma-1}}^{s_{j,\gamma}} \sigma_{j,\gamma} e^{\int_0^t \alpha_{j,\gamma} dx} dB_{j,\gamma}(t) \right) \right] \\ = E \left(\sum_{\gamma=1}^{\kappa(ij)} \int_{s_{ij,\gamma-1}}^{s_{ij,\gamma}} \sigma_\gamma e^{\int_0^t \alpha_\gamma dx} dB_{i,\gamma}(t) \right) \\ \times E \left(\sum_{\gamma=\kappa(ij)+1}^{\kappa(j)} \int_{s_{j,\gamma-1}}^{s_{j,\gamma}} \sigma_{j,\gamma} e^{\int_0^t \alpha_{j,\gamma} dx} dB_{j,\gamma}(t) \right) \\ = 0. \end{aligned}$$

Similarly, we have

$$E \left[\left(\sum_{\gamma=\kappa(ij)+1}^{\kappa(i)} \int_{s_{i,\gamma-1}}^{s_{i,\gamma}} \sigma_{i,\gamma} e^{\int_0^t \alpha_{i,\gamma} dx} dB_{i,\gamma}(t) \right) \times \left(\sum_{\gamma=1}^{\kappa(ij)} \int_{s_{ij,\gamma-1}}^{s_{ij,\gamma}} \sigma_{j,\gamma} e^{\int_0^t \alpha_{j,\gamma} dx} dB_{j,\gamma}(t) \right) \right] = 0$$

Also by independence of species *i* and *j* after they diverge, we have

$$E \left[\left(\sum_{\gamma=\kappa(ij)+1}^{\kappa(i)} \int_{s_{i,\gamma-1}}^{s_{i,\gamma}} \sigma_{i,\gamma} e^{\int_0^t \alpha_{i,\gamma} dx} dB_{i,\gamma}(t) \right) \times \left(\sum_{\gamma=\kappa(ij)+1}^{\kappa(j)} \int_{s_{j,\gamma-1}}^{s_{j,\gamma}} \sigma_{j,\gamma} e^{\int_0^t \alpha_{j,\gamma} dx} dB_{j,\gamma}(t) \right) \right] = 0.$$

Thus, the covariation for OU_{MVA} is

$$Cov [X_i(T), X_j(T) | X_i(0) = X_j(0) = \theta_0] = e^{-\left(\sum_{\gamma=1}^{\kappa(i)} \alpha_{i,\gamma}(s_{i,\gamma}-s_{i,\gamma-1}) + \sum_{\gamma=1}^{\kappa(j)} \alpha_{j,\gamma}(s_{j,\gamma}-s_{j,\gamma-1})\right)} \left(\sum_{\gamma=1}^{\kappa(ij)} \sigma_{\gamma}^2 \frac{e^{2s_{ij,\gamma}} - e^{2s_{ij,\gamma-1}}}{2\alpha_{\gamma}} \right)$$

Special case 2, OU_{MV} :

Derivation for OU_{MV} case is technically the same as the general case (A2.6) by assuming

$$\alpha_i(t) = \alpha_j(t) = \alpha,$$

$$0 \leq t \leq T$$

Hence, the covariation between species *i* and *j* is

$$Cov [X_i(T), X_j(T) | X_i(0) = X_j(0) = \theta_0] = \frac{e^{-2\alpha T}}{2\alpha} \sum_{\gamma=1}^{\kappa(ij)} (e^{2\alpha s_{ij,\gamma}} - e^{2\alpha s_{ij,\gamma-1}}) \sigma_{\gamma}^2. \tag{A3}$$

Supporting Information

The following supporting information is available for this article:

Table S1. Fossil information and minimum age estimates for the clades calibrated in our divergence time analysis of Monocotyledonae (monocots).

Figure S1. Comparisons of model fit based on AIC weights (w_i) for balanced trees comprised of 32, 64, 128, and 512 taxa.

Figure S2. Comparisons of model fit based on AIC weights (w_i) for pectinate trees comprised of 32, 64, 128, and 512 taxa.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.