The Role of Subjectivity in Reconstructing Ancestral Character States: A Bayesian Approach to Unknown Rates, States, and Transformation Asymmetries

TED R. SCHULTZ[1,3] AND GARY A. CHURCHILL[2]

[1] *Department of Entomology, MRC 165, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20560, USA; E-mail: schultz@onyx.si.edu*

[2] *The Jackson Laboratory, Bar Harbor, Maine 04609; Department of Plant Breeding and Biometry, Warren Hall, Cornell University, Ithaca, New York 14853, USA*

The synthesis of data from ecology and systematics makes possible, at least in principle, the reconstruction of features of extinct ancestors. The reliability of such reconstructions, however, is a matter of considerable concern. For example, the commonly used method of parsimony character-state optimization (Farris, 1970; Hartigan, 1973; Swofford and Maddison, 1987) often (but not always) assigns unambiguous character states to internal (ancestral) nodes of fully resolved trees, but it is intuitively clear that some such assignments are more reliable than others (Fig. 1). Various probabilistic methods (including maximum-likelihood approaches) for determining the relative reliability of ancestral state assignments have been proposed, all depending on various model assumptions about character evolution (e.g., Frumhoff and Reeve, 1994; Maddison, 1995; Schluter, 1995; Schultz et al., 1996; Schluter et al., 1997; reviewed in Cunningham et al., 1998). These general models require plausible values for the parameters describing the probability of change between character states, and these values can either be estimated from the data or derived from independent sources. For example, maximum-likelihood reconstructions of nucleotide evolution (which are not necessarily explicitly concerned with ancestral states) usually derive the probabilities of change between the four nucleotides from the relative base frequencies across all of the sequences under consideration (Felsenstein, 1981; reviewed in Swofford et al.,
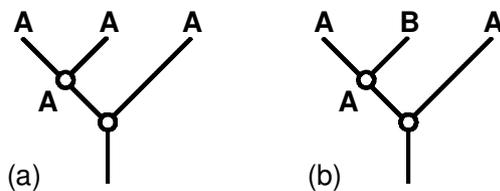

FIGURE 1. Parsimony character-state optimization reconstructs the state in the common ancestor of the sister-species pair as state A in both trees (Fig. 1a and 1b). This method provides no means for judging relative reliability, however, even though the reconstruction in 1a seems intuitively more reliable than the one in 1b.

1996). In contrast, reconstructions of amino acid evolution usually derive the probabilities of amino acid transitions from generalized substitution matrices based on empirical observations across a wide variety of proteins (Dayhoff et al., 1978; Jones et al., 1992).

Obviously, disagreement over choice of models can lead to disagreement about the reliability of ancestral states. As we will demonstrate below, however, even when researchers agree about the general model, if they disagree about the relative plausibility of certain parameter values over others, then they will sometimes disagree about what the data have to say about ancestral states. For example, in maximum likelihood approaches to ancestral state reconstruction (Schluter, 1995; Yang et al., 1995; Koshi and Goldstein, 1996; Schluter et al., 1997; Zhang and Kumar, 1997; Zhang and Nei, 1997), the values of unknown parameters (e.g., rate, branch lengths, state-transition probabilities) are estimated from the data, assumed to be true, and then used to infer the probability

---

[3] Author for correspondence.

of the ancestral state. Because this strategy treats a maximum likelihood–estimated parameter as a known quantity and uses it to calculate a derivative quantity (the probability of the ancestral state), it fails to account for uncertainty in the former and leads to overestimated confidence in the latter. In fact, different researchers can continue to disagree about the probability of the ancestral state simply by disagreeing about the (uncertain) values of the unknown parameters. As we will argue below, this disagreement may persist even if the uncertainties surrounding these values are quantified under a modified approach using, for example, confidence limits. We will further argue that this sort of subjective disagreement about parameter values (rather than about the data or the general model) explains the majority of current disagreement about ancestral state reconstruction. Finally, we will propose a fully Bayesian statistical method for calculating ancestral state probability that properly accounts for all aspects of uncertainty (within the context of the assumed model) and thus more accurately reflects our state of knowledge regarding an ancestral state.

The Data and Model

The following discussion focuses on the mapping of a single character onto a given phylogeny. The character is here considered to be ecological or behavioral, since such characters are most frequently associated with mapping exercises, but the discussion applies equally to any type of atomistic binary or (with modifications) multistate character, e.g., single nucleotides or amino acids. Topology, branch lengths, and the distribution of character states in terminal taxa are all assumed to be "known," although in fact significant errors usually accompany these estimates and further reduce the reliability of ancestral state reconstruction. The additional assumptions that the character evolves at an average rate and with a constant state-transition matrix across the phylogeny could, if violated, also alter our probability estimates. Such departures from our simplifying assumptions, however, will serve only to amplify the effects of the sub-

jective disagreement that we will demonstrate below.

Suppose we are confronted with the pattern in Figure 2. Some number of species N simultaneously diverge from a common ancestor at an unknown time in the past. In the present we observe that all N species are fixed for state A of a two-state character of interest. Suppose further that we know that this character has been evolving according to a Markov chain model, and that the rate of evolution has been, on average, the same in all N lineages. We wish to answer
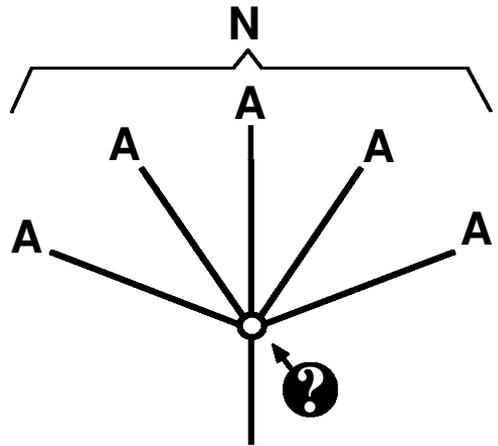

FIGURE 2. A polytomous phylogenetic tree ("bush") in which N species lineages arise simultaneously from a common ancestor and in which all are subsequently observed to be fixed for character state A.

the question: What was the state in the common ancestor; i.e., what is the probability that the ancestral state is the same as the descendant state? For any given time elapsed since divergence (t) and given rate matrix, we can easily calculate the probability of a given pattern in the N descendants, including the observed pattern $\{A, A, A, \ldots, A, A\}$. If we assume that a priori all possible values of t are equally plausible, then we will find upon examining the data that smaller values of t confer greater likelihood on the observed outcome than do greater values of t; in fact, the likelihood will always peak at

$t = 0$. If, additionally, we accept that some non-zero amount of time has passed since divergence, i.e., that $t > 0$, and thus allow for the possibility of multiple B $\rightarrow$ A changes, then, for most N, we can identify a maximum value for t above which the probability of the observed pattern drops below some critical value (e.g., 0.05). Put simply, in the extreme case in which there are a large number of descendant species and in which transitions A $\rightarrow$ B and B $\rightarrow$ A are equiprobable (symmetrical exchange), when enough time t has passed, we expect to observe a roughly 50% mixture of As and Bs in the N descendant species rather than a uniform fixation of one or the other state. With this approach and a given rate matrix, Schultz et al. (1996) found the probability to be nearly 1.00 that the ancestral state is the same as the descendant state when $N = 8$ (i.e., the reconstruction under both parsimony and maximum likelihood methods). Using the same model and parameter values, but assuming that higher rates of change were plausible, Frumhoff and Reeve (1994:Fig. 2) found this probability to be as low as nearly 0.5 (equivalent to chance expectation).

Subjective Disagreement

The difference between these conclusions occurs because Schultz et al. (1995) allowed information in the data to set a limit on the plausible values of the rate of change. This confidence-level approach constitutes a statistical test in which the null hypothesis can be stated as follows: The common ancestor was in state B, and t is sufficiently large that multiple state changes have occurred, at least one per branch. The alternative hypothesis is this: The common ancestor was in state A, and t is sufficiently small that change on any branch is improbable. For small values of N, it is not possible to distinguish between these two hypotheses at a statistically significant level; for larger values, however, the expectation associated with the null hypothesis (that of a mixture of states in the descendants, as discussed above) is sufficiently different from that of the alternative hypothesis (fixation in the descendants for the ancestral state), and the

data can be used to generate a significant P-value. As statisticians frequently emphasize, a P-value is not the probability that the null hypothesis is true, however desirable such a statement might be; instead, the proper interpretation of a statistically significant P-value "is logically that of the simple disjunction: Either an exceptionally rare chance has occurred, or the theory of random distribution [i.e., the null hypothesis] is not true" (Fisher, 1956:39).

Of these choices, it is clear that, based on prior knowledge, some biologists favor the former (null) hypothesis: that high rates obtain, and that a "rare chance" (i.e., improbable event) of fixation for state A in N descendant species has occurred. Referring to such prior knowledge, Frumhoff and Reeve (1994:174) write: "[G]iven the well-documented potential for within-lineage character evolution, parsimonious inferences of the ancestral states of a given character may often not be robust and should be interpreted with caution." That they are willing a priori to give greater credence to high rates (corresponding to high values of t in their discussion and to large θ in the discussion below) is clear: "If there are non-zero probabilities of within-lineage transitions from A to B and B to A and there has been sufficient time for even a few transitions to occur, then the probabilities of seeing A or B *currently* in these species can be essentially independent of their starting states, that is, the character state of their most recent common ancestor."

In contrast, Hennigian cladists are inclined to give greater a priori credence to low rates. Citing "Hennig's Auxiliary Principle" (Hennig, 1966:121–122), Brooks and McLennan (1991:35) emphasize that a similarity between two characters should, whenever possible, be interpreted as homology (translating into low rates in the present context), stating: "Never assume convergent or parallel evolution; always assume homology in the absence of contrary evidence." This principle is applied in the inference of phylogenies from data sets consisting of multiple characters under the parsimony criterion; it is also commonly used in the parsimonious mapping of single characters onto

given phylogenies (e.g., Brooks and McLennan, 1991).

This fundamental disagreement over what evolutionary rates are a priori most reasonable lies in a region outside the bounds of classical statistics, because it implies a difference of opinion on the part of individual researchers over subjective parameters that are independent of the data or the model. This situation is most naturally accommodated by one of the approaches of Bayesian statistics, in which the common frequentist interpretation of probability is replaced by the epistemic interpretation that probability numerically expresses degrees of uncertainty and belief in light of data. As stated by Edwards et al. (1963:194), "Probability is orderly opinion, and ... inference from data is nothing other than the revision of such opinion in the light of relevant new information." This interpretation is especially appropriate to phylogenetics, addressing the complaint that frequentist statistics cannot be applied to the single-case, historical particulars that are the primary focus of phylogenetic inquiry (Wenzel and Carpenter, 1994; Wenzel, 1997; see also Sokal and Rohlf, 1995:70).

BAYESIAN QUANTIFICATION OF SUBJECTIVE DISAGREEMENT

If we interpret probability as degree of belief, then we can use Bayes' theorem (Bayes, 1763) to calculate the relationship between prior belief and posterior belief. In general form, Bayes' theorem may be written as

$$P(\mathrm{H}_i|\mathrm{data}) = \frac{P(\mathrm{data}|\mathrm{H}_i) \cdot P(\mathrm{H}_i)}{P(\mathrm{data})}$$

Posterior probability, $P(\mathrm{H}_i|\mathrm{data})$, is the degree of belief a given researcher has in a hypothesis after viewing the data; i.e., it is the probability of the hypothesis given the data and given the information available prior to viewing the data. As expressed in Bayes' theorem, this posterior probability is equal to the product of the likelihood, $P(\mathrm{data}|\mathrm{H}_i)$, and the prior probability, $P(\mathrm{H}_i)$, as a proportion of the total probability accounted for by all the competing hypotheses, $P(\mathrm{data})$. The likelihood is just the probability conferred

on the data under the hypothesis without regard to prior information or belief. The prior probability is the degree of belief a given researcher has in a hypothesis, H_i, prior to viewing the data, given just the prior information available. This degree of belief is subjective in that it may vary across researchers, since each may draw from a different information background. Bayes' formula, then, allows researchers to quantify their individual preconceptions about the way things will turn out and to incorporate this information into their final conclusions. The posterior probability of a hypothesis will tend to be high if either the data are more strongly predicted by that hypothesis than by other, competing hypotheses (i.e., if its likelihood is high), or, all other factors being equal, if the prior probability of that hypothesis is higher than that of other, competing hypotheses. The degree to which likelihood and prior belief are balanced in posterior belief is determined by the strength of conviction of the prior belief.

Practically speaking, Bayes' theorem can be used in two ways: First, it can be used to determine how decisive the data are in altering a range of prior belief about character evolution. In fact, it has been shown for many classes of problems that when data are decisive, Bayes' theorem funnels a wide range of prior belief into a quite narrow posterior (Edwards et al., 1963; Howson and Urbach, 1989). Second, Bayes' theorem can be used to quantify the difference in prior belief necessary to explain the situation in which researchers agree about the model and the data, but disagree about which hypothesis is favored thereby.

As introduced in Schultz et al. (1995) and outlined in the *Appendix*, we will reparameterize the general model of Tavaré (1986) to express character evolution in a binary (two-state) character with two parameters, which together summarize the rate matrix: (a) the rate ("speed") of the process (θ), i.e., the expected number of transitions per unit time; and (b) the symmetry of the process (ω), i.e., the extent to which transitions favor one state (A) over the other (B). We will use Bayes' theorem to express the posterior probability of rate (θ) and symmetry (ω) as a

function of the likelihood of the data and of the joint prior probability of rate and symmetry:

$$P(\theta, \omega | \text{data}) \propto P(\text{data} | \theta, \omega) \cdot P(\theta, \omega)$$

We will use this quantity, in turn, to calculate the posterior probability of the state in the common ancestor of the ingroup. Prior beliefs about rate and symmetry will be represented as gamma and beta distributions, respectively, i.e., as continuous probability distributions. This choice of parametric models is somewhat arbitrary but allows us to study a range of prior means and variances (corresponding to "best guesses" and "degrees of conviction," respectively). Note that our use of the gamma distribution here has absolutely no connection with its use in modeling rate heterogeneity across nucleotide sites in maximum likelihood molecular phylogenetic analyses.

THE BAYESIAN APPROACH APPLIED

Consider the case of two researchers, both interpreting the same data. These data consist of a distribution of states of a two-state character across terminal taxa on an unrooted phylogenetic tree (Fig. 3), which is composed of an ingroup of N species related by a bush, and an outgroup composed of two species. The researchers agree about the accuracy of the character-state assignments, the tree topology, and the tree branch lengths. They also entirely agree in their prior belief about the symmetry of the process of character evolution (ω), described as a beta distribution with B(4, 4); i.e., they share a belief, held with a moderate level of conviction, that A $\rightarrow$ B and B $\rightarrow$ A transitions are equally likely.

The two researchers disagree, however, in their prior beliefs about the rate of evolution (θ) in the character. First we will consider the case in which the prior beliefs about rate of both researchers are loosely distributed: Researcher 1's prior belief centers on a mean of 1/8 change across the entire tree (i.e., from ingroup node to outgroup node), but she is open-minded about other possible values; this researcher's prior belief is modeled as a
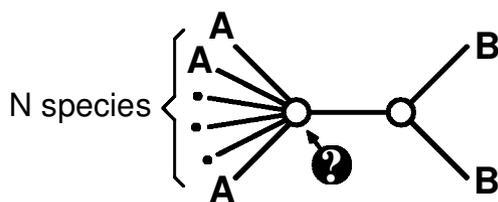

FIGURE 3. The basic unrooted tree topology used to generate the plots in Figure 5. The ingroup (on left) has variable number N species, all in state A; the outgroup (on right) contains two species, both in state B. All branches are of equal length.

gamma distribution with $\Gamma(1/8, 1)$ (Fig. 4a). Researcher 2's prior belief centers on a mean of two changes across the tree, but he is likewise open-minded about other possible rates; this researcher's prior belief is modeled as a gamma distribution with $\Gamma(2, 1)$ (Fig. 4a).

The data consist of a variable number of ingroup species N, each found to be fixed for the same state, state A. When the ingroup contains only two species, Researcher 1 adjusts her belief about low rates slightly upwards, whereas Researcher 2 adjusts his prior belief in high rates downwards (Fig. 5a). As the ingroup size N increases, the evidence obviously swings in favor of low rates and the posterior beliefs of both researchers start moving toward lower rates. These posteriors converge with increasing N but, even when there are 20 ingroup species, the two researchers do not entirely agree about rate. This lack of complete agreement indicates that there is a limited amount of information in the data with regard to rate of change.

Now consider the same two researchers, again agreeing in prior belief about symmetry (ω) and again with the same differing mean prior beliefs about rate (θ), but now with much greater degrees of conviction about those rates, i.e., with tight probability distributions around those prior beliefs (Researcher 1: $\Gamma(10, 80)$, Researcher 2: $\Gamma(100, 50)$; Fig. 4b). In this case the same data produce strikingly different outcomes, with the result that increasing N within a reasonable range does not result in converging posterior belief (Fig. 5b). When there are
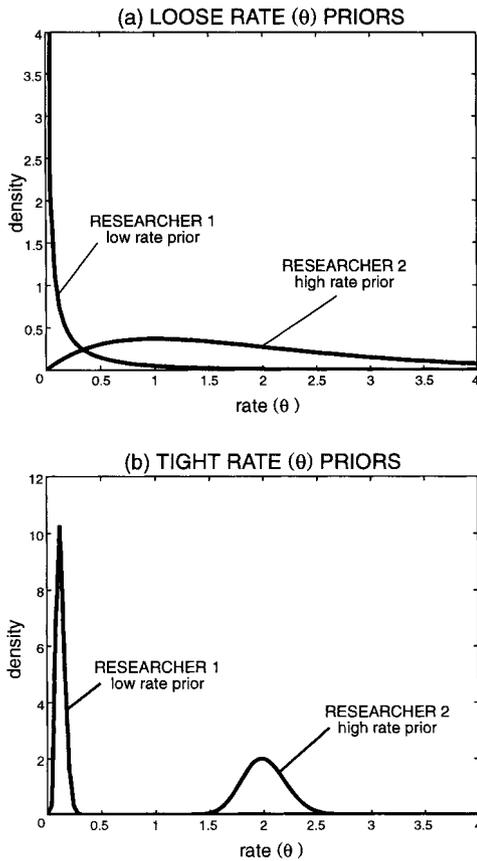
FIGURE 4. Probability distributions of the prior beliefs of two researchers about average rate (θ). Researcher 1 has a prior belief that rates are slow, with a mean value of $1/8$ change across the greatest span of the tree (from ingroup node to outgroup node) in Figure 3. Researcher 2 has a prior belief that rates are fast, with a mean value of two changes across the span of the tree in Figure 3. (a) The prior beliefs of both researchers are loosely held; i.e., both are willing to consider a wide range of values around their mean beliefs. Researcher 1's prior belief is gamma-distributed as $\Gamma(1/8, 1)$; Researcher 2's prior belief is gamma-distributed as $\Gamma(2, 1)$. (b) The prior beliefs of both researchers are tightly held; i.e., they are willing to consider only a narrow range of values around their mean beliefs. Researcher 1's prior belief is gamma-distributed as $\Gamma(10, 80)$; Researcher 2's prior belief is gamma-distributed as $\Gamma(100, 50)$.

20 ingroup species all fixed for the same state, Researcher 1 remains convinced that rate is low and change is symmetrical; in contrast, Researcher 2 remains convinced that rate is high, but alters his belief about

symmetry, with a posterior belief in asymmetrical exchange favoring state A. Thus, even in the face of data that were decisive under the confidence-level approach of Schultz et al. (1995), Researcher 2 is able to retain a belief in high rates simply by altering his moderately held belief about the symmetry of the process.

We believe that this Bayesian interpretation of the disagreement represented by Researchers 1 and 2, with two different prior subjective beliefs about rate of evolution, addresses the current state of affairs in historical ecology more accurately than did the confidence-level approach of Schultz et al. (1995). In the remainder of this paper we therefore adopt a fully Bayesian approach for calculating probabilities of ancestral states that allows for the reality of a range of prior subjective belief across researchers. This approach takes as its input a continuum of prior belief about the rate (θ) and symmetry (ω) of the process of character evolution and, for a given set of data (tree topology, branch lengths, and character-state distribution), generates a range of posterior probabilities for the ingroup ancestral state. The only previous applications of Bayesian statistics to ancestral state reconstruction (Yang et al., 1995; Koshi and Goldstein, 1996; Zhang and Kumar, 1997; Zhang and Nei, 1997) do not treat rate or symmetry as subjective parameters; instead, these parameters are either estimated from the data or assigned values derived from frequentist models (e.g., the empirically derived Jones et al. [1992] amino acid substitution matrix). The availability of frequency data about nucleotides and amino acids, along with the use of these data to construct plausible estimates for the rate and symmetry of a focal character, gives these previous studies a distinct advantage over the single-character situation explored here, for which no obvious independent frequency-based information exists. But even when seemingly plausible Bayesian prior probabilities are based on independent frequency data, there is an uncertainty associated with such parameter estimates that is either not addressed or addressed only indirectly (using, e.g., sensitivity analyses or simulations) in these
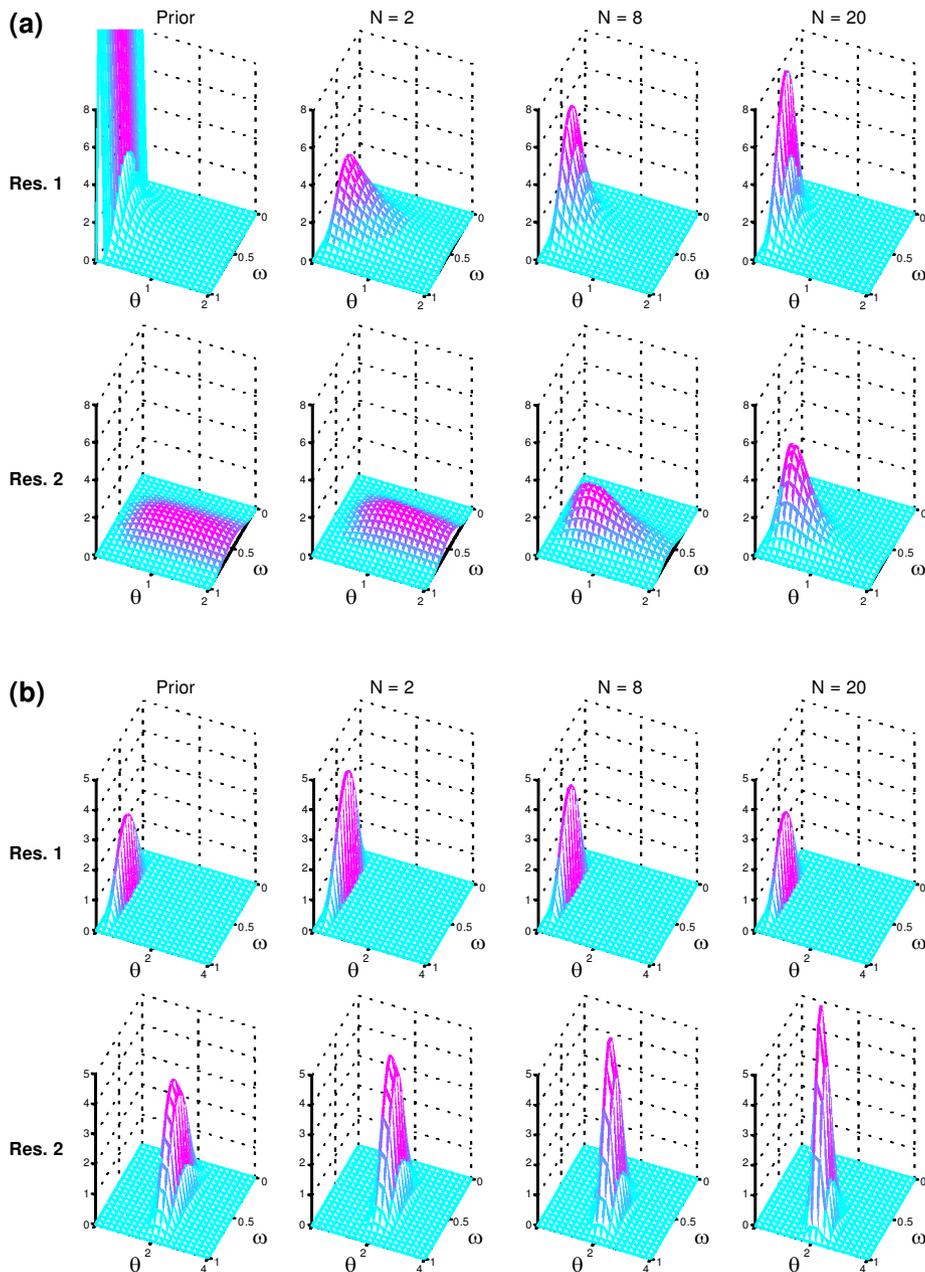
FIGURE 5. Probability densities of prior (first column) and posterior beliefs about rate (θ) of two researchers, as a function of the number of ingroup species ($N = 2$, 8, and 20) observed to be fixed for the same character state on the tree in Figure 3. Both researchers agree in their prior beliefs that symmetry (ω) is distributed as B(4, 4), i.e., that exchange between states is, on average, symmetrical. (a) The effect on posterior belief about average rate (θ) of observing 2, 8, and 20 ingroup species fixed for the same character state when prior belief is loosely held (Fig. 4a). As N increases, Researchers 1 and 2 gradually converge in their posterior beliefs. (b) The effect on posterior belief about average rate (θ) of observing 2, 8, and 20 ingroup species fixed for the same character state when prior belief is tightly held (Fig. 4b). Researchers 1 and 2 never converge in their posterior beliefs; instead, with increasing N, Researcher 2 alters his belief about symmetry.

previous applications of Bayesian statistics to the problem of ancestral character states. In the fully Bayesian approach adopted here, the uncertainty associated with unknown parameter values, which provides the grounds for subjective disagreement, is directly addressed by considering a range of prior belief.

Most of our Bayesian analyses utilize a general unrooted tree topology (Fig. 6a) consisting of an ingroup bush connected to an outgroup bush by a central branch. The number of species in the ingroup, the number of species in the outgroup, and the relative length of the central branch are varied. Results of these experiments are presented in Figures 7a-d and 8a-d. One set of analyses examines the effect of differing resolutions of ingroup topology, with ingroup size fixed at 5 species (Figs. 6b to 6e). The results of these experiments are presented in Figures 7e, 7f and 8e, 8f. In all cases the y-axis refers to the degree of posterior belief that the ingroup ancestral state is the same as the ingroup descendant state (state A), i.e., the ancestral state assignment under both parsimony and ML criteria.

In Figure 7, prior belief about rate (θ) is varied on the x-axis, ranging from 0 to 4 changes per unit of time, where 1 unit of time is the span across the tree from outgroup tip to ingroup tip. Prior belief is modeled as a gamma distribution, $\Gamma(\alpha, \beta)$; in the top row of each set of six plots, the prior belief about rate is tightly held ($\alpha = 32$); in the bottom row the prior belief is loosely held ($\alpha = 1$). In the first column of each set of six plots, the prior belief about symmetry is that state transitions favor state B three times as strongly as state A ($\omega = 0.25$); in the second column, the process is believed to be perfectly symmetrical ($\omega = 0.50$); and in the third column, the process is believed to favor state A three times as strongly as state B ($\omega = 0.75$). These prior beliefs about symmetry are modeled as a beta distribution, $B(v_1, v_2)$, and are held with a moderate level of conviction ($v_1 + v_2 = 16$). In Figure 8, prior belief about symmetry is varied on the x-axis, with symmetrical exchange at the midpoint, asymmetry favoring B increasing to the left of center, and asymme-

try favoring A increasing to the right of center. Prior belief is modeled as a beta distribution, $B(v_1, v_2)$; in the top row of each set of six plots, prior belief about symmetry is tightly held ($v_1 + v_2 = 1000$), in the bottom row the prior belief is much more loosely held ($v_1 + v_2 = 4$). In the first column of each set of six plots, the prior belief is that rate (θ) is low, 0.5 changes from ingroup to outgroup terminal; in the middle column prior belief about rate is 1, a moderate value; and in the right column, prior belief is that rate is relatively fast, 2 changes across the tree. These prior beliefs about rate are modeled as a gamma distribution, $\Gamma(\alpha, \beta)$, and held with a moderate level of conviction ($\alpha = 4$).

Regardless of prior belief about symmetry, the same general trend obtains across all plots in Figures 7 and 8: Researchers who have prior beliefs in low rates will disagree with those who have prior beliefs in high rates about the probability that the ancestor was in state A, with the former assigning higher probabilities to the ancestral state than the latter do. Obviously, as rampant change becomes increasingly probable, the character-state distribution in the ingroup descendant species retains little information about the ancestral state, and the reliability of an ancestral state reconstruction approaches the chance expectation (equal to the posterior value of ω). When prior belief about rate is tightly held (top rows, Figs. 7a–f), the data have less influence on posterior belief than when prior belief about rate is loosely held (bottom rows, Figs. 7a–f). Thus, as the number of ingroup species (N), all fixed for state A, increases in the bottom row of Figure 7d (where prior belief is loosely held), the posterior probability that the ancestor was also in state A rapidly increases. But when prior belief is tightly held (top row, Figure 7d), the same data have much less effect on the posterior probability.

Similarly, researchers who have a prior belief in transformation asymmetries favoring state B will usually disagree with researchers who have prior beliefs in transformation asymmetries favoring state A, with the former assigning higher probabilities to the ancestral state than the latter. (This ef-
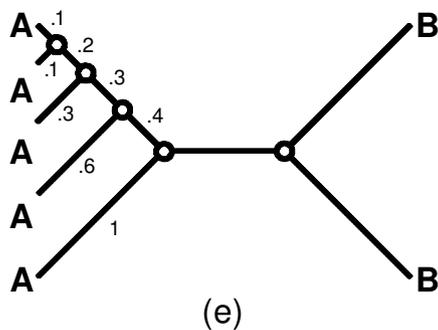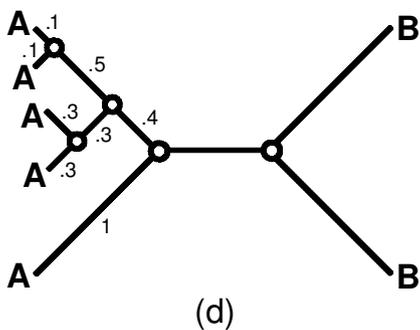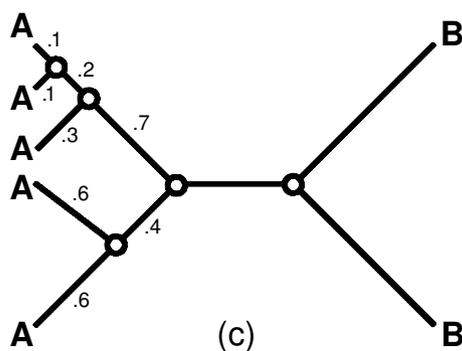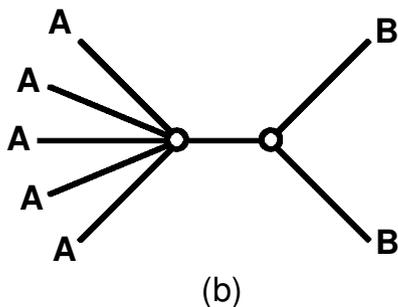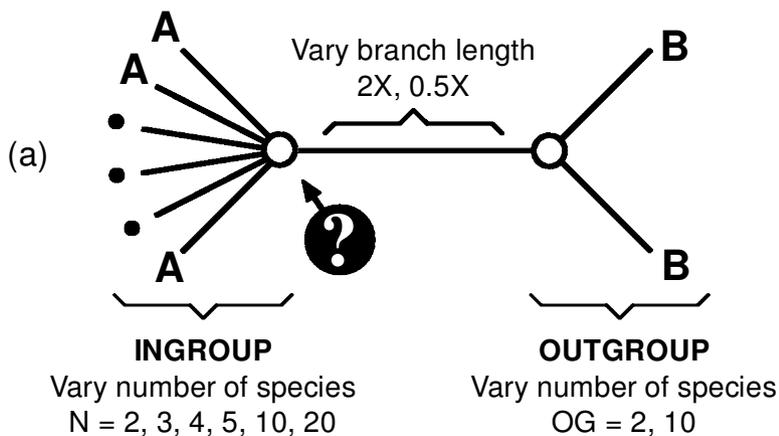
FIGURE 6. Basic tree topologies used in Figures 7 and 8. (a) The tree topology used in Figures 7a–d and 8a–d, in which the number of ingroup species (N) is varied from 2 to 20, the number of outgroup species is varied (2 vs. 10), and the length of the central branch is varied (long vs. short). (b–d) The tree topologies used in Figures 7e, f and 8e, f, in which the length of the central branch is varied. The indicated ingroup branch lengths were multiplied by a factor of 0.40 for the short-central-branch case, and by 0.25 for the long-central-branch case.

fect may be noted by comparing the first column of Figs. 7a–f, in which state B is favored, with the third column, in which state A is favored, or by comparing the left and
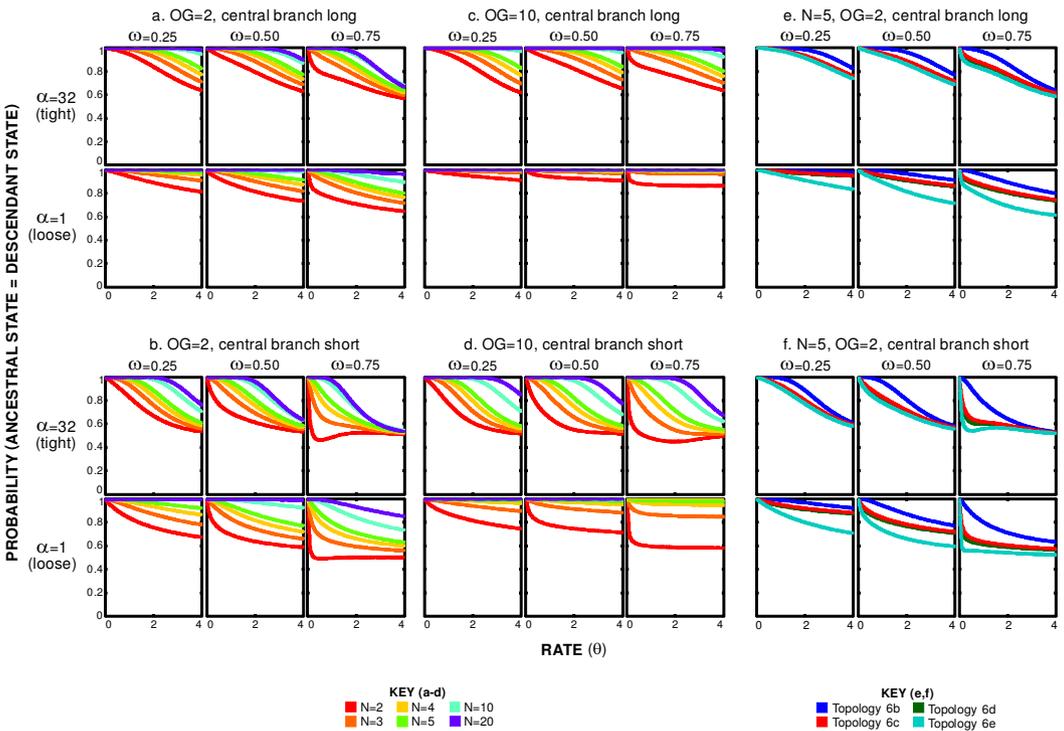
FIGURE 7. A fully Bayesian approach to estimating the reliability (posterior belief, y-axis) of the parsimony reconstruction that the ingroup ancestral state is the same as the ingroup descendant state on the trees in Figure 6, given the reality of a range of prior belief about rate (θ) across researchers, varied from 0 to 4 on the x-axis. Prior belief is modeled as a gamma distribution, $\Gamma(\alpha, \beta)$; in the top row of each group of six plots (panels a–f), prior belief about rate is tightly held ($\alpha = 32$); in the bottom row, it is loosely held ($\alpha = 1$). Prior belief about symmetry (ω), modeled as a beta distribution with $B(v_1, v_2)$ is varied across the columns of each six-plot group (panels a–f) and is held with a moderate level of conviction ($v_1 + v_2 = 16$). In the first column, the prior belief is that the process favors state B ($\omega = 0.25$); in the second column, the prior belief is that the process is perfectly symmetrical ($\omega = 0.50$); in the third column, the prior belief is that the process favors state A ($\omega = 0.75$). In the upper six-plot group of each vertical pair, the central branch is twice as long as the ingroup and outgroup branches; in the lower group, it is half as long. (a, b) The outgroup contains two species ("OG = 2"); (c, d) the outgroup contains ten species ("OG = 10"). (e, f) The resolved trees in Figures 6b–e are used, with $N = 5$ species in the ingroup and 2 species in the outgroup ("OG = 2"). (a–d) Curves correspond to number of ingroup species N, as indicated in the key; (e, f) curves correspond to the variously resolved topologies in Figures 6b–d.

the right portions of the plots in Figs. 8a–f.) This disagreement occurs because when prior belief about symmetry favors state B and especially when rates are believed to be low, there is a high posterior probability that the ancestor was in state A. When prior belief about symmetry favors state A, however, the probability that the ancestor was also in state A decreases because of the increasing probability of multiple B → A changes along the ingroup branches, especially when the number of ingroup species N is small. In Figure 8, the probability that the ancestor was in state A is generally higher when prior belief about symmetry is loosely held (row 2 in each set of six plots) than when prior belief is tightly held (row 1 of each set). This difference again demonstrates the greater influence of the data on a loosely held prior belief in comparison with the influence of the same data on a tightly held prior belief. Though persistent, these disagreements about posterior probabilities generally become less pro-
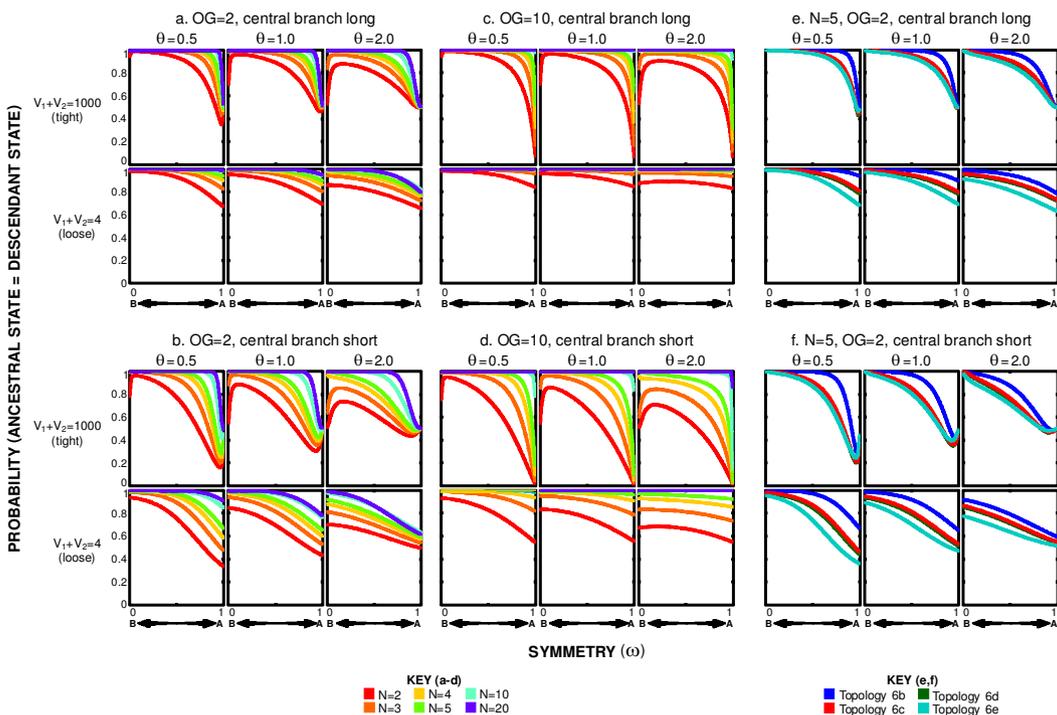
FIGURE 8. A fully Bayesian approach to estimating the reliability (posterior belief, y-axis) of the parsimony reconstruction that the ingroup ancestral state is the same as the ingroup descendant state on the trees in Figure 6, given the reality of a range of prior belief about symmetry (ω) across researchers, with symmetrical exchange at the midpoint, asymmetry favoring B increasing to the left of center, and asymmetry favoring A increasing to the right of center. Prior belief is modeled as a beta distribution, $B(v_1, v_2)$; in the top row of each group of six plots (panels a–f), prior belief about symmetry is tightly held ($v_1 + v_2 = 1000$); in the bottom row, it is more loosely held ($v_1 + v_2 = 4$). Prior belief about average rate (θ), modeled as a gamma distribution with $\Gamma(\alpha, \beta)$, is varied across the columns of each 6-plot group (panels a–f), and is held with a moderate level of conviction ($\alpha = 4$). In the first column, the prior belief is that rate is slow ($\theta = 0.5$); in the second column, the prior belief is that rate is intermediate ($\theta = 1.0$); in the third column, the prior belief is that rate is fast ($\theta = 2.0$). As in Figure 7, in the upper six-plot group of each vertical pair, the central branch is twice as long as the ingroup and outgroup branches; in the lower group, it is half as long. (a, b) The outgroup contains two species ("OG = 2"); (c, d) the outgroup contains ten species ("OG = 10"). (e, f) The resolved trees in Figures 6b–e are used, with $N = 5$ species in the ingroup and two species in the outgroup ("OG = 2"). (a–d) Curves correspond to number of ingroup species N, as indicated in the key; (e, f) curves correspond to the variously resolved topologies in Figures 6b–d.

nounced with increasing number of ingroup species N.

In general, across a wide range of prior belief about rate and symmetry, the probability that the ingroup ancestor was in state A decreases as the length of the central branch decreases. The reason for this is that a shorter central branch affords less opportunity for the required B $\rightarrow$ A change between the ingroup and the outgroup.

The polytomous ingroup topology is expected to provide a "best case" scenario for ancestral state reconstruction, because descendants are maximally informative about the state in their shared ancestor. Figures 7e, f, and 8e, f summarize the results of a set of experiments in which the number of ingroup species was held constant at $N = 5$ but ingroup topology was varied. In these plots the curves correspond to four separate topologies (Figs. 6b–d) rather than to the number of ingroup species. As expected, under a range of prior belief about rate and symmetry, the polytomous ingroup topol-
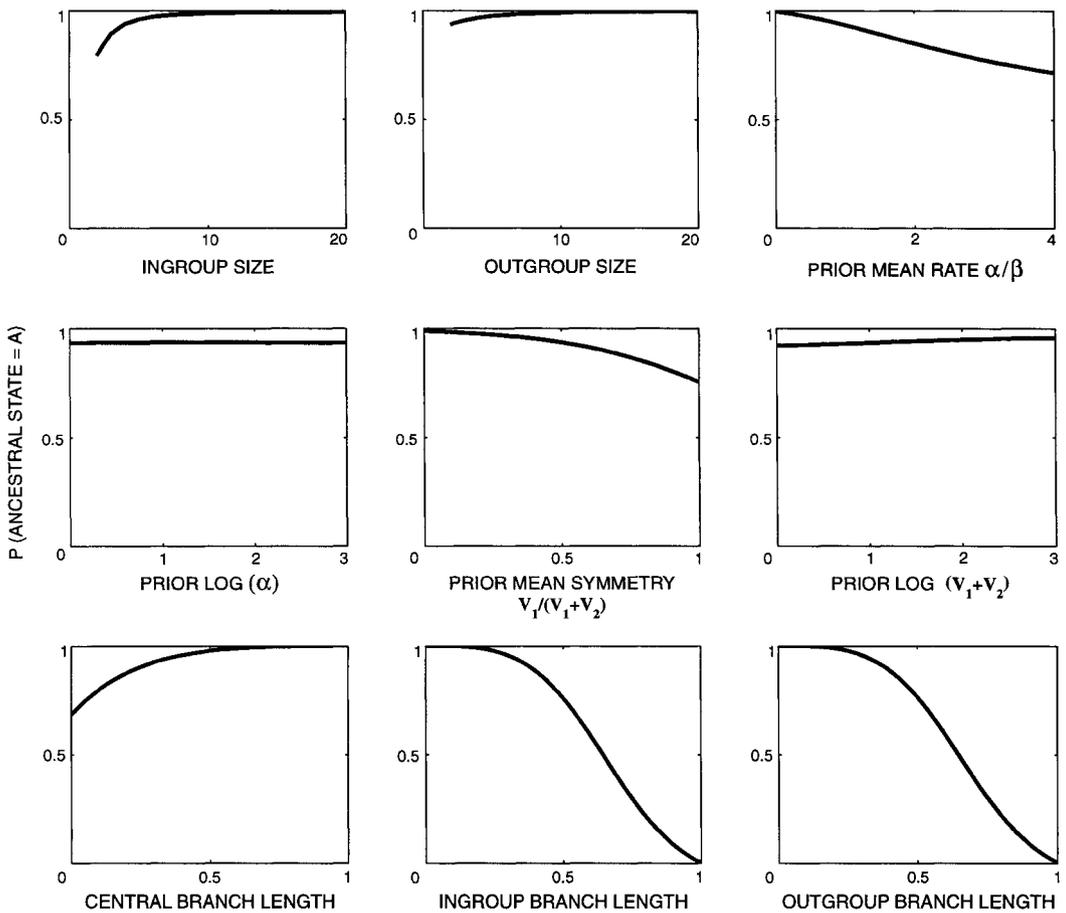
FIGURE 9.   The effect of varying parameters in isolation on the degree of posterior belief (*y*-axis) that the parsimony reconstruction is correct, i.e., that the ingroup ancestral state is the same as the ingroup descendant states, state A. Calculations employed the "default" tree topology in Figure 6a, in which there are four ingroup species (i.e., $N = 4$), two outgroup species, and all branch lengths are equal. Default prior belief about rate (θ) is modeled as $\Gamma(2, 2)$; default prior belief about symmetry (ω) is modeled as $B(1.5, 1.5)$.

ogy always confers the highest probability on the ancestral state. Even more interesting, the three resolved topologies sort out in a consistent rank order across the range of parameter values, with the most balanced or symmetrical topology (Fig. 6c) always following the bush (Fig. 6b) in conferring the second most reliable reconstruction of the ingroup ancestral state.

CONCLUSION

As is demonstrated by the exercises discussed above and by Figure 9, under a Bayesian approach that incorporates a range

of prior belief, the posterior probability associated with an ancestral state assignment is increased by decisive data in the form of: (*a*) unambiguous outgroup information, (*b*) a relatively large number of ingroup species, (*c*) relatively short ingroup branches, and (*d*) polytomous versus various resolved topologies. It is thus encouraging to note that nonsubjective information (i.e., the data) can have a strong effect on the reliability of ancestral state reconstruction regardless of prior belief.

We emphasize, however, that the artificial character universe modeled here is quite biased in favor of ancestral state recon-

struction, whereas many real-life situations are far less favorable. For example, real ingroups and outgroups are not necessarily as large as some of those modeled here, are not necessarily fixed unambiguously for a single state, and are usually dichotomously (rather than polytomously) resolved, as in (e.g.) Figures 7e, f and 8e, f, in which increased resolution is associated with decreased reliability of ancestral state reconstruction. Under such conditions, prior beliefs about rate and symmetry are likely to have a greater effect on posterior probability than are the objective data; i.e., prior (subjective) belief about character change will dominate the posterior probability.

The Bayesian approach outlined here allows us to identify such cases and thus to localize the argument to the arena of prior belief about characters and about character evolution, rather than to the arena of the information content of the character-mapping exercise. Unless methods for reconstructing ancestral character states improve, in such cases presumably "reasonable" people starting with different assumptions about how a character evolves must either defend those prior beliefs or else simply agree to disagree.

REFERENCES

BAYES, T. 1763. An essay towards solving a problem in the doctrine of chances. Philos. Trans. R. Soc. 53:370–418 (reprinted in 1958. Biometrika 45:293–315).

BROOKS, D.R. AND D.A. MCLENNAN. 1991. *Phylogeny, Ecology, and Behavior.* Chicago: University of Chicago Press.

CUNNINGHAM, C. W., K. E. OMLAND, AND T. H. OAKLEY. 1998. Reconstructing ancestral character states: A critical reappraisal. Trends Ecol. Evol. 13:361–366.

DAYHOFF, M. O., R. M. SCHWARTZ, AND B. C. ORCUTT. 1978. A model of evolutionary change in proteins. Pages 345–352 in Atlas of protein sequence and structure, vol. 5, suppl. 3 (M. O. Dayhoff, ed.). National Biomedical Research Foundation, Silver Spring, Maryland.

EDWARDS, W., H. LINDMAN, AND L. J. SAVAGE. 1963. Bayesian statistical inference for psychological research. Psych. Rev. 70:193–242.

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

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

FISHER, R. A. 1956. Statistical methods and statistical inference. Oliver and Boyd, Edinburgh.

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.

HARTIGAN, J. A. 1973. Minimum evolution fits to a given tree. Biometrics 29:53–65.

HENNIG, W. 1966. Phylogenetic systematics. University of Illinois Press, Urbana.

HOWSON, C., AND P. URBACH. 1989. Scientific reasoning: The Bayesian approach. Open Court, LaSalle, Illinois.

JONES, D. T., W. R. TAYLOR, AND J. M. THORNTON. 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8:275–282.

KOSHI, J. M., AND R. A. GOLDSTEIN. 1996. Probabilistic reconstruction of ancestral protein sequences. J. Mol. Evol. 42:313–320.

MADDISON, W. P. 1995. Calculating the probability distributions of ancestral states reconstructed by parsimony on phylogenetic trees. Syst. Biol. 44:474–481.

SCHLUTER, D. 1995. Uncertainty in ancient phylogenies. Nature 377:108–109.

SCHLUTER, D., T. PRICE, A. Ø. MOOERS, AND D. LUDWIG. 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51:1699–1711.

SCHULTZ, T. R., R. B. COCROFT, AND G. A. CHURCHILL. 1996. The reconstruction of ancestral character states. Evolution 50:504–511.

SOKAL, R. R., AND F. J. ROHLF. 1995. Biometry: The principles and practice of statistics in biological research, 3rd edition. W. H. Freeman and Co., New York.

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

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, AND D. M. HILLIS. 1996. Phylogenetic inference. Pages 407–514 in Molecular systematics, 2nd edition (D. M. Hillis,

C. Moritz, and B. K. Madle, eds.). Sinauer Associates, Sunderland, Massachusetts.

TAVARÉ, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lect. Math. Life Sci. 17:57–86.

WENZEL, J. W. 1997. When is a phylogenetic test good enough? *In* The origin of biodiversity in insects: Phylogenetic tests of evolutionary scenarios (P. Grandcolas, ed.). Mém. Mus. Nat. Hist. Nat. 173:31–45.

WENZEL, J. W., AND J. M. CARPENTER. 1994. Comparing methods: Adaptive traits and tests of adaptation. Pages 79–101 *in* Phylogenetics in ecology (P. Eggleton and R. Vane-Wright, eds.). Harcourt Brace & Co., London.

YANG, Z., S. KUMAR, AND M. NEI. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics 141:1641–1650.

ZHANG, J., AND S. KUMAR. 1997. Detection of convergent and parallel evolution at the amino acid sequence level. Mol. Biol. Evol. 14:527–536.

ZHANG, J., AND M. NEI. 1997. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. J. Mol. Evol. 44 (suppl. 1):S139–S146.

APPENDIX

"RATE" AND "SYMMETRY" PARAMETERS

The 2-state continuous-time, time-homogeneous Markov-process model of evolution is typically described in terms of rate parameters α and β, which refer to the relative rates of exchange between state A and state B such that the expected number of changes over an interval of duration t is (Tavaré, 1986):

$$t(\beta\pi_A + \alpha\pi_B) = t\,\frac{2\alpha\beta}{\alpha+\beta}$$

We can reparameterize this model by setting

$$\theta \equiv \beta\pi_A + \alpha\pi_B = \frac{2\alpha\beta}{\alpha+\beta} \ \text{ and } \ \omega \equiv \frac{\alpha}{\alpha+\beta}$$

Now θ is the average rate ("speed") of change and ω is a measure of the symmetry of the process. The state-transition probabilities become

$$P_{AA}(t) = \omega + (1-\omega)e^{-\theta t/2\omega(1-\omega)}$$

$$P_{AB}(t) = (1-\omega) - (1-\omega)e^{-\theta t/2\omega(1-\omega)}$$

$$P_{BA}(t) = \omega - \omega e^{-\theta t/2\omega(1-\omega)}$$

$$P_{BB}(t) = (1-\omega) - \omega e^{-\theta t/2\omega(1-\omega)}$$

When $\omega > 1/2$, the process tends toward state A and when $\omega < 1/2$, the process tends toward state B.

Prior Distributions

The prior distribution on the rate of the process, i.e., the average number of changes per unit of time, is modeled as a gamma distribution $\theta \sim \Gamma(\alpha, \beta)$. Thus, $P(\theta) \propto e^{-\beta\theta}\theta^{\alpha-1}$ has prior mean α/β and becomes more concentrated with increasing α. The prior distribution on the symmetry of the process, i.e., the tendency of the process to favor one or the other of states A and B, is modeled as a beta distribution $\omega \approx B(v_1, v_2)$. Thus, $P(\omega) \propto \omega^{v_1-1}(1-\omega)^{v_2-1}$ has prior mean $v_1/(v_1 + v_2)$ and becomes more concentrated with increasing $v_1 + v_2$.

Likelihood and Posterior Probabilities

Let x = data observed and s = the (unobserved) distribution of ancestral states. The likelihood is

$$P(x|\theta, \omega) = \sum_{s\in S} P(x, s|\theta, \omega) = \sum_{s\in S} P(x|s, \theta, \omega)P(s|\theta, \omega).$$

The posterior distribution of the model parameters θ, ω is proportional to the product of the likelihood and the prior probability of those parameters

$$P(\theta, \omega|x) \propto P(x|\theta, \omega)P(\theta, \omega)$$
$$= P(x|\theta, \omega)P(\theta)P(\omega)$$

The posterior distribution on the ancestral state S is

$$P(s|x) = \int\int P(s, \theta, \omega|x)\partial\theta\partial\omega$$
$$= \int\int P(x|s, \theta, \omega)P(s|\theta, \omega)P(\theta)P(\omega)\partial\theta\partial\omega$$

The multiple integration is easily computed numerically. The likelihood $P(x|\theta, \omega)$ and the conditional (on the ancestral state) likelihood are computed by standard pruning algorithms (Felsenstein, 1981).