Chapter 16
Hidden Markov Models for Studying the Evolution of Binary Morphological Characters

Jeremy M. Beaulieu and Brian C. O'Meara

Abstract Biologists now have the capability of building large phylogenetic trees consisting of tens of thousands of species, from which important comparative questions can be addressed. However, to the extent that biologists have applied these large trees to comparative data, it is clear that current methods, such as those that deal with the evolution of binary morphological characters, make unrealistic assumptions about how these characters are modeled. As phylogenies increase both in size and scope, it is likely that the lability of a binary character will differ significantly among lineages. In this chapter, we describe how a new generalized model, which we refer to as the "hidden rates model" (HRM), can be used to identify different rates of evolution in a discrete binary character along different branches of a phylogeny. The HRM is part of a class of models that are more broadly known as Hidden Markov models because it presupposes that unobserved "hidden" rate classes underlie each observed state and that each rate class represents potentially different transition rates to and from these observed states. As we discuss, the recognition and accommodation of this heterogeneity can provide a robust picture of binary character evolution.

16.1 Introduction

Underlying many important discoveries in ecology, evolution, and behavior is the use of a phylogenetic tree. Phylogenies allow for the non-independence of taxa to be accounted for while also opening up new ways of examining how traits change

J. M. Beaulieu (✉)
National Institute for Biological and Mathematical Synthesis,
University of Tennessee, Knoxville, TN 37996, USA
e-mail: jbeaulieu@nimbios.org

B. C. O'Meara
Department of Ecology and Evolutionary Biology,
University of Tennessee, Knoxville, TN 37996, USA
e-mail: bomeara@utk.edu

through time, detect correlations between ecological and morphological characters, and better understand patterns of lineage diversification. Prior studies of these questions, particularly of larger, older, and widespread clades, have tended to rely on very sparse taxon sampling or on a representative sample of the major lineages contained within the group. The traits amenable to these types of trees can neither change too little (lest there be no variation to examine) nor too much (lest there be no signal left to detect) and so are naturally a biased set. The amount of sequence data available has grown rapidly. In just 20 years, what has been considered a "large" tree has gone from 500 taxa (Chase et al. 1993) to those that contain more than 50,000 species (Smith et al. 2011, also see Bininda-Emonds Chap. 4 this volume). By scaling up analyses to include many more taxa, we can analyze traits with a much wider range of evolutionary rates, and this includes many traits of great ecological and evolutionary importance.

At the same time, however, it is clear that these large comprehensive phylogenies present new challenges for comparative biology. For decades, the dominant models have always assumed a homogeneous process through time and across taxa. As biologists, we know that life is not evolving according to a homogeneous process. Mass extinction events change the ecological context for evolution. Evolution within lineages can lead to different selective regimes. Such heterogeneity requires that different models be applied to different parts of a phylogeny. Recent attention has been focused almost exclusively on solving this problem for continuously varying characters. It is now possible to apply parameter-rich models for detecting meaningful differences in phenotypic evolution among clades, among specific branches, or even pieces of branches under processes such as Brownian motion (O'Meara et al. 2006; Thomas et al. 2006; Revell 2008) or the Ornstein–Uhlenbeck process (Butler and King 2004; Beaulieu et al. 2012). Models of discrete binary character evolution, on the other hand, have received fairly little attention, and biologists are still forced to rely on conventional models that apply uniform rates of change to all branches in a tree.

Simple models of binary character evolution may make sense for less inclusive clades, such as the traditional genus or family levels, which often contain relatively few instances of character change. But, they are not likely to adequately explain the evolution of a discrete binary character in very large and very old clades. At these phylogenetic scales, it is hard to ignore evident distributions of observed character states. Some parts of the tree will often only exhibit one character state, while other parts apparently have undergone frequent state changes and appear rather labile. Obviously, not accounting for heterogeneity in different parts of a tree can lead to problems with estimates of transition rates and/or the inference of the likeliest ancestral states that accompany them. More broadly, however, when phylogenetic trees can contain many thousands of species, we miss important opportunities to discover patterns that could not previously have been recognized and quantified.

In this chapter, we describe a generalized model, which we refer to as the "hidden rates model" (HRM) (Beaulieu et al. 2013a) that allows for the identification of different rates of evolution in a discrete binary character along different

branches of a phylogeny. This model was inspired by the covarion model (COncomitant VARIable codON; Fitch and Markowitz 1970) of nucleotide substitution, which presupposes that unobserved rate classes underlie each nucleotide state at a site in an alignment, and that each rate class represents potentially different transitions to and from observed states (Penny et al. 2001; Galtier 2001). All these models are part of a class that are more broadly known as Hidden Markov models because the different rate classes are treated as "hidden" states in the Markov process. As we will discuss, our HRM provides a powerful tool for detecting various forms of branch-specific heterogeneity when it exists, and can adequately pinpoint where underlying, but unobserved or unmeasured, factors have influenced the evolution of a binary character, even though the HRM is formally time homogeneous.

16.2 The Hidden Rates Model

Although they emphasize Bayesian methodology, Currie and Meade (Chap. 13 this volume) provide a thorough introduction to the underlying theory of continuous-time Markov models and the various procedures used for calculating their likelihood and estimating transition rates. But briefly, under a likelihood-based approach, the likelihood is defined as being proportional to the probability of observing the data given a model of evolution and a specific tree,

$$L(\mathbf{Q}) \propto P(\mathbf{D}|\mathbf{Q}, \mathbf{T}) \tag{16.1}$$

where the data, $\mathbf{D}$, is a vector of observable character states at the tips of a phylogenetic tree, $\mathbf{T}$, whose branch lengths and topology are assumed to be known. The model of evolution, defined by $\mathbf{Q}$, is an instantaneous rate matrix describing the possible transition rates between character states. For a single binary character that has two observable states, 0 and 1, $\mathbf{Q}$ is a 2×2 matrix,

$$\mathbf{Q} = \begin{bmatrix} - & q_{0 \to 1} \\ q_{1 \to 0} & - \end{bmatrix} \tag{16.2}$$

which we can then use to compare the fit of two models: one where we assume equal transition rates between the two states ($q_{0 \to 1} = q_{1 \to 0}$), or one where we assume two distinct transition rates ($q_{0 \to 1} \neq q_{1 \to 0}$). This matrix is transformed into a transition-probability matrix, symbolized as $P(t)$, and is equal to $e^{\mathbf{Q}t}$, where t represents the length a branch. In general, we use this matrix to calculate conditional likelihoods, which are defined as the sum of the probability of observing everything descended from a focal node given that the focal node is in each character state. These likelihoods are computed for every node in the tree starting with the tips and working down toward the root. Thus, the conditional likelihood at the root represents the likelihood in Eq. (16.1) (for the more details about the dynamic programming

algorithm used to carry out this computation, see Felsenstein 1981). In order to complete the calculation at the root, however, an additional step is needed which involves of weighting the conditional likelihood by the prior probability of the possible states at the root. By default, we assume that each possible character state is weighted equally. Other approaches weight the conditional likelihood 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). This probability is calculated by dividing the likelihood that the root is in each character state and rate combination by the sum of the likelihoods of all possible character states.

Looking at $\mathbf{Q}$ in Eq. (16.2), it is easy to see how unsatisfying it might be to assume that, at most, two transition rates (a forward and backward rate between the two states for our character) govern the evolution of a binary character, particularly when applying such a model to a very broad assemblage of species. As one zooms out, including more and more clades, the factors that are associated with transitions between states are unlikely to be consistent. For example, the frequency of transitions between fleshy and dry fruit types in flowering plants will differ depending on whether or not clades occur in regions where biotic dispersal is more likely (i.e., tropics). In insects, the loss of flight will vary based on the environment, costs of dispersal, or any other correlated factor that can change across a tree. More broadly, processes that can also affect rates of character evolution include generation time, effective population size, the underlying genetic architecture of the trait, and/or mutation rates. However, going into the analysis, we may be unaware or even unclear what these specific factors might be—we just know that these rates could change in different portions of the tree. Thus, the HRM is designed as a means to effectively "paint" areas of a phylogeny where transitions happen frequently or infrequently due to unmeasured characters that affect the rates.

Conceptually, the HRM is a generalized form of the covarion model (Fitch and Markowitz 1970), which is used to infer phylogenies from sequence data by allowing the rate of nucleotide substitutions to not only vary by site, but also along branches. The covarion model assumes that there are two stochastic processes at a site: one for transitions between specified rate classes; and the other for transitions between character states within a given rate class. However, because only nucleotide states can be observed, these rate classes are considered "hidden" states in the model, and therefore, we have to treat each observed nucleotide at a site as an ambiguous observation of the different unobserved rate classes. In the formulation of Penny et al. (2001), the unobserved rate classes are instances when the mutation rate of a nucleotide base is either turned "on," and transitions among the four nucleotide states are possible, or turned "off," where the mutation rate is set to zero.

The unobserved rate classes under the HRM are similar to the model described by Galtier (2001), where they need not be considered "on" or "off," but can represent distinct transition models (i.e., rate class A, rate class B, etc.). In other words, these rate classes may differ in being considered "on" or "off," "fast" or

"slow," or in the direction of asymmetry of transitions between states. As with the covarion model, we assume that each observed character state is an ambiguous observation of the different unobserved rate classes, and we can use the same likelihood framework as in Eq. (16.1) to calculate the conditional probability of all ancestral nodes including the root (Felsenstein 1981). That is, we begin at the tips by summing over the probabilities that are compatible with our observed character state: for example, assuming two rate classes, A and B, the probability is set to 1 for both $0A$ and $0B$ given our observation of a tip being in state 0. We define a new model, $\mathbf{Q}$, to account for the process of transitioning between all character state and rate class pairs:

$$\mathbf{Q} = \begin{bmatrix} - & q_{0A \to 1A} & q_{0A \to 0B} & 0 \\ q_{1A \to 0A} & - & 0 & q_{1A \to 1B} \\ q_{0B \to 0A} & 0 & - & q_{0B \to 1B} \\ 0 & q_{1B \to 1A} & q_{1B \to 0B} & - \end{bmatrix} \tag{16.3}$$

Notice in this particular case that the entries in $\mathbf{Q}$ describing dual transitions (state and rate class transitions occur simultaneously) are set to zero to force such transitions to either pass through the same state to a different rate, or to pass through a different state in the same rate (see Pagel 1994). This assumption can be relaxed, of course, by simply adding these transitions back into the model. As with the likelihood model described above, a nonlinear optimization routine is used to find estimates for the entries in $\mathbf{Q}$ that maximizes the conditional likelihood at the root.

16.3 HRM Model Space

Under the HRM framework, the $\mathbf{Q}$ matrix described in Eq. (16.3) can easily be modified to allow any number of hidden rate classes (Fig. 16.1). In fact, model complexity can range from a model with a single rate class, which is the same as the familiar time-homogeneous model in Eq. (16.2), to a model that includes an infinite number of rate classes. One extreme case, where a hidden rate class is assumed for every branch, is similar in effect to the "no-common mechanism" model, which is a parsimony equivalent (Tuffley and Steel 1997). Of course, it is unlikely that such a model would ever fit the data well, because the number of parameters would far exceed the number of data points (see Holder et al. 2010). Nevertheless, we can use model selection methods, such as Akaike's information criterion (AIC) (Akaike 1974), to obtain the model that best fits the data, or calculate the relative weight for a set of models that can be taken as information about the evolutionary process. As with Huelsenbeck et al. (2004), Pagel and Meade (2006), a reversible jump MCMC approach could also be used to get the posterior probabilities of various models.
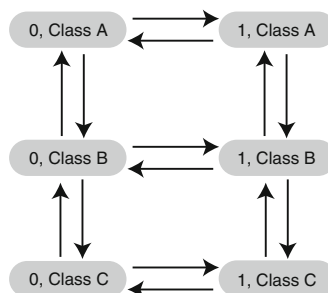
The discussion thus far has focused exclusively on evaluating models that contain increasing numbers of hidden rate classes. However, these represent a very
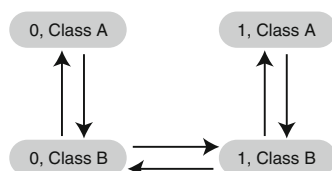
(a) HRM of increasing complexity
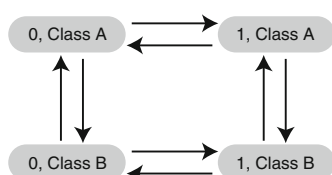
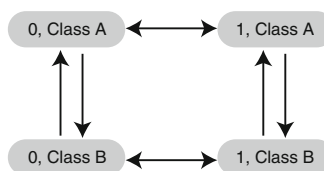**(b)** Examples of subset HRM+2 rate classes
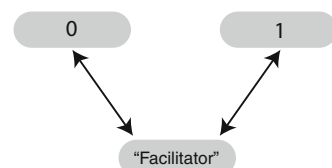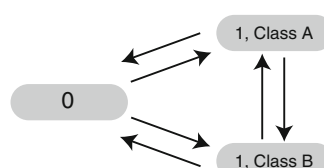
Fig. 16.1 a Graphical representation of HRM's with increasing number of hidden rate classes. Under the HRM framework, model complexity can range from a model with a single rate class, which is the same as a model that assumes a homogeneous process, to a model that includes an infinite number of rate classes. Here, we highlight HRM's with two- and three-ordered hidden rate classes, with the *black arrows* denoting the directions of the possible transition among the different state and rate class combinations. **b** For a given number of hidden rate classes, there are many models that contain a subset of the possible parameters (see HRM MODEL SPACE). For an HRM with two rate classes such subset models include, but are certainly not limited to, the "precursor" model of Marazzi et al. (2012), models where hidden rate classes underlie one state as opposed to both, models where particular transitions are set to zero as in the covarion model, or various combinations of models where transition rates vary among, but not within, each rate class

small subset of the possible models that could be evaluated. For instance, the "precursor model" of Marazzi et al. (2012) illustrates an efficient use of parameters in creating a specific model in the HRM framework (Fig. 16.1). Under the precursor model, there are two rates that correspond to transitions between one

observed state and a "hidden" precursor state, and between the hidden state and the other observed state. This model essentially describes the hidden state as "facilitating" transitions to and from the observed states.

To get a sense of the complete model space provided by the HRM framework, we can use Stirling numbers of the second kind (Abramowitz and Stegun 1972) to count the different combinations of models for a given a number of parameters. The Stirling numbers are computed as

$$S(n,k) = \frac{1}{k!} \sum_{i=0}^{k} (-1)^{k-i} \frac{k!}{(i-k)!i!} (i^n) \tag{16.4}$$

where n parameters are partitioned into all possible k subsets (i.e., 1, 2, …, n parameters). The Bell (1934) number, which is the sum of these Stirling numbers, can be used to count the total number of distinct model combinations contained within a given HRM. For an HRM that contains two hidden rate classes, there are 4,140 distinct models that could be evaluated (i.e., the sum of 8 total parameters partitioned into all possible subsets of 1, 2, …, 8 parameters); for three hidden rate classes there are 190, 899, 322 distinct models. Note however that this calculation assumes that the rate classes are ordered (see Fig. 16.1). This need not always be the case, and an HRM could easily be constructed that assumes an unordering of the different rate classes. Of course, the addition of several more parameters to account for transitions between all rate classes would make the model even more parameter rich and contain even more distinct subset models than in the ordered case. In any event, parameter subsets of a three-ordered hidden rate classes include, but are not limited to, models where hidden rate classes only underlie one state as opposed to both, models where particular transitions are set to zero (as in the covarion model), or various combinations of models where transition rates vary among, but not within, each rate class (Fig. 16.1). Again, we can use model selection methods to either obtain the relative weight for each of these models and average the parameters (i.e., model averaging approach), or simply determine which model is the best fit among the set.

16.4 HRM in Relation to Other Models

HRM is not the only model that deals with heterogeneity. As mentioned above, the covarion model (Penny et al. 2001; Galtier 2001) allows states to switch from off to on: effectively an extreme form of the HRM model where one of the categories has no transitions. An important set of models developed by Yang et al. (1995) fit different rates to different branches of a tree. These are typically applied for codon models but can in principle be applied to any discrete data. Yang (1994) also developed the use of a discrete gamma to deal with rate heterogeneity across sites. The likelihood of the data is calculated under each of several rates pulled from a

distribution described by a single parameter. Pagel and Meade (2004) develop a phylogenetic mixture model that is essentially a generalization of this; rather than just allowing likelihoods to be calculated across different overall rates, they allow the rate matrix to also vary.

The threshold model (Wright 1934; Felsenstein 2005, 2012) represents a different approach to dealing with heterogeneity. Rather than a hidden discrete trait affecting the rate of the observed trait, it allows for a hidden continuous trait to set the state of the observed trait. When this continuous trait, termed the liability, crosses the threshold, the discrete character changes state. This has been extended for multi-state-ordered characters (Revell 2014). The behavior of the model is grossly similar to HRM in that the frequency of discrete state changes varies over the tree; with the HRM, due to different hidden rates, and in this case, due to distance of the liability from the threshold. There are some important differences, however. With the canonical threshold model, there is the expectation over time that the liability moves away from the threshold, as it evolves with unbounded Brownian motion, and so the long-term expectation is that the transition rate eventually becomes zero. Revell (2014) has largely addressed these issues by modifying the threshold model so that the liability evolves with bounds or with a strong attraction (i.e., Ornstein–Uhlenbeck process, see O'Meara and Beaulieu Chap. 16 this volume) back to the threshold. However, the HRM also allows this in that it could have absorbing states, but it does not require it. The threshold model also assumes that near a change, where the liability is close to the threshold, the rate of gain or loss of a trait is equal, though the overall gain and loss rates could still be unequal by starting with a liability greater or lower than the threshold. The HRM can allow unequal rates on parts of the tree or over the whole tree. An advantage of the threshold model is that it only has parameters for the starting liability, the threshold value, and rate of movement of the liability; making it a relatively efficient way to fit changing rates over a tree.

16.5 Application of the HRM

The development of the HRM was motivated by the desire to understand rates of evolution between two growth habit states, woody and herbaceous, within campanulid angiosperms, a large flowering plant clade containing some 35,000 species, including the familiar composites (sunflowers and relatives), umbels (carrots and relatives), and Dipsacales (honeysuckles and relatives). Historically, growth habit has been considered by botanists to be far too labile to be understood across larger, more traditional taxonomic ranks (Cronquist 1968). This is because the vegetative features of a plant are thought to be intimately tied to the environment in which they exist, and since species contained within larger, older, and globally distributed clades can occur in a range of environments, growth habit should vary considerably at larger phylogenetic scales. Furthermore, there is increasing genetic evidence that transitions between woody and herbaceous forms should be fairly easy (Groover 2005), involving the suppression and re-expression

of only a few genes regulating both the cambial growth necessary for wood formation and the onset of flowering (Lens et al. 2012).

Such genetic integration of both vegetative and reproductive development suggests that the capacity for growth habit evolution should be ever present within flowering plants. However, it is quite puzzling that the distribution of particular habits is non-uniform across the angiosperm phylogeny. For example, some large and old clades contain only woody species (e.g., Fagales, which contain oaks and their relatives), others contain only herbaceous species (e.g., Brassicaceae, which contain mustards and their relative), and still others show extreme variation in habit presumably through many transitions between woody and herbaceous states (e.g., Asteraceae, sunflowers, and their relatives). These observations alone call into question the use of conventional likelihood-based methods for understanding the evolution of growth habit.

Here, we show results from a recent study of 8,911 campanulid species where growth habit was scored and where a large comprehensive phylogeny was used (Beaulieu et al. 2013a). We compare the fit of five models of evolution. The two simplest models are the conventional time-homogeneous models, where we either assume there are equal transition rates between woody and herbaceous states, or there are two distinct transition rates, one for transitions from woody to herbaceous and another for transitioning from herbaceous to woody. We also assessed the fit of HRM's that assume two, three, and four hidden rate classes underlying each observed woody and herbaceous state (see Fig. 16.1 for how these models are graphically structured).

When comparing the fit of the time-homogenous models to the HRM's with different numbers of hidden rate classes, it is clear that models of branch-specific rates of evolution are by far the better fit to the growth habit data. For instance, the addition of just a single hidden category provides an extraordinary improvement in the likelihood over both the one-rate and two-rate time-homogeneous models (just over 400 log-likelihood units; Fig. 16.2). With three hidden rate classes the likelihood is improved by an additional 100 log units, but begins to plateau, where the fit of a model with four hidden rate classes did not substantially improve the likelihood any further.

A HRM with three rate classes was the best-fit model overall based on AIC and with the estimated parameters indicating a complicated process of transitioning between woody and herbaceous states in an ancestrally woody clade. Incidentally, the model suggests that the three hidden rate classes represent transition models of increasing rate: "slow," "medium," and "fast." In practice, such structured rate classes will not always result, and describing rate classes in such terms as "slow" or "fast" will be inappropriate, though it works in this particular case. In the slow-rate class, the asymmetry in transition rates is similar to the time-homogenous model in that it is more likely for herbaceous species to re-evolve a woody habit than the reverse ($q_{WS \to HS} = 0.0000$, s.e. $= \pm 0.0002$; $q_{HS \to WS} = 0.0009$; s.e. $= \pm 0.0007$). In fact, being woody in the slow-rate class represents an absorbing state: herbaceous may not evolve again once a lineage transitions into this state and rate combination. However, it is important to emphasize that rates
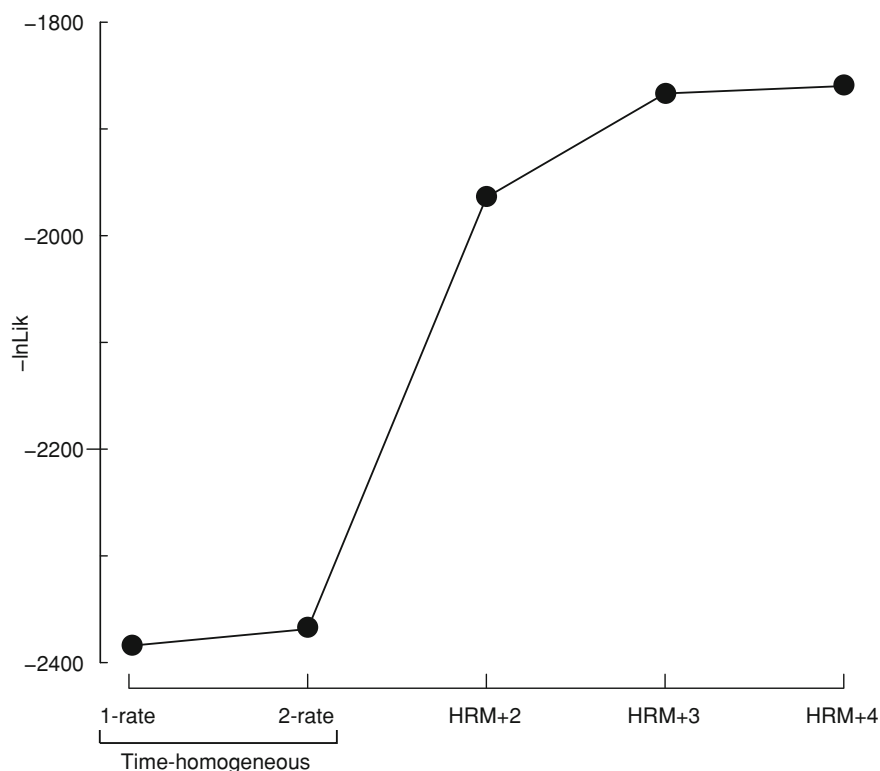
Fig. 16.2 Plot of the log-likelihoods for the different models fit to the growth habit data of Campanulidae (campanulids) from Beaulieu et al. (2013a, b). The addition of a single hidden rate (HRM + 2) improved the likelihood by just over 400 log units over both time-homogenous models. When three rate classes were allowed (HRM + 3), the likelihood was improved by another 100 log units. Four rate classes (HRM + 4) did not substantially improve the likelihood any further

under the HRM are equivalent to substitution rates, not mutation rates, and therefore, only provide an indication of what happened over evolutionary time and not what was proposed by mutation. Thus, in this particular case, the capacity to transition to herbaceous may still exist in these plants, unless, of course, this rate class represents instances where clades have lost the genetic machinery to shut down cambial activity. Finally, in the medium ($q_{WM \to HM} = 0.0383$, s.e. $= \pm 0.0346$; $q_{WM \to HM} = 0.0012$, s.e. $= \pm 0.0149$) and fast ($q_{WF \to HF} = 99.8$, s.e. $= \pm 2.9$; $q_{HF \to WF} = 39.9$, s.e. $= \pm 1.9$) rate categories, it is far more likely for a lineage to transition from woody to herbaceous. The model also indicates that transitions among the different rate classes follow a general trend in which higher rates are inferred for transitions toward a slower rate class.

The general picture emerging within campanulids, particularly when "painting" the likeliest state and rate combinations onto internal nodes clearly supports the view
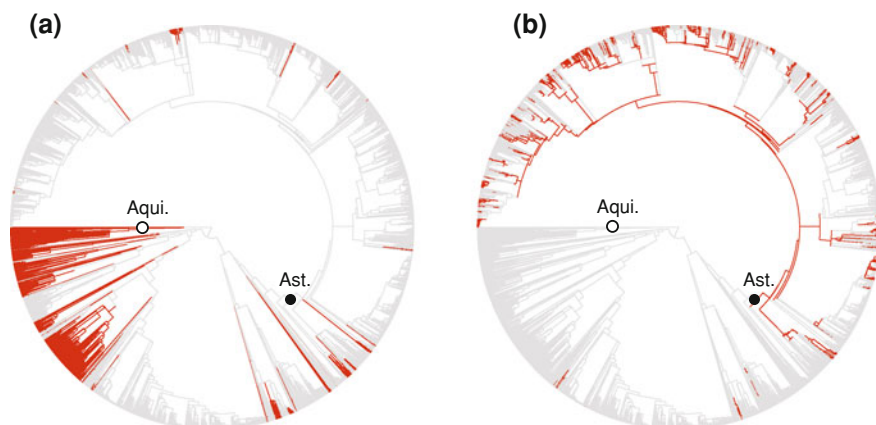
Fig. 16.3 Examples of branches "painted" as being in the different state and rate classes in the HRM with 3 rate classes applied to a phylogeny of 8,911 species of Campanulidae (campanulids) from Beaulieu et al. (2013a). Branches are *colored* based on whether the marginal probability is >0.75 (*dark*) of being in each growth habit state and rate class combination. The evolution of growth habit in campanulids clearly varies among clades. For example, woody clades such as Aquifoliales (denoted by an *open dot*) have slower, and the herbaceous clades such as Asteraceae (denoted by a *black dot*) generally have faster, rates of growth habit evolution, **a** woody, slow, **b** herb, medium + fast

that the rate at which growth habit evolves varies enough among clades to be biologically meaningful (Fig. 16.3). For example, most of the branches inferred to be in the faster rate classes appear to be confined only to Asteraceae, a geographically widespread clade (Bremer and Gustafsson 1997; Beaulieu et al. 2013b), where changes in growth habit as an adaptive response to new environments is well documented within the group (Carlquist 1974). In other groups, the evolution of growth habit is clearly limited, which may also point to their ecology, but also may reflect other additional underlying genetic factors. Within the Aquifoliales (e.g., hollies and their relatives), for example, the herbaceous habit has either never evolved or is a strategy that has not been successful, even though the group is currently widely distributed across both tropical and temperate regions. Whatever factors may be underlying these differences observed among clades, the use of the HRM clearly demonstrate that even though the capacity for growth habit evolution may be ever present, it is clearly expressed in fundamentally different ways.

16.6 Future Directions and Conclusions

Incorporating different rate classes of transition rates are not limited to binary characters. In a general sense, the concept of including "hidden" states can be used in any instance where the number of observed states is less than the number of

actual states. In other words, it is rather straightforward to extend the HRM to include characters that take on multiple states (i.e., >2 states). A good example comes from Maddison (1993): Consider a set of taxa whose observed states are red tails, blue tails, and no tails. In the HRM framework, one could treat this as having four hidden states—tail red, tail blue, tailless red, tailless blue—with three displayed states with tailless red and tailless blue both being present because tailless color is unknown. The HRM could then be used to address questions such as whether particular tailless species are more likely to have genes for red or blue color.

Similarly, biologists are not always interested in the transitions back and forth between states in just one binary character, but rather how the state of one binary character can affect the probability of change in another. There are ways to do this, of course, which have been highly influential (see Pagel 1994), but they still assume that the evolutionary process is homogeneous across the tree. Future extensions of the HRM will include models of correlated evolution between two or more binary characters in order to provide a new way of understanding the overall strength of character correlations and how it can change in various portions of phylogeny.

Future extensions will also include developing a HRM that, along with estimating transition rates, will also estimate the effect of a character state on speciation and extinction rates (i.e., BiSSE approach; Maddison et al. 2007; FitzJohn et al. 2009). Often these types of methods are used to test whether a character state is a "key innovation," as indicated by one state being correlated with higher net diversification rates (i.e., speciation–extinction) relative to another. At greater phylogenetic scales, however, the real effect between a character state and diversification rates is not always clear. What may seem like a causal connection may actually be due to other unmeasured factors or because the analysis included a nested clade that exhibits both the focal character and "something" else (Beaulieu and Donoghue 2013). The development of an HRM that includes the estimation of speciation and extinction rates as they relate to character states and rate classes would provide a powerful extension and allow for a more refined understanding of how particular character states influence the diversification process.

For now, the hidden rates framework can be used as a means of detecting differences in the evolution of a binary character and for the identification of models that can dramatically improve the fit to the underlying data. Unlike most 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. 2012), the HRM uses the observed character data directly to infer where the evolutionary model shifted in a phylogeny. In this way, the HRM is inherently exploratory, much like methods such as AUTEUR (Eastman et al. 2011) and SURFACE (Ingram and Mahler 2012; Mahler and Ingram Chap. 22 this volume) are for understanding clade-specific differences in continuous trait evolution. With the availability of such exploratory tools, we no longer have to restrict analyses, or describe results, in terms of a particular taxonomic rank (genus, family, order, etc.). Instead, we may begin to discover that

important evolutionary events better correspond to groups of taxa that do not have a formal name (e.g., Smith et al. 2011). It is in this way that methods such as the HRM will afford us with a far better understanding of evolution.

References

Abramowitz M, Stegun IA (1972) Handbook of mathematical functions, with formulas, graphs, and mathematical tables. National Bureau of Standards, Washington (DC)

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

Beaulieu JM, Donoghue MJ (2013) Fruit evolution and diversification in campanulid angiosperms. Evolution 67:3132–3144

Beaulieu JM, Jhwueng D-C, Boettiger C, O'Meara BC (2012) Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. Evolution 66:2369–2383

Beaulieu JM, O'Meara BC, Donoghue MJ (2013a) Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. Syst Biol 62:725–737

Beaulieu JM, Tank DC, Donoghue MJ (2013b) A Southern Hemisphere origin for campanulid angiosperms and traces of the break-up of Gondwana. BMC Evol Biol 13:80

Bell ET (1934) Exponential polynomials. Ann Math 35:258–277

Bremer K, Gustafsson MHG (1997) East Gondwana ancestry of the sunflower alliance of families. Proc Natl Acad Sci USA 94:9188–9190

Butler MA, King AA (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am Nat 164:683–695

Carlquist S (1974) Island biology. Columbia University Press, New York and London

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD et al (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Ann Mo Bot Gard 80:528–580

Cronquist A (1968) The evolution and classification of flowering plants. Houghton Mifflin, Boston

Eastman JM, Alfaro ME, Joyce P, Hipp AL, Harmon LJ (2011) A novel comparative method for identifying shifts in the rate of character evolution on trees. Evolution 65:3578–3589

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 Trans 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

Fitch WM, 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 RG, Maddison WP, Otto SP (2009) Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. Syst Biol 58:595–611

Galtier N (2001) A maximum-likelihood phylogenetic analysis under a covarion-like model. Mol Biol Evol 18:866–873

Groover AT (2005) What genes make a tree a tree? Trends Plant Sci 10:210–214

Holder MT, Lewis PO, Swofford DL (2010) The Akaike Information Criterion will not choose the no common mechanism model. Syst Biol 59:477–485

Huelsenbeck JP, Larget B, Alfaro ME (2004) Bayesian phylogenetic model selection using reversible jump Markov Chain Monte Carlo. Mol Biol Evol 21:1123–1133

Ingram T, Mahler DL (2012) SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. Methods Ecol Evol 4:416–425

Lens F, Smets E, Melzer S (2012) Stem anatomy supports *Arabidopsis thaliana* as a model for insular woodiness. New Phytol 193:12–17

Maddison WP (1993) Missing data versus missing characters in phylogenetic analysis. Syst Biol 42:576–581

Maddison WP, Midford PE, Otto SP (2007) Estimating a binary character's effect on speciation and extinction. Syst Biol 56:701–710

Marazzi B, Ane C, Simon MF, Delgado-Salinas A, Luckow M, Sanderson MJ (2012) Locating evolutionary precursors on a phylogenetic tree. Evolution 66:3918–3930

O'Meara BC, Ane C, Sanderson MJ, Wainwright PC (2006) Testing for different rates of continuous trait evolution. Evolution 60:922–933

O'Meara BC (2007) Estimating different rates of gene loss on a tree. Genetics 117:1415–1416

Pagel M (1994) Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc R 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

Penny D, McCormish BJ, Charleston MA, Hendy MD (2001) Mathematical elegance with biochemical realism: the covarion model of molecular evolution. J Mol Evol 53:711–723

Revell LJ (2008) On the analysis of evolutionary change along single branches in a phylogeny. Am Nat 172:140–147

Revell LJ (2014) Ancestral character estimation under the threshold model from quantitative genetics. Evolution 68:743–759

Smith SA, Beaulieu JM, Stamatakis A, Donoghue MJ (2011) Understanding angiosperm diversification using small and large phylogenetic trees. Am J Bot 98:1–12

Thomas GH, Freckleton RP, Szekely T (2006) Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. Proc R Soc Lond B 273:1619–1624

Tuffley C, Steel M (1997) Links between maximum likelihood and maximum parsimony under a simple model of site substitution. Bull Math Biol 59:581–607

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, Kumar S, Nei M (1995) A new method of inference of ancestral nucleotide and amino acid sequences. Genetics 141:1641–1650