

# Models of morphological evolution

Fredrik Ronquist

September 21, 2005

In this lecture, we will be covering models of morphological evolution. These fall into two different classes: those that model morphology in terms of discrete states and those that treat morphology as quantitative characters. Both are applicable to other types of data; for instance, the discrete-state models can be applied to restriction site data and other data of a presence/absence nature, and the quantitative character models can be used as approximations of allele frequency evolution (see previous lecture).

## 1 Morphology as discrete states

In many cases, morphological variation falls naturally into discrete states. Typical examples include the presence or absence of major features like wings and feathers. Although it is currently standard practice to use parsimony methods, like Fitch optimization, to infer phylogenies from such discrete-state morphological data, we can also apply discrete-state continuous-time Markov models like the ones we have seen previously for molecular sequence data. A few minor complications arise but they are all possible to address.

### 1.1 The binary model

For a morphological character with two states (0 and 1) we can simply use a Markov model with the instantaneous rate matrix (unscaled)

$$Q = \{q_{ij}\} = \begin{Bmatrix} - & 1 \\ 1 & - \end{Bmatrix}$$

This is obviously an analogue of the Jukes-Cantor model of DNA evolution. Lewis (2002) recently referred to this model as the *M2 model*; it has been in use for morphological and other types of discrete data at least since the 80's.

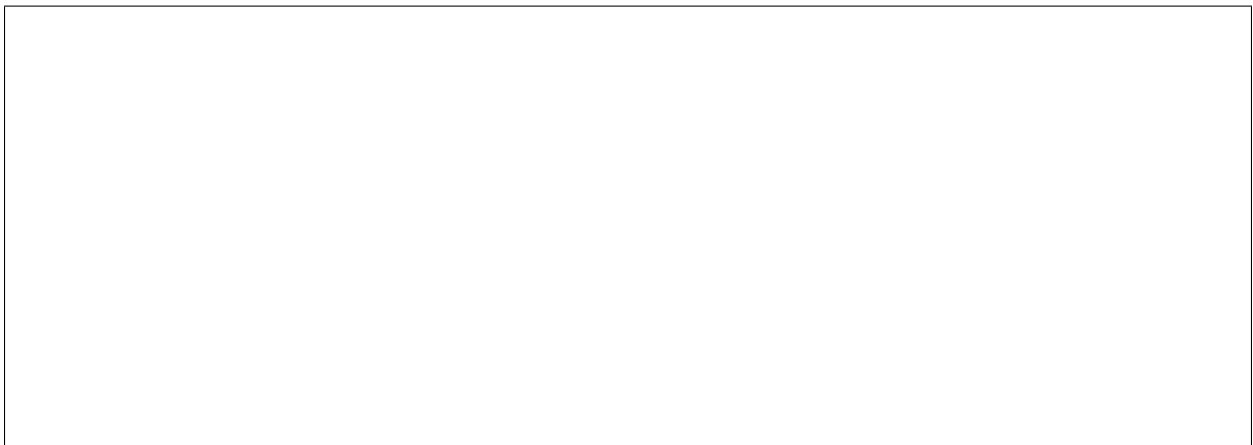
In principle, the M2 model can be easily extended to multi-state characters, but the issue of state ordering comes up. Recall that parsimony methods distinguish between ordered and unordered multi-state characters (Fig. 1). In the simplest case, we assume that all changes between states are possible (Fig. 1a); the alternative is to order the states in a linear series such that, in a three-state character, changes between the two end states have to go through the intermediate state (Fig. 1b). In a stochastic model, we would simply set the instantaneous rate of the impossible changes to zero and use a uniform rate for the other events. Thus, the stochastic model for a three-state unordered character, we can refer to it as the M3u model, is:

$$Q = \{q_{ij}\} = \begin{pmatrix} - & 1 & 1 \\ 1 & - & 1 \\ 1 & 1 & - \end{pmatrix}$$

The equivalent for the three-state ordered character, the M3o model, is:

$$Q = \{q_{ij}\} = \begin{pmatrix} - & 1 & 0 \\ 1 & - & 1 \\ 0 & 1 & - \end{pmatrix}$$

Figure 1: Stochastic models for unordered (a) and ordered (b) morphological characters.



Note the similarity between the M3o model and the codon models we discussed earlier. As with

the latter models, the zero entries in the instantaneous rate matrix of the M3o model do not result in transitions between the end states being impossible but they do force those transitions to go through the intermediate state, thus lowering the transition probability. An interesting but somewhat counterintuitive property of the M3o model is that it has the stationary state frequencies equal for all states; the rate asymmetries only affect the intensity with which the state changes occur. Thus, if you start a set of characters in the intermediate state, you expect to approach equilibrium faster than if you start with the same set of characters in one of the extreme states.

The M3o and M3u models can easily be extended to any number of states. However, this raises the question of how to determine the appropriate state space for a given morphological character. The state space may be obvious for some characters but for the majority it is difficult to determine the state space in any other way than by simply recording the number of observed states in some set of organisms. If the number of states is based on observations, we may wish to let the size of the state space of each character be a parameter of the model. However, the probability of the state space being larger than the number of observed states appears to be small for most real data sets; it is only slowly evolving characters on small phylogenies that are likely to have unobserved states.

## 1.2 Ascertainment bias [JF: Chapter 15:234-235]

An issue that arises with both morphological and restriction site data is *ascertainment bias*, also called coding bias. When calculating the probability of observing some data on a tree, we typically assume that all character patterns can be observed. For instance, with a four-species tree and a four-by-four model of DNA evolution for  $n$  characters, we could calculate the probability of observing all the  $4^4 = 256$  character patterns from AAAA over AAAC to TTTT on a given tree. If the probability of pattern  $i$  is  $p_i$ , we have that  $\sum p_i = 1$ . The likelihood  $L$  (or the probability of the data given some parameter values) can be calculated by taking the number of observations  $f_i$  of each pattern  $i$  into account in the product over sites:  $L = \prod p_i^{f_i}$ .

If certain types of character patterns cannot be observed, then the likelihood of the observed data calculated naively according to this formula will be too high, with potentially serious effects on parameter estimates. In restriction site data, for instance, we typically cannot observe sites that are absent from all of the studied taxa. To compensate for this, we simply divide the probability of each character by the probability of the unobservable pattern. If the probability is  $p_0$  for the all-absence pattern, then the corrected likelihood is  $L^* = \prod p_i/p_0$ .

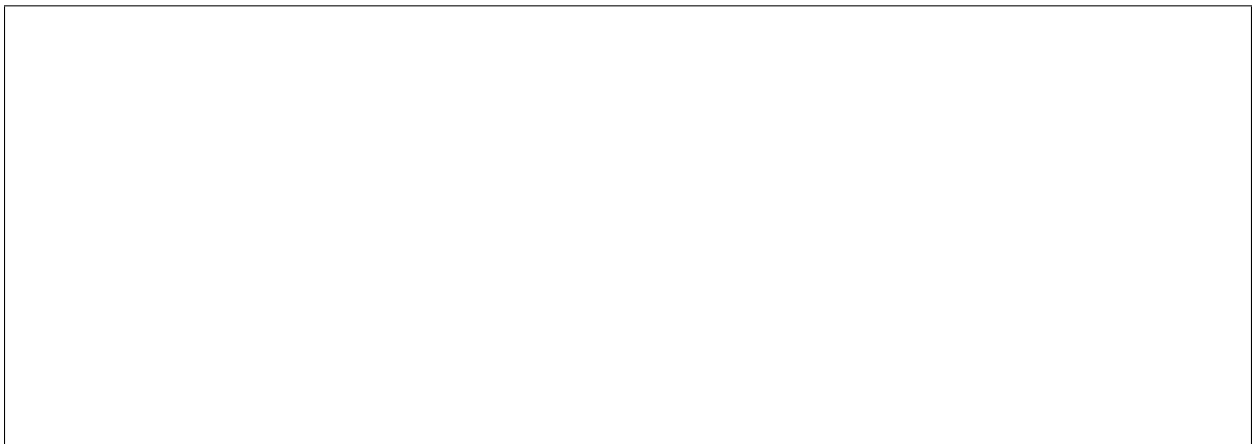
With morphological data, the difficulty is to include invariable characters. There is simply no

straight-forward way of sampling invariable characters with an intensity comparable to that with which variable characters are scored. Again, the solution is to calculate the probability of the unobservable patterns, in this case the all-zero and all-one patterns, and correct for the ascertainment bias when calculating the likelihood value. If the probability is  $p_0$  for the all-zero pattern and  $p_1$  for the all-one pattern, then the corrected likelihood is  $L^* = \prod p_i / (p_0 + p_1)$ .

It is common practice in morphological studies to omit not only invariable characters but also characters that are not parsimony-informative. The latter are characters that have the same length on all trees using the parsimony criterion; typically they are unique features of single terminals in the tree. In principle, this ascertainment bias can also be corrected for, but the correction is more complicated because there are many more character patterns that are left out of the character matrix.

The total probability of all character patterns is summarized in Figure 2.

Figure 2: Cumulative probabilities of different types of character patterns



### 1.3 Rate asymmetry

We can easily accommodate transition rate asymmetry in the M2 model by assuming that the stationary state frequencies of the two states are different, giving us an analogue of the Felsenstein 81 model:

$$Q = \{q_{ij}\} = \begin{Bmatrix} - & \pi_1 \\ \pi_0 & - \end{Bmatrix}$$

There is only one free parameter in this model, since  $\pi_0 + \pi_1 = 1$ . An alternative parameterization of the same model would use the ratio ( $\kappa$ ) of the forward (0 to 1) rate to the backward (1 to 0) rate:

$$Q = \{q_{ij}\} = \begin{Bmatrix} - & \kappa \\ 1 & - \end{Bmatrix}$$

We can go from one representation to the other by noting that  $\kappa = \pi_1/(\pi_0 + \pi_1)$ ,  $\pi_1 = \kappa/(\kappa + 1)$  and  $\pi_0 = 1/(\kappa + 1)$ .

There is one implicit assumption of this type of model that can cause problems with morphological data. By introducing a stationary state frequency parameter for the 0 state and one for the 1 state, we are assuming that state labels are non-arbitrary if there is more than one character. In other words, given that we know the states of one arbitrarily chosen binary character, there must be a way of determining how to apply the state labels to all other characters. For instance, if one state is taken to mean ‘absence’ and the other state ‘presence’ of a particular type of trait, then state labels are non-arbitrary, the assumption is satisfied, and we can reasonably infer the stationary state frequency of the two states. This would be true, for instance, for restriction site data where the two states can be used to denote presence or absence of the each restriction site.

For many morphological characters, however, state labels are truly arbitrary. For instance, how can we translate the presence or absence of wings to black or yellow body color? The solution to this dilemma is to model variation in stationary state frequencies across sites using some suitable distribution, for instance a symmetric beta distribution (Fig. 3). The symmetric beta distribution is like a normal beta distribution where the two parameters (referred to as  $\alpha_0$  and  $\alpha_1$ ) are assumed to be equal and replaced by a single value  $\alpha = \alpha_0 = \alpha_1$ . The symmetric beta distribution can model a wide variety of scenarios, from extremely asymmetric rates ( $\alpha \ll 1$ ) over moderately asymmetric rates ( $\alpha \approx 1$ ) to equal rates ( $\alpha = \infty$ ).

## 2 Quantitative characters [JF: Chapters 23-25]

Many morphological and other characters are best understood as evolving on a continuous scale. Most of the work on stochastic models for quantitative characters has focused on Brownian motion, more precisely a mathematical model that approximates Brownian motion. In this model, a particle takes a large number of steps on an axis, each step being independent of the others and displacing

Figure 3: Symmetric beta distribution with some different values of  $\alpha$ 

the particle by a random amount drawn from a normal distribution with variance  $s^2$ . The model is obtained as the limit condition when the number of steps  $n$  goes to infinity and the variance  $s^2$  goes to zero while the product  $ns^2$  remains constant. If we start a Brownian motion process with value  $x_0$ , and the process accrues the variance  $\sigma^2$  per time unit, then the expected value at time  $t$  is normally distributed with mean  $x_0$  and variance  $\sigma^2 t$ .

Calculating the likelihood under the Brownian motion model is relatively straight-forward except for a problem with the degrees of freedom at the root of the tree. Felsenstein explains the procedure in detail in his Chapter 23.

A disadvantage of the Brownian motion model is that it has no limits (Fig. 4). A character can easily take on an extreme value and then it is unlikely to come back to the ancestral value. This may not be a realistic model for the evolution of morphological characters, which tend to have physical bounds as well as some central tendency in their values. An alternative to Brownian motion that may better reflect these circumstances is the Ornstein-Uhlenbeck (OU) process. The OU process is Brownian motion but with a force continually pulling towards a central point (Fig. 4). If the returning force is pushing a character towards a central point at the rate of  $a$  per unit time, and the process accumulates  $\sigma^2$  variance per unit time, then an OU process starting at  $x_0$  will find itself in position  $x_t$  after  $t$  time units, where  $x_t$  is normally distributed with mean

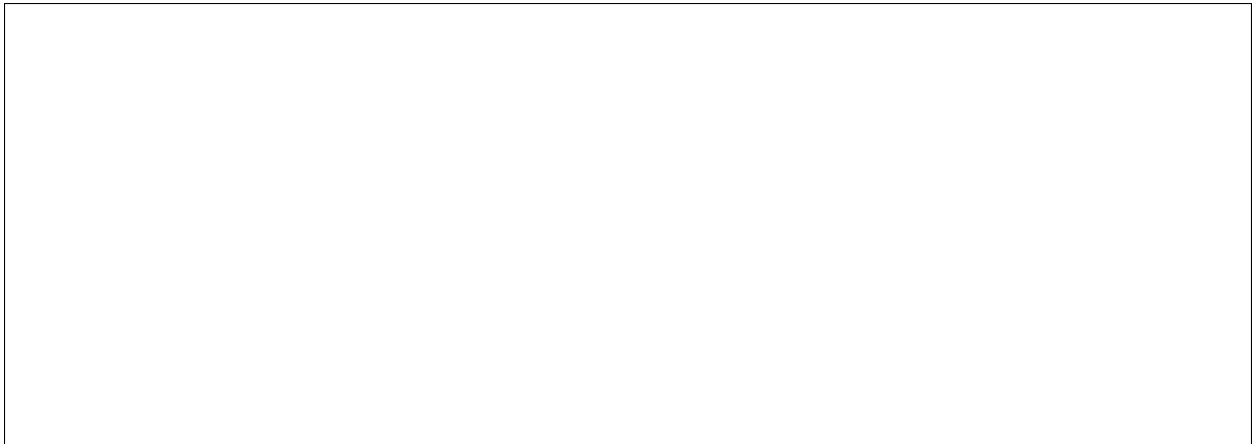
$$\text{Exp}[x_t] = x_0 e^{-at}$$

and variance

$$\text{Var}[x_t] = \frac{(1 - e^{-2at})\sigma^2}{2a}$$

The Ornstein-Uhlenbeck process is difficult to handle in likelihood inference but it has been used to model rate variation over time in a Bayesian context.

Figure 4: Examples of Brownian motion (a) and the Ornstein-Uhlenbeck process (b).



### 3 Study questions

1. What is ascertainment bias and how can it be corrected?
2. How can ordered and unordered discrete morphological characters be modeled?
3. What are the stationary state frequencies of the M3u model and the M3o model?
4. When is it appropriate to assume transition rate asymmetry in a binary character?
5. How can you address transition rate asymmetry when state labels are arbitrary?
6. What is the difference between the Brownian motion model and the Ornstein-Uhlenbeck process?