Identifying Hidden Rate Changes in the Evolution of a Binary Morphological Character: The Evolution of Plant Habit in Campanulid Angiosperms

JEREMY M. BEAULIEU^{1,*}, BRIAN C. O'MEARA², AND MICHAEL J. DONOGHUE¹

¹Department of Ecology and Evolutionary Biology, Yale University, P.O. Box 208106, New Haven, CT 10620-8106, USA; and ²Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996-1610, USA

*Correspondence to be sent to: National Institute for Mathematical and Biological Synthesis, University of Tennessee, 1122 Volunteer Blvd, Ste. 106 Knoxville, TN 37996, USA; E-mail: jbeaulieu@nimbios.org.

> Received 14 January 2013; reviews returned 18 March 2013; accepted 6 May 2013 Associate Editor: Mark Fishbein

Abstract.—The growth of phylogenetic trees in scope and in size is promising from the standpoint of understanding a wide variety of evolutionary patterns and processes. With trees comprised of larger, older, and globally distributed clades, it is likely that the lability of a binary character will differ significantly among lineages, which could lead to errors in estimating transition rates and the associated inference of ancestral states. Here we develop and implement a new method for identifying different rates of evolution in a binary character along different branches of a phylogeny. We illustrate this approach by exploring the evolution of growth habit in Campanulidae, a flowering plant clade containing some 35,000 species. The distribution of woody versus herbaceous species calls into question the use of traditional models of binary character evolution. The recognition and accommodation of changes in the rate of growth form evolution in different lineages demonstrates, for the first time, a robust picture of growth form evolution across a very large, very old, and very widespread flowering plant clade. [Binary character; Campanulidae; comparative methods; flowering plants; growth habit; herbaceous; Hidden rates model; woody.]

With the availability of larger and larger phylogenetic trees comes the promise of evaluating more complicated models of evolution. Considerable attention has been focused on continuously varying characters, and it is now possible to apply complex, parameter-rich models for detecting differences on different branches of the tree in the processes associated with phenotypic evolution under Brownian motion (O'Meara et al. 2006; Thomas et al. 2006) or the Ornstein–Uhlenbeck process (Butler and King 2004; Beaulieu et al. 2012a). In contrast, discrete binary characters, which are often the focus of evolutionary studies, have received little attention, and we are continually forced to rely on relatively simple models that apply the same rates of state change to all branches in a tree (but see O'Meara 2007).

Simple models of binary character evolution may make sense for small clades, but they are not likely to adequately explain the evolution of such characters in larger, older, and globally distributed clades. In these instances one might expect the lability of a trait—or the propensity to undergo state changes—to differ significantly among clades. In flowering plants, for example, some large and old clades contain only woody species (e.g., Fagales, containing the oaks and their relatives), others contain only herbaceous species (e.g., Brassicaceae, the mustards), and still others show tremendous variation in growth form, apparently with multiple shifts between these growth forms (e.g., Asteraceae, the sunflower and relatives). Failure to account for such differences in transition rates in different portions of the tree could lead to errors in estimating transition rates and the associated inference of ancestral states. In view of the fact that we can now infer extremely large phylogenetic trees (e.g., with over 55,000 tips; Smith et al. 2011) we should have the power to fit models with more than a few parameters.

Here we develop and implement a new method for identifying different rates of evolution in a binary character along different branches of a phylogeny, and we bring this to bear on the inference of ancestral states. We illustrate this approach by exploring the evolution of growth form in Campanulidae, an angiosperm clade containing some 35,000 species, including the familiar composites (sunflowers and relatives), umbels (carrots and relatives), and Dipsacales (honeysuckles and relatives). As we will show, the phylogenetic distribution of woody versus herbaceous species calls into question the use of traditional models of binary character evolution, and the recognition and accommodation of changes in the rate of growth form evolution in different lineages casts a new light on the diversification of campanulids.

METHODS AND DATA

A Hidden Rates Model

Here our primary interest is in understanding the rates of evolution between two habit states—woody (W) and herbaceous (H)—across a very large and variable flowering plant clade. Typically, a likelihood-based ancestral state reconstruction method would be used, and we would fit a model that, at its most complex, would assume only two transition rates governing the probability of change between the two character states (we will refer to this as the TH model, for "time-homogeneous"). The likelihood of this model is defined

SYSTEMATIC BIOLOGY

as being proportional to the probability of the data given a model of evolution:

$$L(\mathbf{Q}) \propto P(\mathbf{D}|\mathbf{Q},\mathbf{T}). \tag{1}$$

In this case, the model, \mathbf{Q} , defines a continuous-time Markov process, and the data, \mathbf{D} , are the observed character states at the tips of a phylogeny \mathbf{T} , whose branch lengths and topology are known. For a binary character, \mathbf{Q} is a 2×2 transition matrix representing the transition rates between the character states:

$$\mathbf{Q} = \frac{W}{H} \begin{bmatrix} - q_{W \to H} \\ q_{H \to W} & - \end{bmatrix}.$$
 (2)

The entries in \mathbf{Q} define the instantaneous transition from W to $H(q_{W \to H})$, and from H to $W(q_{H \to W})$. Implicit in this model is that the transition rates for going between the W and H states are applied to all branches in the phylogeny. That is, if the W to H transition rate is "fast", it will be assumed to be "fast" throughout the entire phylogeny. However, across a very large phylogeny one could easily imagine scenarios wherein the transition rates between the states could differ in different portions of the tree. This is the issue that our hidden rates model (HRM) is intended to address.

Phylogeneticists have long appreciated the need to account for site-specific rate heterogeneity when inferring phylogenies from molecular sequence data, and have developed several methods to do so. For example, Penny et al. (2001) and Galtier (2001) proposed a form of the "covarion" (COncomitantly VARIable codON) model of nucleotide substitution, originally formulated by Fitch and Markowitz (1970), which presupposes that one or more rate classes underlie each state at a site in an alignment. In the covarion model of Penny et al. (2001) there are two such rate classes: a site can either be considered "on", in which transitions among nucleotide states occur freely, or "off", where all transition rates are zero. A more general form of the covarion model, proposed by Galtier (2001), allows for an arbitrary number of rate classes, each with its own transition probabilities among nucleotide states. In either case, the different rate classes represent potentially different transition rates to and from observed states. However, because only the states can be observed, the states are treated as ambiguous observations of the different rate classes. These models are more broadly known as hidden Markov models (HMM) because rate classes are treated as "hidden" states in the Markov process.

Here we describe how a generalization of the covarion model, which we refer to as the "hidden rates model" (HRM), can be used to allow different rate classes to be treated as "hidden" states in reconstructing ancestral character states. For growth form, there may be, for example, two-rate classes, fast (*F*) and slow (*S*), underlying each observed state of *W* and *H*. In this case, an observed character state, such as *W*, might be either *WF* or *WS*. However, since we are only able to observe the character states, we treat each observation as having

a probability of 1 for either being F or S. We can then define a single model, \mathbf{Q} , to account for the process of transitioning between all character state and rate pairs:

$$\mathbf{Q} = \frac{WS}{HS} \begin{bmatrix} - q_{WS \to HS} q_{WS \to WF} & 0\\ q_{HS \to WS} & - & 0 & q_{HS \to HF} \\ q_{WF \to WS} & 0 & - & q_{WF \to HF} \\ 0 & q_{HF \to HS} & q_{HS \to WF} & - \end{bmatrix}.$$
 (3)

In this two-rate example, **Q** is a 4×4 transition matrix with eight possible instantaneous transition rates. The entries in **Q** describing dual transitions, involving changes in both the state and rate (e.g., $WF \rightarrow HS$) are set to zero to force such transitions to either first pass through the same state to a different rate (i.e., $WF \rightarrow HS$), or to pass through a different state in the same rate (i.e., $WF \rightarrow HF \rightarrow HS$). This simplifies the model somewhat by reducing the number of parameters, but it also ensures that the model does not misattribute dual transitions when both state and rate change over longer time periods (Pagel 1994).

Before moving forward, a remark about the biological meaning of the HRM is warranted. In the example above, it could be that HF and HS represent different kinds of herbaceousness; perhaps the former retain a cambial layer that could simply be activated to produce secondary xylem (wood), whereas the latter have completely lost a continuous cambial cylinder, rendering it harder to evolve wood. However, slow or fast could be due to some unmeasured additional character(s). For example, HF could be perennial plants whereas HS could be annual plants. As it is unlikely for annuals to evolve wood, all transitions to woodiness would come from HF plants. Note that if there is a single unmeasured trait, such as annual or perennial, the transition matrix in our HRM approach with two hidden states per observed state is the same as the transition matrix in a standard Pagel (1994) pair of binary traits model. The heterogeneity in transition rates is based on this unmeasured other character, but the HRM should, in theory, be able to account for this. Of course, environmental factors may also affect transition rates. For example, it might be that an herbaceous lineage could evolve woodiness only in the absence of particular competitors. Co-occurrence with particular competitors is not genetically heritable in the classic sense, but descendant generations may nevertheless share this "trait" with their ancestors. In this way, related branches would tend to be similar in transition rate between woody and herbaceous forms based on the presence/absence of competitor. The HF combination could thus correspond to branches that were herbaceous and lacked competitors. There are likely to be many traits or environmental factors that affect transition rates of a focal character. The HRM simply allows the model to effectively "paint" a phylogeny with areas where transitions happen frequently or infrequently due to unmeasured characters that affect the rates.

We also note that the model, \mathbf{Q} , described in equation (3), can easily be modified to allow an arbitrary

number of hidden rate classes. Here we use the Akaike Information Criterion (Akaike 1974) to compare the log-likelihoods of different models. The log-likelihood is equivalent to equation (1) and can be calculated efficiently by making use of conditional likelihoods at each internal node (Felsenstein 1981). The conditional likelihood of an internal node, k, is the product of the probability of observing all descendent character states from its two descendants, node i and node j, given that k is in character state, c_k :

$$L_k(c_k) = \left(\sum_{c_i} P_{c_k c_i}(t_i) L_i(c_i)\right) \left(\sum_{c_j} P_{c_k c_j}(t_j) L_j(c_j)\right).$$
(4)

Under the continuous-time Markov model, the probability of change from one character state to another (e.g., the term $P_{c_k c_i}(t_i)$ for descendent *i*) over a time interval, *t*, can be expressed as, $P(t) = e^{Qt}$, with **Q** representing our model of evolution. Thus, the conditional likelihood calculated at any given internal node is carried out exactly as in the TH model. However, in this case we begin at the tips by summing over the probabilities that are compatible with our observed character state: for example, the probability is set to 1 for both *WS* and *WF* given our observation of a tip being woody. A nonlinear optimization routine is used to find estimates for the entries in **Q** that maximizes the conditional likelihood of the bottom-most internal node, the root of the phylogeny.

For the remainder of the article we focus on evaluating models that contain increasing numbers of hidden rate classes to assess whether portions of a phylogeny where character transitions happen frequently or infrequently are biologically relevant. However, these represent a subset of possible models. For a given rate class, we can compute the number of distinct models using Stirling numbers of the second kind (Abramowitz and Stegun 1972),

$$S(n,k) = \frac{1}{k!} \sum_{i=0}^{k} (-1)^{k-i} \frac{k!}{(i-k)!i!} (i^n)$$
(5)

where n parameters are partitioned into all possible ksubsets. The sum of all possible subsets (i.e., 1, 2, ..., nparameters), or the Bell number (see Pagel and Meade 2006), can be used to calculate the total number of distinct model combinations contained within our HRM framework. Such parameter subsets include, but are not limited to, models where only hidden rate classes only underlie one state as opposed to both (e.g., W, HS, HF), models where particular transitions are set to zero (e.g., $WS \leftrightarrow HS = 0$), or various combinations of models where transition rates vary among, but not within, each rate class (e.g., $WS \leftrightarrow HS \neq WF \leftrightarrow HF$). For an HRM that contains two hidden rate classes, there are 4140 distinct models that could be evaluated (i.e., the sum of eight total parameters partitioned into all possible subsets of $1, 2, \dots, 8$ parameters); for three hidden rate classes there are 190,899,322 distinct models. The relative weight for

each of these models based on AIC can be taken as informative about the evolutionary process, or the best model alone can be selected. Here, we evaluate the fully parameterized models for each rate class, rather than the myriad subsets that set some parameters equal to each other or to zero. For many problems, however, such reduced models may have better fit than more complex ones.

Finally, the observed information matrix (i.e., the Hessian matrix) can be used to test for model nonidentifiability (Formann 1985). The square roots of the diagonals of the inverted Hessian matrix approximate the standard errors of the transition rates. It should be noted, however, that this approach assumes that each parameter estimate is asymptotically normally distributed for the true value as the sample size approaches infinity (Yang 2006). This will not always be the case, especially when the parameter estimate lies near a bound (e.g., q_{WF} =0). Parametric bootstrapping (simulation), which is slower but more accurate, may be more appropriate in such conditions.

Implementation

All methods described above are implemented in the new R package, *corHMM* (pronounced "kor-um") available through CRAN (http://cran.us.r-project.org). Our package requires as input only the observed states at the tips of a phylogeny with branch lengths. Models that contain up to five hidden rate classes can be specified and a bounded subplex routine (NLopt; Johnson 2012) is used to maximize the log-likelihood function to find the optimal parameter estimates for the entries in **Q**. The tree does not have to be fully resolved, and polytomies are allowed by generalizing equation (4) to include as many terms (i.e., probabilities associated with descendant node, *i*,*j*,...,*n*), as there are nodes descending from a focal node.

The marginal reconstruction method of Yang et al. (1995) and Koshi and Goldstein (1996) was implemented to assign the probability of each ancestral state and rate combination to an internal node given the observed tip states and parameter estimates in Q. Our implementation follows the dynamic programming algorithm implemented in Mesquite (Maddison and Maddison 2011) and diversitree (FitzJohn et al. 2009), which starts by traversing the tree from the tips down toward the root computing the conditional probability of each character state for each internal node as described above [equation (4)]. The algorithm then traverses the tree from the root to each tip computing the probability of each character state for the ancestor of a focal node which is taken as the product of two quantities: 1) the probability that the ancestor changed character states along the length of its subtending branch and; 2) the probability a state change occurred along the branch leading from the ancestor to the sister of the focal node. The marginal probability of a focal node is then computed as the probability that a character change occurred along the branch separating the ancestor and the focal node given the above probability for the ancestor and the conditional probability of each state for the focal node. The calculation at the root is similar but involves weighting the conditional likelihood either by weighting each possible character state equally, or weighting by the probability that each character state gave rise to the descendant character states given the transition rates and the tree—a procedure described by FitzJohn et al. (2009) [see equation (10)]. In the latter case, this probability is calculated by dividing the likelihood that the root is in character state *i* by the sum of the likelihoods of all possible character states.

Simulations

We evaluated the performance of the HRM by simulating data sets under two models that contain a single hidden state. The first model assumed three transition rates: two corresponding to separate rates for transitions from state 0 to state 1, and from state 1 to state 0, and one corresponding to transitions from state 1 to a different, hidden and absorbing state (i.e., an evolutionary "dead-end") underlying state 1. The second model was similar to the first, but included an additional rate for transitions to state 1 from the hidden state. The purpose of evaluating these relatively simple models under the HRM was to better understand the behavior of the model and to determine the minimum conditions necessary to adequately infer a hidden state when one indeed exists.

A multiple regression framework was used to examine how several aspects of tree shape and taxon sampling might affect the fit and bias of the HRM. For each model, we simulated 1000 data sets on random trees of various sizes generated under a proportionalto-distinguishable-arrangement model (PDA), and all branch lengths were rescaled so that the total height was 100. Under the PDA model all tree shapes are equiprobable for a given tree size (Slowinski 1990), which ensured that highly imbalanced trees would be obtained at moderate frequency. The shape of each tree was measured using Colless' index (I_c) standardized by the I_c of a completely pectinate tree (a comb) of the same size. Colless' index is the sum of the absolute values of the difference between the numbers of taxa in every pair of clades on the tree (Colless 1982). The size of a randomly generated tree was determined by drawing a value from a uniform distribution that ranged from 32 to 1000 possible taxa. The known parameters for the different transition rates of the generating model were drawn from a uniform distribution in the interval [0, 0.10]. We chose this interval to ensure that, on average, there was > 1 state changes along each root to tip path. We also assessed the overall character change inferred in each of these data sets by calculating the average parsimony score per edge as Pscore/(2n-2), with *n* being the number of taxa. Although parsimony is biased toward underestimating the amount of change, it is straightforward to calculate and could serve as a fast indicator of whether there is enough change to generate sufficient information to fit our model. All simulations were carried out in *corHMM*.

Each simulated data set was evaluated under the generating model, the alternative HRM, and the TH model that assumes no hidden state. 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 2002), was calculated for all models. A generalized linear model (GLM) was used to test for an association between the AIC weight of the true model and tree size, tree shape, and the average number of changes (under parsimony) per edge. Whether or not the generating model had the highest AIC weight was treated as a factor. A binary response required the use of a binomial description of the error distribution in the GLM, which is equivalent to fitting a logistic regression. Finally, we tested for an association between the three predictors and the bias in the parameter estimates of the transitions to and from the hidden state.

Campanulid Phylogeny and Habit

Campanulidae contains four major lineages of angiosperms—Aquifoliales (536 species), Asterales (26,870 spp.), Apiales (5489 spp.), and Dipsacales (1090 spp.)—as well as a number of smaller clades (using the APGIII (2009) for taxonomy): Bruniales (79 spp. from both Bruniaceae and Columelliaceae), Escalloniaceae s.l. (130 spp.), and Paracryphiaceae s.l. (130 spp.) (see Tank and Donoghue 2010). In total, campanulids contain nearly 35,000 species.

We constructed a large sequence-based tree for campanulids using available data from GenBank for 12 gene regions: five coding regions of chloroplast DNA (atpB, matK, ndhF, psbA, and rbcL) and six non-coding regions (rpl16 and rps16 introns, and the intergenic spacers *atpB-rbcL*, *trnK*, *trnS-trnG*, and *trnL*-F), as well as the nuclear ribosomal internal transcribed spacer region (ITS). These regions were chosen because they are among the most commonly used gene regions for molecular phylogenetic studies within campanulids (see Beaulieu et al. 2012b). The chloroplast regions are also relatively less complex with respect to gene duplication and loss. Although ITS can be complicated (see Alvarez and Wendel 2003), it was included here because it is, by far, the best-sampled gene region for campanulids within GenBank. The sequence data set was constructed using the procedures described in Smith et al. (2009) and implemented in the program PHLAWD. It uses a "baited" sequence comparison approach where a set of sequences provided by the user is used to filter GenBank sequences and to determine if sequences are homologous to the gene regions of interest; sequences judged not to be homologous are discarded. For homologous sequences, a saturation analysis is conducted that compares the raw pair-wise sequence distances and those corrected according to a Jukes-Cantor model of molecular substitution. If an alignment appears to be saturated, the matrix is broken up into less inclusive groups based on a user-defined nested taxonomic hierarchy. Here, we used the NCBI taxonomy as our guide for breaking up alignments into smaller matrices. The individual subset alignments are then aligned together using profile-to-profile alignment techniques implemented in MUSCLE (Edgar 2004).

Our final combined sequence matrix of 12,094 sites for 8911 species contained 86.2% gaps or missing sequence data. A maximum likelihood (ML) phylogenetic analysis of the total sequence alignment was performed using RAxML (v.7.2.6; Stamatakis 2006), and tree searches were conducted under the GTR+CAT approximation of rate heterogeneity, partitioned by gene region. A criterion based on the Robinson-Fould (RF) distance was used to assess asymptotic convergence of the log-likelihood score over time; the ML search was terminated if the relative RF distance between two consecutive inferred trees was smaller than 1%. The phylogenetic analysis was restarted 50 times to obtain a set of 50 candidate trees, and for each we estimated branch lengths and identified the tree with the highest likelihood under the $GTR + \Gamma$ model.

The resulting trees were rooted on the branch leading to Aquifoliales based on the analysis of Tank and Donoghue (2010), which found strong support for Aquifoliales as sister to a clade containing the rest of campanulids. We also note that the relationships among the major campanulid lineages in our large sequencebased tree are identical to those of the more focused study of Tank and Donoghue (2010). Relationships among less inclusive clades also mirrored the cumulative results of hundreds of published studies focused on included clades (Beaulieu et al. 2012b). Overall, our tree provides a fairly complete synthesis of the knowledge of relationships within campanulids.

The ML tree was converted to an ultrametric tree using the non-parametric dating method implemented in PATHd8 (Britton et al. 2007). The algorithm in PATHd8 only considers the mean path lengths from a node to its descendants and addresses deviations from the molecular clock locally. We used as fixed age constraints the absolute age estimates obtained from a more focused Bayesian divergence time analysis of 121 campanulid taxa (Beaulieu et al. 2013). In the present analysis, campanulids are estimated to have originated in the Albian (ca. 104 MA) and the primary divergences along the backbone are estimated to have occurred by the Cenomanian (ca. 93 MA). The primary crown clades (i.e., Aquifoliales, Asterales, Apiales, and Dipsacales) are all estimated to have existed by the end of Campanian (\sim 80 MA). In general, these age estimates for campanulids are older than the ages implied directly by the fossil record (e.g., \sim 83.5 MA for crown campanulids, Martínez-Millán 2010), but younger than estimates from previous molecular studies (e.g., ~123 MA for crown campanulids; Bremer et al. 2004).

We scored each of the 8911 species in our sequencebased tree as being either (0) woody or (1) herbaceous. Woody plants (trees or shrubs) produce secondary xylem (true wood) following cell divisions in the vascular cambium. In herbaceous plants the cambium produces little or no secondary xylem tissue. There is ambiguity in some cases. We scored plants that may produce some secondary growth but that die back to the ground each year as herbaceous. Our scoring of habit was based on published accounts, including original taxonomic descriptions and floras. The final tree and character matrix can be downloaded from Dryad (doi:10.5061/dryad.fn50c).

We fit four different models of habit evolution to our campanulid data set. The simplest model assumed separate rates for transitions from W to H, and from Hto W (the TH model). We also assessed the fit of HRM's that assumed two, three, and four hidden rate classes underlying each of the observed W and H states. In each of these models, we treated both W and H observations as being ambiguous for the different hidden rate classes specified in the model. It would be possible, however, to assign some taxa to particular rate categories (e.g., set some species to be HS). In all cases, we performed analyses with short branches collapsed in order to minimize the impact of branch length and phylogenetic uncertainty on the estimates of the transition rates. We also checked for consistency in the fit and estimates of transition rates among the individual ML trees, which were similar from tree to tree. Therefore, for illustrative purposes, we focus on results obtained from the ML tree with the highest likelihood. For the model with the best overall fit based on AIC, we used the estimated parameters to calculate the marginal probabilities of the most likely W and H state and rate at all internal nodes in our dated ML tree. For general comparison, and for our analyses of model adequacy, we also carried out a parsimony analysis using the R package phangorn (Schliep 2011).

Model Adequacy

Model adequacy is a different concern than model fit: the latter determines which of a set of models is least bad (or "best") for the data, whereas the former determines whether a single model adequately describes the data. For example, a model that says humans, chimps, and mice all diverged simultaneously 70 million years ago is a better fit to the data than one that puts that divergence 700 million years ago, but neither model adequately describes the data. Adequacy is often evaluated by simulating under the focal model to see if it generates data indistinguishable from the empirical data using one or more measures.

We checked the adequacy of the HRM in describing the observed growth form data in two ways. In the first test, we examined the expected number of state changes as inferred with parsimony under the HRM model. We simulated 100 random data sets on our ML tree using the parameter estimates from the HRM with three hidden rate classes and for each data set recorded the parsimony score. In the second test, we examined the expected

TABLE 1. Multiple regression results for simulations that assessed how the numbers of taxa, average parsimony score (Pscore) per edge, and the shape of a tree predict the fit and bias of the HRM

Model: $0 \leftrightarrow 1 \rightarrow 1^*$	Slope	P-value
Model fit		
Number of taxa	0.001	0.037
Ave Pscore/edge	1.86	0.096
Tree shape	-5.76	<0.001
Bias in parameter estimates		
Number of taxa	0.000	0.083
Ave Pscore/edge	-0.921	<0.001
Tree shape	-0.047	0.707
Model: $0 \leftrightarrow 1 \leftrightarrow 1^*$		
Model fit		
Number of taxa	0.001	0.002
Ave Pscore/edge	1.25	0.475
Tree shape	-5.97	0.012
Bias in parameter estimates		
Number of taxa	< 0.001	0.111
Ave Pscore/edge	-0.203	<0.001
Tree shape	0.098	0.399

Note: The first HRM (i.e., $0 \leftrightarrow 1 \rightarrow 1^*$) assumed three transition rates: two rates corresponding to and from state 0 to state 1, and one corresponding to transitions from state 1 to a hidden state (i.e., 1*). The second HRM (i.e., $0 \leftrightarrow 1 \leftrightarrow 1^*$) was similar to the first but included an additional rate for transitions to state 1 from the hidden state. Bold italics indicates *P* < 0.05.

distribution of state changes under the model. For each simulated data set, we counted the diversity contained within subclades comprised of species with the same character state, and using a Kolmogorov–Smirnov test, where P > 0.05 indicated a match, we compared the distribution of simulated data sets from the HRM, and those of the TH model, to the observed distribution.

RESULTS

Simulations

Several aspects of the underlying tree were significant predictors of whether or not the generating model was favored. A significant negative association with the shape of the tree indicated that more balanced trees were more likely to favor the generating model than more imbalanced ones (Table 1). There was a significant positive association with the number of taxa sampled (Table 1), although there was a great range in the number of taxa in trees that favored the generating model. At the lower end of the range, there were trees with as few as 60 taxa that favored the generating model when there was a single transition rate to the hidden state. When the model assumed that there were trees with as few as 120 taxa that supported the generating model.

Although the average number of changes per edge was not a significant predictor of model fit, it was the only significant predictor of bias in the parameter estimates associated with the hidden state. The negative relationship between the number of character state changes and the bias was particularly pronounced;



FIGURE 1. Plot of the log-likelihoods for the different models fit to the growth habit data of Campanulidae (campanulids). The addition of a single hidden rate (HRM+2) improved the likelihood by just over 400 log-units over a TH model. When three rate classes were allowed (HRM+3), the likelihood was improved by another 100 log-units. With four rate classes (HRM+4) the likelihood was improved by only an additional seven log-units, and was a model that was too complex for the information contained within the data.

across the range of observed average changes per branch, the bias declined by more than half. However, a more important question for many using these methods may not be "what is the value of a particular transition rate", but rather, "are there differences in the direction of gains and losses." For the model that assumed transitions to and from the hidden state, the transition rate randomly chosen as the higher value was correctly inferred as having a higher value in over 80% of the data sets. Taken together, these results suggest that trees with many state changes will tend to return reliable parameter estimates. Meaningful differences in the parameter estimates can also be detected, which includes estimates associated with transitions to and from a hidden state.

The Evolution of Growth Habit

Comparing the fit of the time-homogeneous (TH) model against the HRM with different numbers of hidden rate classes it is clear that branch-specific rates of evolution have been a major factor underlying growth form evolution in campanulids (Fig. 1). The addition of a single hidden category (i.e., allowing slower and faster rate classes for both the woody and the herbaceous state), and thus only six more parameters, improved the likelihood by just over 400 log-units. When three rate classes (i.e., slow, medium, and fast) were allowed, the likelihood was improved by another 100 log-units. With four rate classes the likelihood was improved by another seven log-units, but contained several parameters that were not identifiable, indicating that this model may be too complex for the information contained within the data. In any case, this model had a \triangle AIC of 1.6 relative to three rate classes, suggesting that this model does not provide a substantial improvement over the model with three hidden rate classes.

The parameters estimated under the different models showed very different patterns in the estimated transition rates between woody (*W*) and herbaceous (*H*) states. Under the TH model, transitions from *H* to *W* ($q_{H\rightarrow W}$ =0.0207; s.e.=±0.0011) were estimated to be more than twice the rate for transitions going from *W* to *H* ($q_{W\rightarrow H}$ =0.0094; s.e.=±0.0013). The most likely ancestral states under this model suggest that all campanulid lineages were ancestrally woody and the herbaceous habit evolved multiple times independently within each of the major clades. In numerous instances, herbaceous lineages are inferred to have re-evolved the woody habit, but the vast majority of these transitions are reconstructed to have taken place within Asteraceae.

The addition of a single rate category into the model effectively creates a slower and a faster rate class for both the woody state and the herbaceous state. The slower rate class (S) followed the same general trend seen in the TH model, with the rate from *H* to $W(q_{HS \rightarrow WS} = 0.0008)$; s.e. = ± 0.0008) being higher than the rate from W to $H(q_{WS \to HS} = 0.0001; \text{ s.e.} = \pm 0.0003)$. In contrast, in the faster rate class (F) the rate from W to $H(q_{WF \rightarrow HF} =$ 0.2023; s.e. = ± 0.0262) was more than twice the rate from *H* to $W(q_{HF \rightarrow WF} = 0.0874$; s.e. = ± 0.0121). The direction of transitions between the different hidden rate classes is also noteworthy. In the W state, transitions between rate classes only occur in one direction, from the faster rate class, *F*, to the slower rate class, $S(q_{WF \rightarrow WS} = 0.0814;$ s.e. = ± 0.0196); the rate from rate class S to rate class *F* was estimated to be zero ($q_{WS \rightarrow WF} = 0.0000$; s.e. = ± 0.0007). In the *H* state, transitions from rate class *F* to rate class $S(q_{HF \rightarrow HS} = 0.0368; \text{ s.e.} = \pm 0.0091)$ were nearly three times higher than transitions from rate class S to rate class $F(q_{HS \to HF} = 0.0131; \text{ s.e.} = \pm 0.0037)$.

The parameters estimated in the HRM with three rate classes for both the woody and herbaceous states-slow (S), medium (M), and fast (F)—revealed even stronger differences in the patterns of transition between the W and *H* states. In rate class *S* the rate from *W* to *H* was estimated to be zero ($q_{WS \rightarrow HS} = 0.0000$, s.e. = ± 0.0002) and the rate from *H* to *W* nearly so $(q_{HS \rightarrow WS} = 0.0009;$ s.e. = ± 0.0007). In rate class *M*, *W* to *H* transitions $(q_{WM \to HM} = 0.0383, \text{ s.e.} = \pm 0.0346)$ were an order of magnitude higher rate than transitions going from H to $W(q_{WM \to HM} = 0.0012, \text{ s.e.} = \pm 0.0149)$. In rate class *F*, the transition rate estimates are very high, but transitions from *W* to $H(q_{WF \rightarrow HF} = 99.8, \text{ s.e.} = \pm 2.9)$ are nearly three times faster than transitions from H to $W(q_{HF \rightarrow WF} = 39.9)$, s.e. ± 1.9). Transitions among the different rate classes follow a general trend in which higher rates are inferred for transitioning toward the slowest rate class (i.e., WS or HS; Fig. 2). Thus, it is generally more likely for a particular branch to be found in either rate class S or rate class *M*, than to be in rate class *F*.

Under the HRM with three rate classes, it is noteworthy that being *W* in rate class *S* represents an absorbing state; that is, herbaceousness may not evolve again once a lineage transitions into this state and rate combination. The most likely ancestral states and rates inferred at each node in our campanulid phylogeny imply a fairly complicated history in the evolution of growth form. All the earliest campanulid lineages were inferred to be woody, including Apiidae (Fig. 2), the least inclusive clade containing Asterales and Dipsacales. The woody habit was retained for much of the early evolution of Apiidae, leading to the primary crown clades (Asterales, Apiales, and Dipscales). The herbaceous habit evolved many times independently from the woody state. Predominantly woody clades such as Rousseaceae (Roussea+Carpodetoideae) and the clade comprised of Alseuosmiaceae-Argophyllaceae-Phelliniaceae within Asterales, Pennantiaceae, Torricelliaceae, Griseliniaceae, and Araliaceae within Apiales, and Adoxaceae and Caprifolieae within Dipsacales, were all inferred to be in this absorbing woody state. Regarding ancestral reconstructions, we note that the TH analysis yielded very similar results (Fig. 2). Inferences based on parsimony were also generally similar, although in this case the woody state was retained throughout the early splits within Araliaceae.

Rates under HRM are equivalent to substitution rates rather than mutation rates: that is, the rates describe what has actually happened over evolutionary time given selective pressures and available variation, as opposed to rates describing what was proposed by mutation. For example, although transitions from the woody state to the herbaceous state are likely more difficult in some lineages than in others, there is at least some possibility of such a transition occurring in every lineage given sufficient time. Even if all extant members of a lineage are woody, we presume that the capacity exists to shut down cambial activity and shift to an herbaceous growth form. It is important to appreciate that the HRM relies strictly on observed state frequencies for inferring rates of transitions within and among different "hidden" rate classes, and may therefore be subject to assigning very low or zero rates under some circumstances.

Much of the action in the evolution of growth form in campanulids has been within the exceptionally diverse Asteraceae (Fig. 3). It appears that for nearly its entire history, Asteraceae has been characterized by higher rates of growth form evolution. The three-rate model suggests that Asteraceae was ancestrally herbaceous, and that the earliest transitions to woodiness occurred within several small lineages (i.e., Barnadesioideae, Mutisioideae, Stifftioideae, and Gochnatioideae; Panero and Funk 2008) that today are largely confined to South America (Bremer and Gustafsson 1997; Funk et al. 2005; Panero and Funk 2008). A majority of the branches associated with higher rates of growth form evolutionespecially transitions from herbaceous to woody forms-are concentrated in the "out of South America" clade (sensu Panero and Funk 2008), which makes up the bulk of Asteraceae diversity. This clade includes several major lineages, such as Carduoideae (2780 species), Cichorioideae (3600 species), and Asteroideae (16,360 species), and its origin appears to coincide with the radiation and worldwide spread of Asteraceae. Each of the major lineages within this group shows elevated



FIGURE 2. Time-calibrated phylogeny of Campanulidae (campanulids). The phylogeny is taken from a ML analysis of 8911 species based on a combined analysis of 11 chloroplast genes and 1 nuclear gene. The major clades of campanulids are labeled and estimates of the most likely growth habit (woody=brown; herbaceous=green) based on the marginal reconstruction using the parameters estimated from the TH model and an HRM with three rate categories (HRM+3). Esc/Brun represents Escalloniales and Bruniales. Transitions between the two states in the TH model and the six combinations of states and rates estimated from the HRM with three rate categories (slow = S, medium = M, and fast = F) are also shown. The thickness of the arrows corresponds to the rates, with thicker arrows denoting higher rates.



FIGURE 3. The distribution of branches inferred as being in the different rate categories in the HRM with three rate categories. Branches are colored based on whether the marginal probability is > 0.75 (dark) of being in each state and rate combination. The evolution of growth habit in campanulids clearly varies among clades, with Asteraceae (denoted by the black dot) generally having higher rates of growth form evolution than any other clade of campanulids.

rates of growth form evolution, and it is especially noteworthy that there have been repeated transitions from herbaceous to the woody habit, often associated with movements onto oceanic islands (Carlquist 1974).

A Note on Model Adequacy

We checked the adequacy of the HRM in describing the observed growth form data in two ways. In the first test, we examined the expected number of state changes as inferred with parsimony under the HRM model. We simulated 100 random data sets on our ML tree using the parameter estimates from the HRM with three hidden rate classes and for each data set recorded the parsimony score. The observed parsimony score for our campanulid data set (obs. Pscore = 502) was not significantly different (P = 0.16) from the parsimony scores from the simulated data (simulated Pscore 95% $CI = 482.7 < \mu < 702.2$). However, when simulating data using the estimated parameters from the TH model, the observed parsimony score far exceeded the range of the parsimony scores from the simulated data (simulated Pscore 95% CI = $157.8 \le \mu \le 271.1, P < 0.01$).

The first test examined whether the overall number of changes matched, but if the HRM model is appropriate, changes should be "clumped" in characteristic ways on the tree. In the second test, we examined the expected distribution of state changes under the model. Within our campanulid data set, a majority of the subclades comprised of only woody or herbaceous states contained few species (<50 species), but there were several that contained as many as 330 species. For each simulated data set, we counted the diversity contained within subclades comprised of species with the same character state. Based on a Kolmogorov–Smirnov test, where P >0.05 indicated a match, 99% of the simulated data sets from the HRM had distributions that matched the observed, as opposed to <5% with the TH model. Thus, the HRM seems to predict the observed clumping much better than the TH does.

On the whole, these tests indicate that an HRM with three rate classes adequately fits the observed data set. It is a model that generally expects a lot of character state transitions that are non-uniformly distributed across the phylogeny. Had we relied on the commonly used TH model, these tests indicate that we would be fitting a model that expects far too little change that is too evenly distributed as compared to what was observed.

DISCUSSION

For decades the dominant models in comparative biology have been homogenous through time and across taxa. As trees have grown dramatically in scope and size, an assumption of homogeneity becomes less defensible. There are undoubtedly unseen factors that affect the evolution of a phenotypic character and this will often presumably manifest as differences in the rate of evolution of the character across lineages. The HRM developed here provides a means of detecting such branch-specific differences in the evolution of a binary character. Unlike existing approaches, for both continuous and discrete characters, which require *a priori* assignment of models to different branches (O'Meara et al. 2006; Thomas et al. 2006; O'Meara 2007; Beaulieu et al. 2012a), the HRM uses the observed character data directly to infer where the evolutionary model shifted in a phylogeny. The HRM can detect branch-specific rate heterogeneity when it exists and, in such instances, the model fits patterns of character change in observed data sets more adequately than simpler models of binary character evolution.

There are several other models that relate to the HRM. The HRM was inspired by the covarion models of nucleotide evolution (Fitch and Markowitz 1970; Galtier 2001; Penny et al. 2001), and can be viewed as a generalization of that approach. These models limit **Q** to contain one quadrant with zero transition rates, one where transition rates are allowed to be non-zero, and the other two quadrants relating to transitions from one matrix to the other. Felsenstein and Churchill (1996) developed a specific HMM for inferring rates from DNA sequences. These models assume that an underlying discrete and "hidden" rate category evolves along a sequence under a Markov process, with all taxa having the same rate category for a given site. In contrast, the HRM uses a single character (although an extension to multiple characters is possible) and allows the rate model to evolve along a phylogeny rather than along a sequence. Phylogenetic mixture models, like the discrete gamma (Yang 1994), or the more general Pagel-Meade model (2004), appear similar to the HRM in allowing sites to be evaluated under different transition models. However, these methods work by summing likelihoods across different models rather than using one model with hidden rates.

The threshold model (Wright 1934; Felsenstein 2005; Felsenstein 2012) is similar to the HRM in allowing different effective transition rates in different clades. Under the threshold model, a continuous trait (the "liability") evolves along a phylogeny under a Brownian motion process, and the state of the observed discrete trait depends on whether the liability is above or below a certain threshold. Where the liability is near the threshold, transitions between discrete states may happen frequently, but where the liability is far from the threshold, these transitions occur more infrequently. This allows for continuous change in transition rates unlike the HRM, which has a fixed number of rates. In theory, the ancestral state of the liability could be inferred (with some uncertainty), allowing different parts of the phylogeny to be distinguished as evolving at higher or lower rates, just as the ancestral state of rate categories can be inferred with the HRM. One somewhat odd feature of the threshold model is that the transition rate will typically decrease as one moves away from a discrete transition both forward and backward in time. At a transition, the liability crosses the threshold: through time, its mean remains at the threshold, but its standard deviation (expected distance from this threshold) increases with the square root of time, making it less and less likely to cross the threshold again. There is, therefore, the expectation under the threshold model that an interesting clade (one with multiple discrete character transitions) will slow down in the future (and was slower in the past). HRM is not constrained in this way: clades may increase in transition rate through time up to some maximum (e.g., if the fast rate category is absorbing), decrease, or remain constant. We note, however, that a modification of the threshold model to include attraction of the liability to the threshold (using, say, an Ornstein–Uhlenbeck process) would reduce this behavior of the threshold model. Another important property of the HRM is that it allows for asymmetry in the rates of gains and losses, and for the evaluation of other specific rate hypotheses, such as whether a rate between two states is zero.

The model that is the most directly comparable to the HRM is the "precursor model" of Marazzi et al. (2012). The precursor model specifies a model of evolution, **Q**, that contains only two transition rates: one corresponding to transitions between an observed state and a hidden state, and one corresponding to transitions between the hidden state and the other observed state. In other words, the model describes the hidden state as "facilitating" transitions to an observed state. This model is nested within the more general HRM, and illustrates an efficient use of parameters in creating a specific, biologically meaningful model in the HRM family.

An important thing to note is that the HRM does not just apply to cases in which there are two observable states that may have hidden rates. In a general sense, it can be used in any instance where the number of observed states is less than the number of actual states, even if rates are homogenous. For example, Maddison (1993) provides a hypothetical example of a set of taxa where the observed states are red tails, blue tails, and no tails. One could treat this in the HRM framework as having four hidden states (tail red, tail blue, tailless red, tailless blue) with three displayed states (tailless red and tailless blue both present because tailless color is unknown). One would be able to use the HRM model here to address questions such as whether particular tailless species are more likely to have genes for red or blue color. For some parameter values there is even some information in the model on whether color can change when in the tailless state. Similarly, HRM models could be used for nucleotide or amino acid data, doing the same sort of fitting as a covarion model (which is one kind of HRM) but allowing more flexibility. For example, there could be three rate categories: fast (no selection), slow (stabilizing selection with occasional new optima), and off (strong stabilizing selection to a fixed optimum).

The development of this model takes place under the assumption that diversification rates are independent of the focal trait (though not necessarily constant). As shown by Maddison (2006), methods can be

misled if this assumption is not true, and the HRM model developed here is no exception. We are developing an implementation that also takes into account diversification, but the approach developed here may still be useful. Users should take note, however, of the potential issues that may occur if the focal traits, or any of their correlated traits, are themselves associated with differential diversification rates. However, for our growth form data set, it seems the HRM adequately fits the data even without taking differential diversification into account, and this is likely to be the case in many empirical applications.

Our analysis of growth form evolution illustrates what we imagine to be the primary use for the HRM by comparative biologists. The goal is to evaluate a number of models, choose between these using the AIC and potentially other factors, and then use the best model(s) to make biological inferences. Having selected a model, we obtained ML parameter estimates from the model, and centered our discussion on the biological significance of these parameter estimates and the corresponding inferred ancestral states. We offer a few notes of caution, however. First, from a model comparison perspective, the HRM will often be difficult to fit to empirical data, largely as a consequence of the large number of parameters included in the model. This is not to say that a complex HRM should not be applied to smaller data sets. As our simulations suggest, there will be circumstances in which data sets containing relatively few species can appropriately fit a more complex model. Second, in the presence of a single underlying "hidden" state, meaningful differences in the associated transition rates can be detected, as long as there are enough changes observed in the data. However, as the number of rate classes is increased, observed character changes may be insufficient within less inclusive clades. In this regard, it may be prudent to interpret the results in light of the standard error associated with each of the parameters, or perhaps to focus on qualitative statements about differences in transition rates. Note that HRM models may be embedded in Bayesian approaches as well, to incorporate priors and to use reversible jump Markov Chain Monte Carlo or similar approaches to weight models.

One general implication of the HRM concerns the way in which patterns of character evolution are addressed across larger, older, and globally distributed clades. Most studies of these questions, at least within flowering plants, have tended to rely on trees with very sparse taxon sampling, usually including one or a few representative species from each of the major lineages within the group (e.g., Sargent 2004; Vamosi and Vamosi 2010; Geeta et al. 2012). Generally speaking, such limited taxon sampling has biased the analysis of broad evolutionary patterns toward more conserved characters that vary little within the large clades represented by the exemplars. More complex and parameter-rich models of character evolution, such as the HRM, coupled with improvements in our ability to sample taxa more extensively, will help to remove this constraint and facilitate studies of characters showing a wider range in evolutionary lability.

This perspective relates directly to the study of growth form evolution in flowering plants. Traditionally, despite evident broader patterns, growth form has been considered too labile to be analysed across large angiosperm clades that tend to occupy a wide range of environments (e.g., Cronquist 1968; see Donoghue 2008). Furthermore, transitions between woody and herbaceous growth forms appear to be fairly simple genetically and developmentally (Groover 2005), involving the suppression and re-expression of only a few genes regulating the cambial activity necessary for wood formation (Lens et al. 2012). Our analysis of the evolution of growth form in campanulids clearly supports the view that growth form is highly labile, but, more importantly, it shows that the rate at which growth form evolves varies enough among clades to be relevant to our understanding of the evolution of growth habit. Thus, growth form appears to be highly constrained in some major clades, such as Aquifoliales, whereas it is much more labile in other clades, especially Asteraceae. Such observations suggest important differences in factors such as the geographic and environmental ranges in different clades and/or genetic/developmental factors that have limited or promoted transitions in certain regions of the tree. Multiple interacting factors may be the rule, as suggested by Asteraceae. By virtue of their seed dispersal and establishment mechanisms, they appear to have a higher likelihood of occupying isolated oceanic islands, where the woody habit may have been favored in multiple lineages (Carlquist 1974). At the same time, it is important to note that shifts back and forth between woody and herbaceous forms may be especially likely in clades such as Asteraceae where many species are somewhat intermediate in form, being either perennial herbs or sub-shrubs, sometimes dying back to a persistent woody base. Likewise, shifts may be much less likely in clades, such as Aquifoliales, comprised exclusively of larger shrubs and trees.

Whatever biological factors might underlie differences in rates of growth form evolution, our results clearly demonstrate that major differences exist and that the introduction of hidden rate classes allows the identification of models that can dramatically improve the fit to the underlying data. The fact that the location of shifts between rate categories can be mapped onto the phylogeny makes it possible to more accurately pinpoint where underlying factors may have changed (cf. Marazzi et al. 2012). The examination of multiple rate shifts has the potential to narrow down on one or a few common causal factors. Importantly, such studies can now be conducted by direct reference to phylogenetic trees, and we no longer have to circumscribe such analyses, or describe the results, by reference to traditional taxonomic groups. That is, we at least have the possibility of discovering that the location of a rate shift does not correspond precisely to some previously named clade, but instead to a (perhaps unnamed) less inclusive or

more inclusive clade (Smith et al. 2011). We suspect that phylogenetic analyses focused on the "hidden" factors promoting or constraining character change will ultimately yield a far richer understanding of the evolutionary process.

SUPPLEMENTARY MATERIAL

Supplementary material, including data files and/or online-only appendices, can be found at http://datadryad.org and in the Dryad data repository (DOI:10.5061/dryad.fn50c).

Funding

This work was supported by the iPTOL program (to J.M.B. and B.C.O.) within the NSF-funded iPlant Collaborative (http://www.iplantcollaborative.org/), and for J.M.B. from NSF [IOS-0842800 to M.J.D.].

ACKNOWLEDGMENTS

Thanks are due to Jeff Oliver, Andrew Leslie, Mark Fishbein, and two anonymous reviewers for helpful suggestions on the manuscript and method implementation. Conversations with Joe Felsenstein reinforced our concern over heterogeneity of processes in large trees.

References

- Abramowitz M., Stegun I.A. 1972. Handbook of mathematical functions, with formulas, graphs, and mathematical tables. Washington (DC): National Bureau of Standards.
- Akaike H. 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19:716–723.
- Alvarez I., Wendel J.F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29:417–434.
- The 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., Jhwueng D.-C., Boettiger C., O'Meara, B.C. 2012a.
- Beaulieu J.M., Jhwueng D.-C., Boettiger C., O'Meara, B.C. 2012a. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. Evolution 66:2369–2383.
- Beaulieu J.M. Ree R.H., Cavender-Bares J., Weiblen G.D., Donoghue M.J. 2012b. Synthesizing phylogenetic knowledge for ecological research. Ecology 93:S4–S13.
- Beaulieu J.M., Tank D.C., Donoghue M.J. 2013. A Southern Hemisphere origin for campanulid angiosperms and traces of the break-up of Gondwana. BMC Evolutionary Biology 13:80.
- Bremer K., Friis E.M., Bremer B. 2004. Molecular phylogenetic dating of asterid flowering plants shows Early Cretaceous diversification. Systematic Biology 53:496–505.
- Bremer K., Gustafsson M.H.G. 1997. East Gondwana ancestry of the sunflower alliance of families. Proc. Natl Acad. Sci. USA 94: 9188–9190.
- Britton T., Anderson C.J., Jacquet D., Lundqvist S., Bremer K. 2007. Estimating divergence times in large phylogenetic trees. Syst. Biol. 56:741–752.
- Burnham K.P., Anderson D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer.

- Butler M.A., King A.A.. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am. Nat. 164: 683–695.
- Carlquist, S. 1974. Island biology. New York and London: Columbia University Press.
- Colless D.H. 1982. Review of phylogenetics: the theory and practice of phylogenetic systematics. Syst. Zool. 31:100–104.
- Cronquist A. 1968. The evolution and classification of flowering plants. Boston: Houghton Mifflin.
- Donoghue M.J. 2008. A phylogenetic perspective on the distribution of plant species. Proc. Natl Acad. Sci. USA 105:11549–11555.
- Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accurate and high throughput. Nucleic Acids Res. 32:1792–1797.
- Felsenstein J. 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. Biol. J. Linn. Soc. 16:183–196.
- Felsenstein J. 2005. Using the quantitative genetic threshold model for inferences between and within species. Philos. T R Soc B 360: 1427–1434.
- Felsenstein J. 2012. A comparative method for both discrete and continuous characters using the threshold method. Am. Nat. 179:145–156.
- Felsenstein J., Churchill G.A. 1996. A hidden Markov model approach to variation among sites in rate of evolution. Mol. Biol. Evol. 13: 93–104.
- Formann A.K. 1985. Constrained latent class models: theory and applications. Brit. J. Math. Stat. Psy. 38:87–111.
- Fitch W.M., Markowitz E. 1970. An improved method for determining codon variability and its application to the rate of fixation of mutations in evolution. Biochem. Genet. 4:579–593.
- FitzJohn R.G., Maddison W.P., Otto S.P. 2009. Estimating traitdependent speciation and extinction rates from incompletely resolved phylogenies. Syst. Biol. 58:595–611.
- Funk V.A., Bayer R.J., Keeley S., Chan R., Watson L., Gemeinholzer B., Schilling E., Panero J.L., Baldwin B.G., Garcia-Jacas N., Susanna A., Jansen R.K. 2005. Everywhere but Antarctica: using a supertree to understand the diversity and distribution of the Compositae. Biol. Skr. Dan. Vid. Sel. 55:343–374.
- Galtier N. 2001. A maximum-likelihood phylogenetic analysis under a covarion-like model. Mol. Biol. Evol. 18:866–873.
- Geeta R., Davalos L.M., Levy A., Bohs L., Lavin M., Mummenhoff K., Sinha N., Wojciechowski M.F. 2012. Keeping it simple: flowering plants tend to retain, and revert to, simple leaves. New Phytol. 193:481–493.
- Groover A.T. 2005. What genes make a tree a tree? Trends Plant Sci. 10:210–214.
- Johnson S.G. 2012. The NLopt nonlinear-optimization package. Version 2.3, http://ab-initio.mit.edu/nlopt.
- Koshi J.M., Goldstein R.A. 1996. Probabilistic reconstruction of ancestral protein sequences. J. Mol. Evol. 42:313–320.
- Lens F., Smets E., Melzer S. 2012. Stem anatomy supports *Arabidopsis thaliana* as a model for insular woodiness. New Phytol. 193: 12–17.
- Maddison W.P. 1993. Missing data versus missing characters in phylogenetic analysis. Syst. Biol. 42:576–581.
- Maddison W.P. 2006. Confounding asymmetries in evolutionary diversification and character change. Evolution 60:1743–1746.
- Maddison W.P., Maddison D.R. 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75, http://mesquiteproject.org.

- Marazzi B., Ane C., Simon M.F., Delgado-Salinas A., Luckow M., Sanderson M.J. 2012. Locating evolutionary precursors on a phylogenetic tree. Evolution 66:3918–3930.
- Martinez-Millán M. 2010. Fossil record and age of the Asteridae. Bot. Rev. 76:83–135.
- O'Meara B.C. 2007. Estimating different rates of gene loss on a tree. Genetics 117:1415–1416.
- O'Meara B.C., Ane C., Sanderson M.J., Wainwright P.C. 2006. Testing for different rates of continuous trait evolution. Evolution 60: 922–933.
- Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc. Roy. Soc. B 255:37–45.
- Pagel M., Meade A. 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. Syst. Biol. 53:571–581.
- Pagel M., Meade A. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. Am. Nat. 167:808–825.
- Panero J.L., Funk V.A. 2008. The value of sampling anomalous taxa in phylogenetic studies: major clade of the Asteraceae revealed. Mol. Phylogenet. Evol. 47:757–782.
- Penny D., McCormish B.J., Charleston M.A., Hendy M.D. 2001. Mathematical elegance with biochemical realism: the covarion model of molecular evolution. J. Mol. Evol. 53:711–723.
- Sargent R.D. 2004. Floral symmetry affects speciation rates in angiosperms. Proc. Roy. Soc. B 271:603–608.
- Schliep K.P. 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27:592–593.
- Slowinski J.B. 1990. Probabilities of *n*-trees under two models: a demonstration that asymmetrical interior nodes are not improbable. Syst. Zool. 39:89–94.
- Smith S.A., Beaulieu J.M., Donoghue M.J. 2009. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. BMC Evol. Biol. 9:1–12.
- Smith S.A., Beaulieu J.M., Stamatakis A., Donoghue M.J. 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.
- Tank D.C., Donoghue M.J. 2010. Phylogeny and phylogenetic nomenclature of the Campanulidae based on an expanded sample of genes and taxa. Syst. Bot. 35:425–441.
- Thomas G.H., Freckleton R.P., Szekely T. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. P R Soc.London B 273:1619–1624.
- Vamosi J.C., Vamosi S.M. 2010. Key innovations within a geographical context in flowering plants: towards resolving Darwin's abominable mystery. Ecol. Lett. 13:1270–1279.
- 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. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39:306–314.
- Yang Z. 2006. Computational molecular evolution. London: Oxford.
- Yang Z., Kumar S., Nei M. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics 141: 1641–1650.