CALCULATING THE PROBABILITY DISTRIBUTIONS OF ANCESTRAL STATES RECONSTRUCTED BY PARSIMONY ON PHYLOGENETIC TREES

Wayne P. Maddison

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA

Abstract.—Parsimony methods are often used to map a character's evolution on a phylogenetic tree by reconstructing the ancestral states at interior nodes of the tree. Although the statistical behavior of methods for reconstructing trees has been studied using stochastic models of evolution, there is little comparable work on methods for reconstructing character evolution. If a discrete character were to begin with an ancestral state at the root of a phylogenetic tree and evolve with specified probabilities of change on the branches of the tree, then the terminal taxa would receive one of many possible distributions of character states. These states in terminal taxa are the data used by parsimony methods in reconstructing character evolution on trees. Given a stochastic model of character evolution that specifies probabilities of change on the branches of a known phylogenetic tree, it is possible to calculate the probabilities of various ancestral state reconstructions at each node in the tree. Exact calculations for these probabilities are presented here. They can be used, for example, to calculate the probability that ancestral states would be reconstructed accurately. [Phylogeny; character evolution; ancestral states; parsimony; mapping characters; probability distribution; accuracy.]

In phylogenetic biology it is common to trace or map a character's evolution on a phylogenetic tree using parsimony methods (Farris, 1970; Fitch, 1971; Sankoff and Rousseau, 1975). Given that reconstructions of ancestral states in a tree are often used in interpreting the processes of character evolution (e.g., Ridley, 1983; Sillén-Tullberg, 1988; Donoghue, 1989; Brooks and McLennan, 1991; Harvey and Pagel, 1991; Maddison and Maddison, 1992; Swofford and Maddison, 1992; Maddison, 1994), it is important to know how accurate they are. However, little work has focused on the accuracy and statistical behavior of methods for reconstructing character evolution (reviewed by Maddison, 1994), in contrast to the extensive attention paid to the statistical aspects of methods for reconstructing the phylogenetic trees themselves (e.g., Felsenstein, 1988; Hendy and Penny, 1989; Huelsenbeck and Hillis, 1993; Hillis et al., 1994). Empirical studies with known ancestral DNA sequences (Hillis et al., 1992) or known codons of extant species (Goodman, 1981) have indicated reasonable accuracy of parsimony reconstructions of character evolution. Theoretical work involving simula-

tions and probability calculations has shown that accuracy is low when rates of evolution are high (Holmquist, 1979; Saitou, 1989; Tateno, 1990; Maddison and Maddison, 1992:288; Collins et al., 1994; Frumhoff and Reeve, 1994), but none of these studies have presented a general and exact method for calculating error probabilities for reconstructed ancestral states.

In this paper, I present methods to calculate various probabilities that are basic to understanding the accuracy and statistical behavior of parsimony reconstructions of character evolution. The probabilities concern the possible outcomes of a parsimony reconstruction of ancestral states when the character evolves according to a stochastic model of change. The basic assumptions are that (1) the phylogenetic tree is known, (2) the evolution of a discrete (binary or multistate) character is being considered, and (3) a stochastic model of evolutionary change is available for the character. The stochastic model specifies the probabilities of character state change on the branches of the tree. If the character were to begin with a state at the root and evolve along the branches of the tree according to the stochastic model,

then the internal nodes and ultimately the terminal taxa would receive various states with particular probabilities, depending on the model. These states in the terminal taxa could be used by a parsimony method to reconstruct ancestral states at all the internal nodes of the tree. Thus, the stochastic model implies probability distributions both for the true (actual) states at the nodes of the tree and for the parsimony-reconstructed states.

Perhaps the best way to clarify the probabilities sought is via an imaginary simulation in which a character begins with a state at the root of the tree and then evolves stochastically from node to node up the tree until the terminal taxa are reached. After evolution has assigned a character state to a node, the simulation moves up to one of the descendant nodes. Along the intervening branch, the simulation chooses whether the state of the character stays the same or changes to some other state by "rolling dice" according to the probabilities of change specified by the model of evolution for the branch. Depending on the outcome, a state is assigned to the descendant node. This is the actual, or true, state at the node; it is the state that the node actually possessed in the simulated evolution. The simulation continues through the tree until the states in the terminal taxa are obtained. Using these terminal states, a standard parsimony algorithm (e.g., Fitch, 1971; Swofford and Maddison, 1987) is then applied to the tree to yield reconstructed states at each of the internal nodes, modeling a biologist reconstructing ancestral states with observed data. (Sometimes the algorithm will reconstruct a set of equally parsimonious states at a node, but this ambiguity is unimportant to the description of the imaginary simulation.) The simulation can then ask, at a particular node, how does the reconstructed state compare with the true state? If the simulation is repeated many times, one could calculate, for example, the proportion of replicates in which the reconstructed state is 0 among those replicates in which the true state at the node had been 1 (implying an incorrect reconstruc-

tion). These proportions are like the probabilities calculated herein, but I performed no simulations because they are unnecessary; the probabilities can be calculated directly without the need of simulation.

Thus, the probability that is the focus of this paper is the probability that a node is reconstructed by parsimony as having had a particular character state given the node's true state. This conditional probability relating true state to reconstructed state can be used to calculate various aspects of the statistical behavior of parsimony reconstructions, including the accuracy of ancestral state reconstructions. In this paper, I am not describing probabilistic methods to estimate ancestral states, for which maximum likelihood calculations (Felsenstein, 1973, 1981) can be used, nor am I describing methods to calculate the accuracy of a particular ancestral state reconstruction given an observed character state distribution on a tree. Instead, the methods described calculate how parsimony reconstructions are expected to behave statistically, in general, under a specified model of evolutionary change.

BACKGROUND AND ASSUMPTIONS

I assume that the phylogenetic tree for a group of species or genes is known and that it is dichotomously branching. The character is assumed to evolve by a Markov process (Fig. 1a), starting with a state at the root and proceeding node by node upward through the tree toward the tips. The probability that a node N (Fig. 2) receives a particular state ($t_N \equiv$ true state at N) depends only on the state at N's immediate ancestral node, which will be called A, and the conditions along the A to N branch. The evolutionary model specifies the probabilities of various events of character evolution on a branch (state changing, staying the same, etc.) as conditional probabilities. Thus, for the A to N branch, the probability that N has true state t_N given that A had state t_A is written $P(t_N \mid t_A)$. In the simplest case the probability of a change of any sort on any branch would be the same, say p (i.e., $P[t_N \mid t_A] = p$ for all $t_N \neq t_A$).
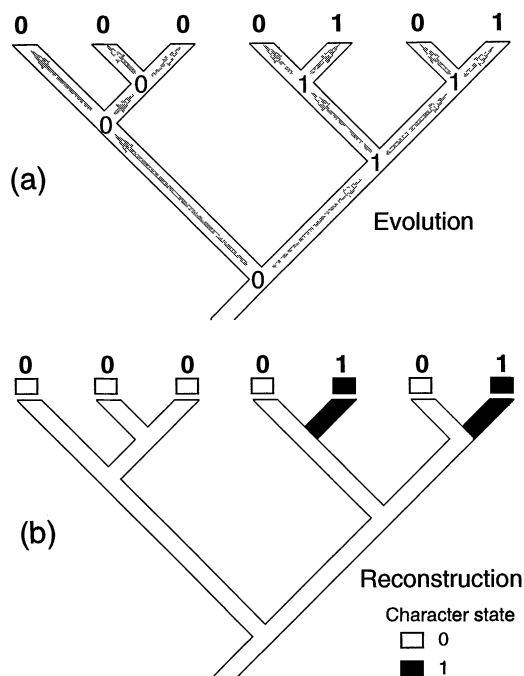
FIGURE 1. Example of a character that evolved along a phylogenetic tree. (a) True history of change, beginning at state 0 in the root of the tree and having three changes, one to state 1 and two reversals to state 0. (b) Ancestral states as reconstructed using parsimony, with state 0 at all the internal branch points and two changes to state 1 (using unordered or ordered parsimony).

FIGURE 2. An arbitrary internal node N in the tree, with its surrounding nodes A (ancestor), S (sister), and L and R (descendants).

Parsimony methods for reconstructing character evolution assign to each node those states that allow for the fewest evolutionary steps throughout the tree (Fig. 1b). Evolutionary steps can be counted in various ways, and I here assume that the ordered (Farris, 1970) or unordered (Fitch, 1971) state methods are used. (A brief review of the algorithm is given here; see Maddison and Maddison [1992] for a more detailed description.) These methods for reconstructing ancestral states first assign a set of states to each node on a pass down through the tree (from the tips to the root; called the "downpass" by Maddison and Maddison, 1992), using the sets previously calculated for the node's two descendants (observed states are used for terminal nodes). Special rules are used to combine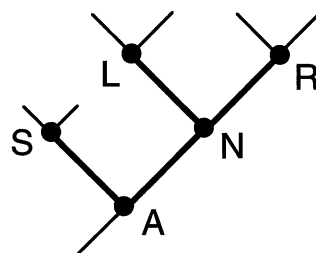 the state sets of the two descendants to yield the state set at the node itself (Farris, 1970; Fitch, 1971; Maddison and Maddison, 1992). On a second pass up through the tree ("uppass"), similar calculations obtain a node's uppass set using its ancestor's uppass set and its sister's downpass set. Finally, the node's final state set (MPR set of Swofford and Maddison, 1987) is calculated using the node's uppass set, its left descendant node's downpass set, and its right descendant's downpass set. The symbols used for the downpass, uppass, and final state sets at node N are D_N, U_N, and F_N, respectively. The final state set contains all those states that can be parsimoniously placed at node N. It might be decisive (e.g., including just a single state) or equivocal (including more than one state as equally parsimonious).

PROBABILITY CALCULATIONS

The probability calculations follow the same pattern as the parsimony calculations themselves, percolating down and back up the tree, calculating and storing information (in this case, probabilities) using the information already calculated at neighboring nodes. The description of the calculations therefore focuses on an arbitrary node N in the tree. If N is an internal node, it has two descendant nodes, L and R, an ancestral node A, and a sister node S (Fig. 2). If N is a terminal node, it lacks descendants.

Given the phylogenetic tree and the model of probabilities of change on each of the branches, the goal is to calculate $P(F_N \mid t_N)$, which is defined as the probability that the final reconstructed state set

at node N is F_N given the true state at N is t_N. The probability calculations first make a pass down and a pass up through the tree. The final probabilities $P(F_N \mid t_N)$ are then calculated at each node using the up and down probabilities.

Probabilities of Downpass Sets

The first task is to calculate $P(D_N \mid t_N)$, which is the probability that a node N will receive downpass set D_N given that its true state is t_N. When N is a terminal node, its assigned downpass set must contain just the observed state, and therefore $P(D_N \mid t_N)$ = 1 if $D_N = \{t_N\}$, $P(D_N \mid t_N) = 0$ otherwise. When N is internal, a particular downpass set D_N can occur if and only if the downpass sets of L and R are appropriate because of the rules used in the parsimony algorithms. Thus one sums over all combinations of D_L and D_R that yield D_N:

$$P(D_N \mid t_N) = \sum_{\substack{D_L D_R \\ \text{yielding } D_N}} P(D_L \ \& \ D_R \mid t_N). \quad (1)$$

For instance, if the character is binary, then the downpass set $D_N = \{0\}$ can be obtained if D_L and D_R are $\{0\}$ and $\{0\}$, respectively, or if they are $\{0\}$ and $\{0, 1\}$, or $\{0, 1\}$ and $\{0\}$. $D_N = \{1\}$ can be obtained from $\{1\}$ and $\{1\}$, or $\{1\}$ and $\{0, 1\}$, or $\{0, 1\}$ and $\{1\}$; $D_N = \{0, 1\}$ can be obtained from $\{0\}$ and $\{1\}$, or $\{1\}$ and $\{0\}$, or $\{0, 1\}$ and $\{0, 1\}$.

Once t_N is given (which it is in the above), the probabilities of downpass sets at L and R are independent, and thus the D_L and D_R probabilities can be separated:

$$P(D_N \mid t_N) = \sum_{\substack{D_L D_R \\ \text{yielding } D_N}} P(D_L \mid t_N) \cdot P(D_R \mid t_N). \quad (2)$$

Consider the term concerning the left descendant, $P(D_L \mid t_N)$. One can add the qualification "and the true state at L is 0 or 1 (or, if there are others, one of the other possible states)":

$$P(D_L \mid t_N)$$
$$= P[D_L \ \& \ (t_L = 0 \text{ or } t_L = 1 \text{ or } t_L \text{ is one of the other possible states}) \mid t_N] \quad (3)$$

because this qualification is necessarily true. But because these possibilities are mutually exclusive, one can rewrite this as a sum over all the possible true states at L:

$$P(D_L \mid t_N) = \sum_{t_L} P(D_L \ \& \ t_L \mid t_N)$$
$$= \sum_{t_L} P(D_L \mid t_L \ \& \ t_N) \cdot P(t_L \mid t_N). \quad (4)$$

Because t_N tells us nothing about the probability of D_L beyond what t_L tells us, this reduces to

$$P(D_L \mid t_N) = \sum_{t_L} P(D_L \mid t_L) \cdot P(t_L \mid t_N). \quad (5)$$

The quantities on the right side are known; $P(D_L \mid t_L)$ is assumed to have already been calculated for L, and $P(t_L \mid t_N)$ is the probability of a t_N to t_L change on L's branch, specified by the model of change. A similar equation can be obtained for $P(D_R \mid t_R)$. Thus, substituting into Equation 2,

$$P(D_N \mid t_N) = \sum_{\substack{D_L D_R \\ \text{yielding } D_N}} \left[\sum_{t_L} P(D_L \mid t_L) \cdot P(t_L \mid t_N) \right.$$
$$\left. \cdot \sum_{t_R} P(D_R \mid t_R) \cdot P(t_R \mid t_N) \right]. \quad (6)$$

These calculations assign the probability of a particular downpass state set for each node. For the root, the downpass state set is the same as its final reconstructed state set, and so these calculations tell us the probability that the root is reconstructed with a particular state set given the true state at the root.

Probabilities of Uppass Sets

After the downpass reaches the root, a pass is made back up through the tree calculating $P(U_N \ \& \ t_N)$, which is the probability that the uppass set at node N is U_N and the true state at N is t_N, and $P(t_N)$, which is the probability that the true state at N is t_N.

$P(t_N)$ is easily calculated using the model of change and the $P(t_A)$ values already calculated at N's ancestral node A by summing over all the possible ways t_N can be achieved at N, namely via change from all possible true ancestral states at A:

$$P(t_N) = \sum_{t_A} P(t_N|t_A) \cdot P(t_A). \qquad (7)$$

$P(U_N \& t_N)$ is calculated in a similar way to the downpass probabilities, except that it uses the probabilities already calculated at N's sister node S and ancestral node A. The derivation of the formula begins by summing over all possible true states at A:

$$P(U_N \& t_N) = \sum_{t_A} [P(t_A) \cdot P(U_N \& t_N|t_A)].$$

Because fixing t_A makes the probabilities of U_N and t_N independent, this is

$$P(U_N \& t_N) = \sum_{t_A} [P(t_A) \cdot P(t_N|t_A) \cdot P(U_N|t_A)].$$

Considering all the ways U_N can be achieved, this becomes

$$P(U_N \& t_N) = \sum_{t_A} \left\{ P(t_A) \cdot P(t_N|t_A) \right.$$
$$\left. \cdot \sum_{\substack{D_S U_A \\ \text{yielding } U_N}} [P(U_A \& D_S|t_A)] \right\}.$$

Finally, one obtains the following by noting that fixing t_A makes U_A and D_S independent and by considering all possible true states at S:

$$P(U_N \& t_N)$$
$$= \sum_{t_A} \left\{ P(t_A) \cdot P(t_N|t_A) \right.$$
$$\cdot \sum_{\substack{D_S U_A \\ \text{yielding } U_N}} \left[P(U_A|t_A) \right.$$
$$\left. \left. \cdot \sum_{t_S} P(D_S|t_S) \cdot P(t_S|t_A) \right] \right\}.$$
$$(8)$$

This calculation considers all possible true states at A and for each considers all possible D_S and U_A that could yield U_N and all possible true states at S. In Equation 8, all quantities are either known from the model or previously calculated. Even though $P(U_A|t_A)$ was not directly calculated pre-

viously, $P(U_A|t_A) = P(U_A \& t_A)/P(t_A)$, both of whose terms were calculated previously.

There is a complication when N's ancestor is the root because the root has no uppass set. In this case, N's uppass set comes directly from its sister's downpass set (D_S), and so there is only one way to achieve a particular U_N, namely, when $D_S = U_N$. Thus,

$$P(U_N \& t_N)$$
$$= P(t_N \& [D_S = U_N]|t_{\text{Root}})$$
$$= P(t_N|t_{\text{Root}}) \cdot P([D_S = U_N]|t_{\text{Root}})$$
$$= P(t_N|t_{\text{Root}}) \cdot \sum_{t_S} P([D_S = U_N]|t_S) \cdot P(t_S|t_{\text{Root}}).$$
$$(9)$$

Probabilities of Final State Sets (MPR Sets)

The probabilities calculated above are used to calculate the probabilities of final state sets, $P(F_N|t_N)$. Because the probabilities have already been calculated for the downpass sets for N's two descendants and for the uppass sets for N, one needs only sum the probabilities of all the ways the final set of F_N can be obtained at N, as follows:

$$P(F_N|t_N)$$
$$= \sum_{\substack{D_L D_R U_N \\ \text{yielding } D_N}} [P(U_N|t_N) \cdot P(D_L|t_N) \cdot P(D_R|t_N)]$$
$$= \sum_{\substack{D_L D_R U_N \\ \text{yielding } D_N}} \left[P(U_N|t_N) \cdot \sum_{t_L} P(D_L|t_L) \cdot P(t_L|t_N) \right.$$
$$\left. \cdot \sum_{t_R} P(D_R|t_R) \cdot P(t_R|t_N) \right]. \qquad (10)$$

This formula considers all possible combinations of downpass sets at L and R and uppass sets at N that would yield F_N, and for each it considers the possible true states at L and R.

EXAMPLES

Two examples are given to illustrate these calculations and how they might be applied to the question of accuracy. Figure 3 shows a worked example for a binary

(a) Probabilities of downpass sets

$D = \{0\}$ $\{1\}$ $\{0,1\}$

| $t=0$ | 0.810 | 0.010 | 0.180 |
| 1 | 0.010 | 0.810 | 0.180 |

$D = \{0\}$ $\{1\}$ $\{0,1\}$

| $t=0$ | 0.786 | 0.042 | 0.172 |

$D = \{0\}$ $\{1\}$ $\{0,1\}$

| $t=0$ | 0.810 | 0.010 | 0.180 |
| 1 | 0.010 | 0.810 | 0.180 |

$D = \{0\}$ $\{1\}$ $\{0,1\}$

| $t=0$ | 0.819 | 0.027 | 0.154 |
| 1 | 0.027 | 0.819 | 0.154 |

(b) Probabilities of uppass sets

$U = \{0\}$ $\{1\}$ $\{0,1\}$

| $t=0$ | 0.666 | 0.096 | 0.139 |
| 1 | 0.074 | 0.011 | 0.015 |

$U = \{0\}$ $\{1\}$ $\{0,1\}$

| $t=0$ | 0.664 | 0.024 | 0.131 |
| 1 | 0.082 | 0.024 | 0.074 |

$U = \{0\}$ $\{1\}$ $\{0,1\}$

| $t=0$ | 0.657 | 0.081 | 0.162 |
| 1 | 0.073 | 0.009 | 0.018 |

(c) Probabilities of final (MPR) sets

$F = \{0\}$ $\{1\}$ $\{0,1\}$

| $t=0$ | 0.943 | 0.029 | 0.028 |
| 1 | 0.143 | 0.829 | 0.028 |

$F = \{0\}$ $\{1\}$ $\{0,1\}$

| $t=0$ | 0.786 | 0.043 | 0.172 |

$F = \{0\}$ $\{1\}$ $\{0,1\}$

| $t=0$ | 0.956 | 0.015 | 0.029 |
| 1 | 0.092 | 0.834 | 0.074 |

$F = \{0\}$ $\{1\}$ $\{0,1\}$

| $t=0$ | 0.917 | 0.028 | 0.055 |
| 1 | 0.138 | 0.715 | 0.148 |

(d) Probabilities of correct reconstruction
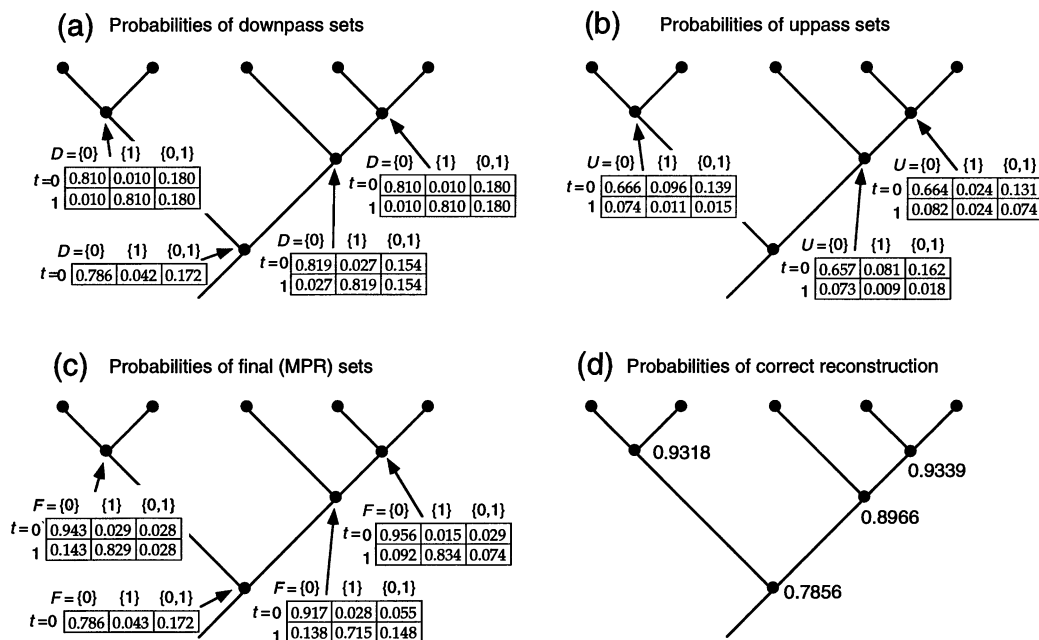
0.9318

0.9339

0.8966

0.7856

FIGURE 3. A worked example showing calculated values for probabilities of state sets in a parsimony algorithm. The character is binary; the probability of a 0 to 1 or 1 to 0 change is 0.1 on each branch. (a) The probabilities calculated for the downpass sets. At each node are the probabilities that the downpass set will be D given a true state of t for the three possible values of D ($\{0\}$, $\{1\}$, $\{0, 1\}$) and the two of t (0, 1). For terminal nodes, $P(D = \{0\} \mid t = 0) = P(D = \{1\} \mid t = 1) = 1.0$; the four other probabilities are 0.0. (b) The probabilities for uppass sets. (c) The probabilities for final (MPR) sets. (d) The probabilities of correct reconstruction.

character, with the assumption of a 0.1 probability of change (state 0 to 1 or state 1 to 0) per branch. Figure 3a is the tree with the values of $P(D_N \mid t_N)$ for each of the six combinations of values of D_N and t_N at each of the nodes (except the root, for which the true state is assumed to be 0). Figure 3b is the tree with values of $P(U_N \mid t_N)$, and Figure 3c is the tree with values of $P(F_N \mid t_N)$. Two of the probabilities listed for the final MPR sets, $P(F_N = \{0\} \mid t_N = 0)$ and $P(F_N = \{1\} \mid t_N = 1)$, refer to a correct reconstruction (the reconstructed state set contains only the true state); two of the probabilities, $P(F_N = \{1\} \mid t_N = 0)$ and $P(F_N = \{0\} \mid t_N = 1)$, refer to an incorrect reconstruction, and the remaining two probabilities refer to ambiguous reconstructions. Thus, the probability of a correct reconstruction (regardless of the true state) can be obtained from the following:

$$P(F_N = \{0\} \mid t_N = 0) \cdot P(t_N = 0) + P(F_N = \{1\} \mid t_N = 1) \cdot P(t_N = 1).$$

Figure 3d shows the probabilities of accurate reconstruction. The probability of accuracy does not depend on which state is assumed to be the true ancestral state at the root because the model of change is symmetrical (the probabilities of gains and losses are assumed to be the same).

Figure 4 shows calculations from a second example. The probabilities of correct reconstruction were calculated for nodes in a fully symmetrical tree of 128 taxa. All branches had the same chance of change, which varied from 0 to 0.5. As the probability of change rises, the accuracy drops, especially for the root node, which suffers only a 50% chance of correct reconstruction when the probability of change is 0.15/branch (the chance of correct recon-
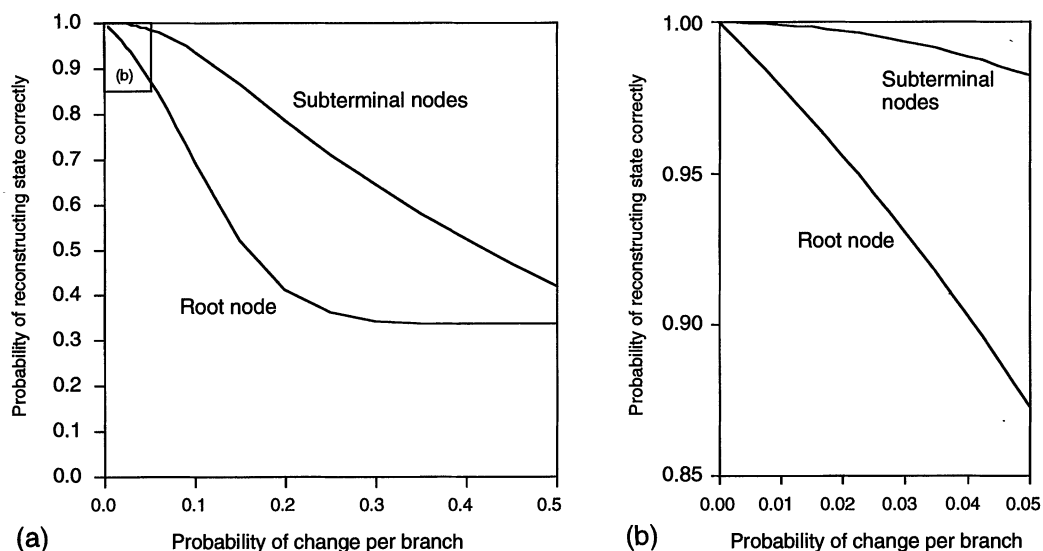
FIGURE 4. Probability of correctly reconstructing the ancestral state in a binary character in a fully symmetrical tree of 128 terminal taxa, for varying rates of evolution. Probability of change is same on every branch. The upper curve shows probabilities for a subterminal node (one with only two descendant terminal taxa), and the lower curve shows probabilities for the root node of the entire tree. (a) Curve for probabilities of change per branch ranging from 0.0 to 0.5. (b) Curve for probabilities of change per branch ranging from 0.00 to 0.05.

struction can drop below 0.5 because ambiguous reconstructions are counted as errors in this example). A character evolving with a 0.15 probability would likely be noticed for its high rate, however, because it would have on average over 30 actual changes on the tree. At lower rates of change the probabilities of correct reconstruction are high, as shown in Figure 4b.

DISCUSSION

In the above examples, the probability calculations were applied to determine the expected accuracy of the reconstructed state at a node. There are, however, many other potential applications of these calculations. Using the probabilities calculated for downpass sets, uppass sets, and MPR sets, as well as related probabilities calculated in a similar style, it is possible to determine, for instance, the probability that changes along a branch are reconstructed correctly, that a certain number of changes is reconstructed along a branch, and that a particular number of steps is counted for a character over the tree. Extensions of the probability calculations to

these diverse applications will be presented and explored in subsequent papers (Maddison and Maddison, in prep.).

Calculations in the literature most similar to these are those of Holmquist (1979), Frohlich (1987), and Maddison and Slatkin (1991). Holmquist sought the probability that an ancestral node would have its state reconstructed correctly given a known tree, given an expected number of changes on each branch, and (unlike the case here) given a set of observed states in the terminal taxa. His probability calculations were approximate because they did not fully consider the effect that parts of the tree below and beside a node have on a node's reconstructed state (i.e., his w's were approximations of this effect). The calculations of Frohlich (1987) and Maddison and Slatkin (1991) sought the probability of reconstructing various states but in circumstances rather different from those described here. Frohlich calculated the probability that a particular state would be reconstructed at the root of a tree, given that the states in the observed taxa are fixed and the tree is chosen randomly.

Maddison and Slatkin extended this by considering also randomly chosen observed character states and a fixed tree. Neither of these results consider a model of evolution on a known tree (although the Maddison and Slatkin case is equivalent to one extreme model: fully saturated evolutionary change on each branch).

For binary characters, the calculations described are fast and use little memory, storing only three 3×2 arrays of real variables at each node (one for probabilities of up, down, and final sets, which can take on one of three values, $\{0\}$, $\{1\}$, or $\{0, 1\}$, and be conditional upon true states of either 0 or 1). For four-state characters, the calculations are considerably more tedious, requiring consideration of the 15 possible state sets and four possible true states, but they would be straightforward to program.

ACKNOWLEDGMENTS

REFERENCES

BROOKS, D. R., AND D. A. MCLENNAN. 1991. Phylogeny, ecology and behavior: A research program in comparative biology. Univ. Chicago Press, Chicago.

COLLINS, T. M., P. H. WIMBERGER, AND G. J. P. NAYLOR. 1994. Compositional bias, character-state bias, and character-state reconstruction using parsimony. Syst. Biol. 43:482–496.

DONOGHUE, M. J. 1989. Phylogenies and the analysis of evolutionary sequences, with examples from seed plants. Evolution 43:1137–1156.

FARRIS, J. S. 1970. Methods for computing Wagner trees. Syst. Zool. 19:83–92.

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

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17:368–376.

FELSENSTEIN, J. 1988. Phylogenies from molecular sequences: Inference and reliability. Annu. Rev. Genet. 22:521–565.

FITCH, W. M. 1971. Toward defining the course of evolution: Minimal change for a specific tree topology. Syst. Zool. 20:406–416.

FROHLICH, M. 1987. Common-is-primitive: A partial validation by tree counting. Syst. Bot. 12:217–237.

FRUMHOFF, P. C., AND H. K. REEVE. 1994. Using phylogenies to test hypotheses of adaptation: A critique of some current proposals. Evolution 48:172–180.

GOODMAN, M. 1981. Globin evolution was apparently very rapid in early vertebrates: A reasonable case against the rate-constancy hypothesis. J. Mol. Evol. 17:114–120.

HARVEY, P. H., AND M. D. PAGEL. 1991. The comparative method in evolutionary biology. Oxford Univ. Press, Oxford, England.

HENDY, M. D., AND D. PENNY. 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38:297–309.

HILLIS, D. M., J. J. BULL, M. E. WHITE, M. R. BADGETT, AND I. J. MOLINEUX. 1992. Experimental phylogenetics: Generation of a known phylogeny. Science 255:589–592.

HILLIS, D. M., J. P. HUELSENBECK, AND C. W. CUNNINGHAM. 1994. Application and accuracy of molecular phylogenies. Science 264:671–677.

HOLMQUIST, R. 1979. The method of parsimony: An experimental test and theoretical analysis of the adequacy of molecular restoration studies. J. Mol. Biol. 135:939–958.

HUELSENBECK, J. P., AND D. M. HILLIS. 1993. Success of phylogenetic methods in the four-taxon case. Syst. Biol. 42:247–264.

MADDISON, D. R. 1994. Phylogenetic methods for inferring the evolutionary history and processes of change in discretely valued characters. Annu. Rev. Entomol. 39:267–292.

MADDISON, W. P., AND D. R. MADDISON. 1992. MacClade version 3: Analysis of phylogeny and character evolution. Sinauer, Sunderland, Massachusetts.

MADDISON, W. P., AND M. SLATKIN. 1991. Null models for the number of evolutionary steps in a character on a phylogenetic tree. Evolution 45:1184–1197.

RIDLEY, M. 1983. The explanation of organic diversity. Oxford Univ. Press, Oxford, England.

SAITOU, N. 1989. A theoretical study of the underestimation of branch lengths by the maximum parsimony principle. Syst. Zool. 38:1–6.

SANKOFF, D., AND P. ROUSSEAU. 1975. Locating the vertices of a Steiner tree in arbitrary space. Math. Progr. 9:240–246.

SILLÉN-TULLBERG, B. 1988. Evolution of gregariousness in aposematic butterfly larvae: A phylogenetic analysis. Evolution 42:293–305.

SWOFFORD, D. L., AND W. P. MADDISON. 1987. Reconstructing ancestral character states under Wagner parsimony. Math. Biosci. 87:199–229.

SWOFFORD, D. L., AND W. P. MADDISON. 1992. Parsimony, character-state reconstructions, and evolutionary inferences. Pages 187–223 in Systematics, historical ecology, and North American freshwater fishes (R. L. Mayden, ed.). Stanford Univ. Press, Stanford, California.

TATENO, Y. 1990. A method for molecular phylogeny construction by direct use of nucleotide sequence data. J. Mol. Evol. 30:85–93.