LIKELIHOOD OF ANCESTOR STATES IN ADAPTIVE RADIATION

DOLPH SCHLUTER,[1,2] TREVOR PRICE,[3] ARNE Ø. MOOERS[1,4] AND DONALD LUDWIG[1,5]

[1]Department of Zoology and Centre for Biodiversity Research, University of British Columbia, Vancouver,
British Columbia V6T 1Z4, Canada
[2]E-mail: schluter@zoology.ubc.ca
[3]Biology Department O-116, University of California—San Diego, La Jolla, California 92093
E-mail: tprice@ucsd.edu
[4]E-mail: mooers@bio.uva.nl
[5]E-mail: ludwig@zoology.ubc.ca

Abstract.—Theories of ecological diversification make predictions about the timing and ordering of character state changes through history. These theories are testable by "reconstructing" ancestor states using phylogenetic trees and measurements of contemporary species. Here we use maximum likelihood to estimate and evaluate the accuracy of ancestor reconstructions. We present likelihoods of discrete ancestor states and derive probability distributions for continuous ancestral traits. The methods are applied to several examples: diets of ancestral Darwin's finches; origin of inquilinism in gall wasps; microhabitat partitioning and body size evolution in scrubwrens; digestive enzyme evolution in artiodactyl mammals; origin of a sexually selected male trait, the sword, in platies and swordtails; and evolution of specialization in *Anolis* lizards. When changes between discrete character states are rare, the maximum-likelihood results are similar to parsimony estimates. In this case the accuracy of estimates is often high, with the exception of some nodes deep in the tree. If change is frequent then reconstructions are highly uncertain, especially of distant ancestors. Ancestor states for continuous traits are typically highly uncertain. We conclude that measures of uncertainty are useful and should always be provided, despite simplistic assumptions about the probabilistic models that underlie them. If uncertainty is too high, reconstruction should be abandoned in favor of approaches that fit different models of trait evolution to species data and phylogenetic trees, taking into account the range of ancestor states permitted by the data.

Key words.—Adaptive radiation, ancestor states, maximum likelihood, evolution of specialization, phylogeny.

Theories of ecological diversification often make predictions about the timing and ordering of character state changes through history. For example, several theories of adaptive radiation and faunal buildup assume that generalist ancestors give rise to more specialized descendants who divide resources increasingly finely as the number of species rises (Mayr 1942; Simpson 1944; Sugihara 1980). A second example is the taxon cycle hypothesis, which proposes that colonists to islands evolve through a series of ecological stages from marginal to stable habitats (Wilson 1961; Ricklefs and Cox 1972) or from one extreme body size to the other (Rummel and Roughgarden 1985). Such ideas are now being tested by "reconstructing" attributes of ancestors from data on contemporary species and a tree of their phylogenetic relationships (e.g., Roughgarden and Pacala 1989; Liebherr and Hajek 1990; Losos 1992a,b, 1996; Jermann et al. 1995). Our aim here is to evaluate the accuracy of such ancestor reconstructions.

The usual approach to reconstructing ancestor states is to choose those estimates requiring the smallest number of state changes through time (maximum parsimony; see Maddison and Maddison 1992). This approach has two potential weaknesses to which evolutionary analyses of ecological traits are especially prone (but which arise in many other circumstances). First, many theories of ecological diversification assume that rates of character change are not low, in which case a parsimony criterion may not be appropriate (e.g., Zhang and Nei 1997). Second, ancestral states are inevitably uncertain if the evolutionary process includes a stochastic (unpredictable) element. When estimating ancestor states we therefore need to know the alternative reconstructions not ruled out by data. Existing maximum-parsimony methods do not quantify uncertainty, although error rates could be calculated in principle (Maddison 1995).

We show that accuracy of ancestral states can be quantified using maximum-likelihood methods. In contrast, other discussions of accuracy have been limited mainly to hypothetical scenarios (Frumhoff and Reeve 1994; Schultz et al. 1996;

1699

Zhang and Nei 1997). Our method for discrete traits was presented briefly by Schluter (1995) and is explained in more detail below. It is similar to an approach developed independently for molecular sequences by Yang et al. (1995), Koshi and Goldstein (1996), and Felsenstein (pers. comm. 1995). We extend the method to continuous traits.

Our examples are motivated by questions about ecological history: did the original Darwin's finch eat mainly seeds, insects, or leaves? Are the founders of adaptive radiations generalists or specialists? We analyze several examples from the literature. These examples illustrate the method, show the kinds of results that may be obtained, illuminate pitfalls of the approach, and suggest possible remedies. We conclude that ancestor states are often highly uncertain, particularly those of early ancestors. We feel that these levels of uncertainty are realistic and not pathological outcomes of statistical assumptions about the methods, but we indicate how these estimates of uncertainty might be improved. We conclude that uncertainty should always be measured and, if it is high, that ancestor state reconstruction be abandoned in favor of approaches that fit alternative models of diversification to evolutionary trees and species measurements.

METHODS

Justification for Evolutionary Models

An explicit model of evolutionary change is needed to calculate likelihoods of ancestor states. We use random walks in continuous time for this purpose. The Markov process is used to model transitions between discrete states (Felsenstein 1993; Frumhoff and Reeve 1994; Pagel 1994; Maddison 1995; Schluter 1995). The analogous process for continuous traits is Brownian motion (Felsenstein 1985; Maddison 1991). The salient features of both models are: (1) probability of change at a point in time along any branch of the phylogenetic tree depends only on the character state at that time, and not on prior character states; (2) transitions along each branch are independent of changes elsewhere in the tree; and (3) rates of change are constant throughout time and along all branches. We estimate these rates from the tree and current species values, rather than from prior knowledge or belief.

Use of random walks to model evolutionary changes in ecologically significant (i.e., adaptive) characters may appear contradictory. However, the random walk is intended to mimic the unpredictable aspect of evolution by natural selection, and using it does not imply that changes are truly "random." Markov and Brownian motion processes are merely simple ways to represent unpredictability. Nevertheless, they may often fail to capture important aspects of evolutionary history. The assumption of constant rates is perhaps the most glaring weakness. More realistic models (e.g., in which rates of evolution may vary) would remedy this problem. We confine ourselves to the simplest scenarios. It is an open question whether use of more realistic (and hence more complex) models would result in estimates less uncertain than those based on the simpler methods presented here.

When analyzing data using these methods we assume that the phylogenetic tree and all of its branch lengths are known with certainty. This is not a required assumption, but it simplifies the presentation. Ultimately it should be possible to take sampling error in tree construction into account when estimating ancestors (e.g., Martins 1996).

Discrete Traits

The discrete method is a straightforward extension of Pagel's (1994) Markov model for estimating rate of evolution of a trait X having two states, 0 and 1. The species values for the trait X, and the tree of phylogenetic relationships among the species, are the data. The parameters q_{01} and q_{10} are the instantaneous transition rates between the two states. The likelihood of a pair of rate parameters $L(q_{01}; q_{10})$ is evaluated in the following way (Pagel 1994): (1) choose a set of character states $X_1, X_2 \ldots X_N$ for all N interior nodes (ancestors) of the tree; (2) use an estimate of q_{01} and q_{10} to calculate for every ancestor i and its two descendants $i' = 1$ and $i' = 2$ the transition probabilities p_{i1} and p_{i2}, where $p_{ii'}$ is the probability that an ancestor beginning in state X_i is in state $X_{i'}$ after time $v_{ii'}$, the branch length between i and i' (descendants i' may be other ancestral nodes or species at the tree tips); (3) calculate the probability of the set of ancestor states as $P(X_1, X_2 \ldots X_N) = \Pi_{i,i'} p_{ii'}$, where the product is over all branches in the tree; (4) repeat the above three steps for each of the 2^N possible combinations of character states for the N ancestors. The likelihood of the parameters q_{01}, q_{10} given the data is obtained by summing the probabilities of each combination of ancestral states

$$L(q_{01}, q_{10}) = \sum_{X_1, X_2 \ldots X_N} P(X_1, X_2 \ldots X_N), \qquad (1)$$

The maximum-likelihood (ML) estimates of the rate parameters, $\hat{q}_{01}$ and $\hat{q}_{10}$, are the values maximizing $L(q_{01}, q_{10})$.

Often limits on the size of the dataset will not warrant the separate fitting of two parameters, and it is best to start with a simpler model in which the forward and backward rates of transition are set equal (i.e., $q_{01} = q_{10}$). This is especially true when the formula is expanded to characters with more than two states, which is accomplished by increasing the number of parameters and possible combinations of ancestor states. A character with three states (0, 1, and 2) would require six parameters ($q_{01}, q_{02}, q_{10}, q_{12}, q_{20}, q_{21}$); a character with four states needs 12. Very few phylogenetic trees in the literature include enough species to warrant fitting a model with six parameters, and a simpler model in which all transition rates are set equal will be justified. An extension of this approach might relax the equality constraint by allowing forward and backward rates to differ, but only within predetermined orders of magnitude. Limiting the number of parameters is more crucial when estimating ancestor states than when estimating phylogenetic trees because here we are interested in a single character. In contrast, likelihood estimates of trees use information on many characters (e.g., base pairs) simultaneously.

To estimate the states at ancestral nodes we condition on the ML rate estimates, $\hat{q}_{01}$ and $\hat{q}_{10}$. The likelihood of an estimate of μ_i, the state of ancestor i, is then obtained as the contribution of that state to $L(\hat{q}_{01}, \hat{q}_{10})$ (Schluter 1995). For example, the likelihood of a particular state for ancestor $i = 1$ is

$$L(\mu_1 | \hat{q}_{01}, \hat{q}_{10}) = \sum_{X_2, X_3 \ldots X_N} P(\mu_1, X_2, X_3 \ldots X_N). \qquad (2)$$

The ancestral state having the highest likelihood is the ML estimate.

The ratio of the likelihoods of the two estimates measures the level of support for the ML estimate. Under normal distribution theory, support limits (analogous to 95% confidence limits) for ML estimates include all values whose *ln*(likelihoods) lie within two units of the maximum (Edwards 1972). These limits correspond to a likelihood ratio of exp(2) = 7.4. The validity of this theory is unclear in the present context since each node has at most two realized descendants. We therefore treat likelihood ratios simply and directly as measures of relative weight of evidence, and the number 7.4 as a rough minimum threshold for gauging level of certainty. An alternative approach judges the statistical significance of a likelihood ratio. Standard likelihood ratio tests are not possible because alternative reconstructions are not nested. A general solution to this problem based on computer simulation is summarized in Goldman (1993).

Finally, the likelihoods may be interpreted in a Bayesian context (e.g., Yang et al. 1995). The alternative states at any node are best regarded as equiprobable a priori (i.e., the prior probability distribution of states is flat). In this case the posterior probability of a state is proportional to its likelihood. For example, a likelihood ratio of nine would correspond to a posterior probability of 0.90 in favor of the ML estimate.

It is easy to show using equation (2) that if the transition rates are very small (i.e., $q_{01} \to 0$ and $q_{10} \to 0$), then the ML estimates of ancestor states are those minimizing the total number of transitions throughout the tree, which is the maximum-parsimony criterion. This implies that maximum parsimony is a limiting case of maximum likelihood, and underscores the central assumption of parsimony in ancestor reconstruction: that change is rare. (The same limit holds when estimating phylogenetic tree topology using maximum likelihood; Felsenstein 1973).

All computations were carried out using DISCRETE 1.0 (Pagel 1994) and a reduced program developed independently. The next release of DISCRETE, expected soon, will provide likelihoods of ancestor states on output (M. Pagel, pers. comm. 1997).

Continuous Traits

The method for continuous traits assumes that each trait evolves according to a Brownian motion process governed by a rate parameter β. Under this model the expected squared difference (variance) between any two species is β multiplied by the time since the species last shared a common ancestor. This model is the basis of the most widely used comparative method in evolutionary biology (phylogenetically independent contrasts; Felsenstein 1985). The ML ancestor states are those minimizing the sum of squared changes along the branches (weighted "squared change parsimony," Maddison 1991). These estimates also have maximum posterior probability when the prior probability distribution is flat (Maddison 1991). We summarize these results below. In the Appendix we derive the marginal distributions of β and the ancestor states (see also Martins and Hansen 1997). The Brownian motion model assumes that the evolutionary pro-

cess includes no trends, since trends cannot be estimated from data solely on contemporary species.

The likelihood function for the rate β and the set of ancestor states $\underset{\sim}{\mu} = \{\mu_1, \mu_2 \ldots \mu_N\}$ has the form

$$L(\underset{\sim}{\mu}, \beta) \propto \frac{1}{\beta^N} \exp\left(\frac{-Q(\underset{\sim}{\mu})}{2\beta}\right), \qquad (3)$$

where Q is

$$Q(\underset{\sim}{\mu}) = \sum_{i,i'} \frac{(\mu_i - \mu_{i'})^2}{v_{ii'}} \qquad (4)$$

(e.g., Guttman et al. 1982). As before, $i' = 1$ and $i' = 2$ are the two descendants of each ancestor i, $i = 1, \ldots N$, and $v_{ii'}$ is the branch length (time) between i and i'. Descendants i' may be other ancestral nodes or they may be tips (species). Q is thus the weighted sum of squared differences between ancestors and descendants, where the sum is over all branches of the tree, and where branch lengths are the weights.

It follows from the above that the set of ancestor states minimizing Q (i.e., minimizing the squared differences between ancestors and descendants) is the ML estimate, $\hat{\underset{\sim}{\mu}}$. In the Appendix we show that

$$t = (\mu_i - \hat{\mu}_i)/C_i \qquad (5)$$

has a t-distribution with N degrees of freedom. C_i is the standard deviation of the marginal distribution of μ_i (equivalent to the "standard error" of $\hat{\mu}_i$). From this we calculate support limits as $\hat{\mu}_i \pm C_i t_{0.05,N}$ (equivalent to 95% confidence limits) for all ancestors in turn. This derivation assumes a flat prior probability distribution for the parameters $\underset{\sim}{\mu}$ and $ln\beta$. The joint marginal distributions of ancestor states at nodes is given in the Appendix. It is also possible to derive joint marginal distributions of several traits simultaneously. These are complex and we do not present them. Instead, we recommend that traits be analyzed separately.

Computations were carried out with a computer program written by adapting routines in PHYLIP 3.5 for the analysis of continuous traits (Felsenstein 1993). This program is available from DS.

RESULTS

Five examples illustrate the method for discrete traits. Change is "rare" in the first two examples. By "rare" we mean that it is possible to reconstruct ancestral states such that at most a single transition occurs between any two states. These are the circumstances under which parsimony methods should work best, and uncertainty about ancestral states should be least. Change is not rare in the remaining examples. Two other examples illustrate the method for continuous traits.

Discrete Traits—Change Rare

Diets of Ancient Galápagos Finches.—Interspecific diversity in diet and beak dimensions in the 13 Galápagos finches is large by continental avian standards (Grant 1986; Schluter 1988). We grouped species into three main (dry season) diet states: granivore (ground finches, *Geospiza*), insectivore (tree finches *Camarhynchus* and warbler finch *Certhidea olivacea*),
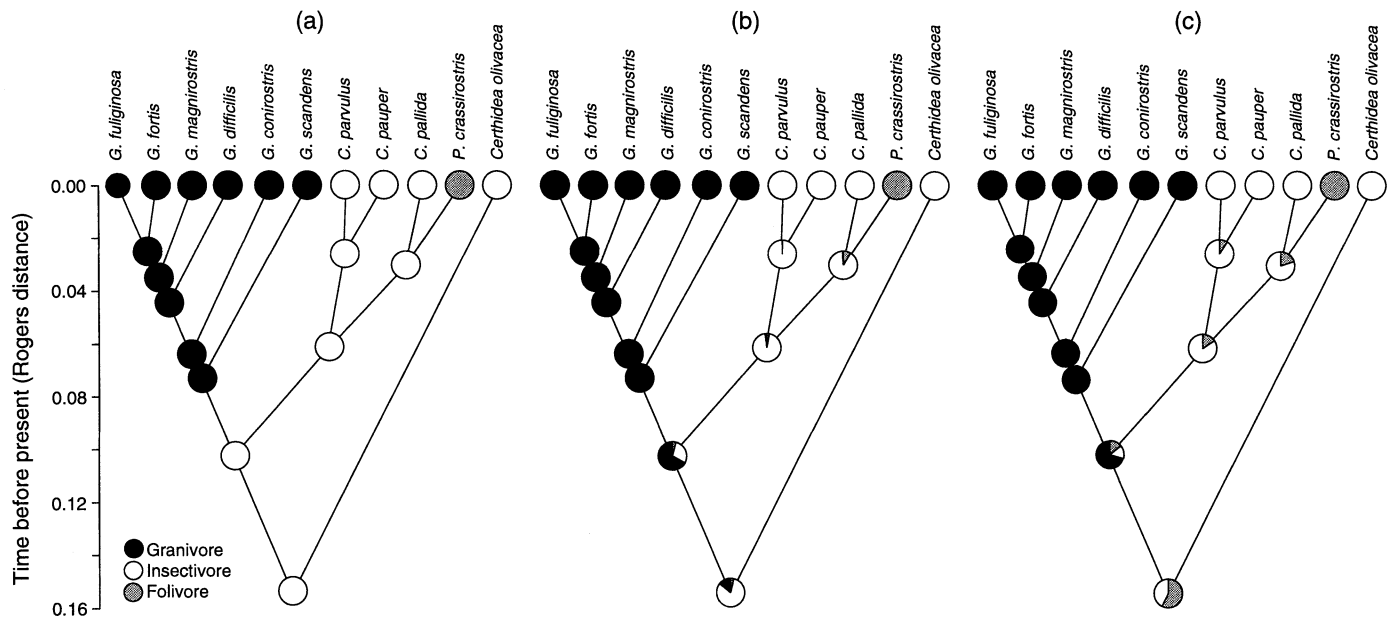
FIG. 1. A phylogeny of 11 of the Galápagos finches based on allozymes (Yang and Patton 1981). The tree is a UPGMA diagram calculated from Rogers's distances. Statistical support for the three groups is fairly strong (ground finches *Geospiza*, tree finches *Camarhynchus-Platyspiza*, and the warbler finch *Certhidea*), but other aspects of tree topology are weakly supported (Stern and Grant 1996). (a) Ancestor states estimated using the method of maximum parsimony (Maddison and Maddison 1992). (b) Maximum-likelihood ancestor states assuming that all transition rates q are equal; $\hat{q} = 3.6$. Area of pies indicate relative support for different ancestor states. (c) Maximum-likelihood ancestor states when transition rates are free to vary.

and folivore (the vegetarian finch *Platyspiza crassirostris*). This classification oversimplifies by ignoring special habits of certain populations (e.g., blood and pollen consumption by some ground finches; Grant 1986; Grant and Grant 1989). A phylogeny of 11 of the species is available from allozymes (Fig. 1); phylogenies based on morphology are similar (Schluter 1986; Grant 1986). Various continental sister species from the Emberizinae have been proposed, but none of them has been confirmed (P. T. Boag and P. R. Grant, pers. comm. 1995).

Reconstructions based on maximum parsimony indicate that two major shifts in diet took place from an insectivorous ancestor state—once to granivory, and once to folivory (Fig. 1a). ML ancestor states are similar (Fig. 1b) with one exception: the common ancestor of the tree finch (*Camarhynchus*) and ground finch (*Geospiza*) clades is most likely a granivore not an insectivore (but this estimate is only weakly supported). The parsimony and ML estimates for this ancestor differ because the ML estimate takes branch lengths into account: less time is available for change along the branch leading to the ground finches (Fig. 1b). Support for an insectivorous root is moderate (supported 5.7:1 over granivore). ML estimates of more recent ancestors are more strongly supported.

We also inspected the results of an unconstrained analysis in which all six transition rate parameters were free to vary (Fig. 1c). We do not recommend this strategy, given that only 11 data points (species) are available, but the results are instructive. The ML estimate for the common ancestor to all the species is folivore, which seems incredible given that only a single descendant (on a short branch) eats leaves. Our explanation for this paradox is as follows: a change that oc-

curs only rarely is unlikely to happen along such a short branch. When all transition rates are unconstrained, a likely explanation for a unique character at the tip of a short branch is that it is a relict whose rarity among contemporary species is the result of a high transition rate away from folivory to other diets. This finding emphasizes that likelihood calculations are not penalized for the number of evolutionary transitions supposed. As a result, a unique state at the tips of short branches will often reappear at the root of a tree accompanied by a high transition rate away from this state. Unconstrained maximum likelihood does not generally work well when all change is truly rare, because individual rate estimates are inevitably poor when transitions are unique. Constraining transition rate parameters to be equal (e.g., Fig. 1b) gives unique states no special weight and leads to more realistic estimates when sample size is small.

Origin of Inquilinism in Gall Wasps.—The larvae of gall wasps develop inside galls while feeding on plant tissues. Most species make their own galls, but a number of species are inquilines: they cannot form galls and instead develop inside those of other species, usually killing the host wasps. Ronqvist (1994) constructed a phylogeny to test whether inquilines evolved once and then radiated, or whether each inquiline species evolved recently from a gall-inducing wasp—such as its current host. The phylogeny showed that all inquilines form a single clade (Fig. 2), such that only a single transition to inquilinism from a gall-inducing ancestor is needed to account for the differences between species.

Ronqvist (1994) rigorously tested tree structure to confirm that the inquilines are monophyletic. In Figure 2 we test his ancestor states. The ML results are identical to those based on parsimony, and each ML node is strongly supported. Even
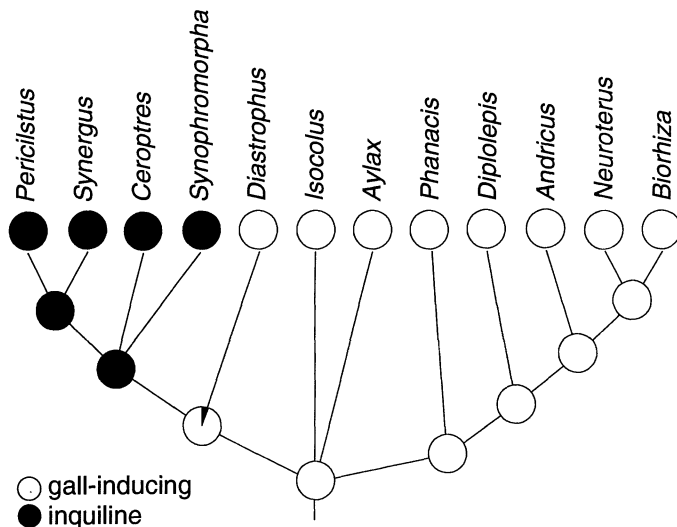
FIG. 2. Phylogeny of 12 species of gall wasps, one species per genus (Ronqvist 1994). This tree is a consensus of five minimum-length trees based on a suite of morphological measurements (modified from Fig. 3 in Ronquist 1994). Branch lengths (divergence times) are unknown and all were set to one. Polytomies were assumed to be real multifurcations (zero branch length between successive speciation events). Pies at nodes indicate relative support for the two ancestral states. $\hat{q}_{01} = \hat{q}_{10} = 0.06$.

at the most uncertain node (the ancestor common to the inquiline clade and the gall-inducing *Diastrophus*) gall induction has 17 times more support than inquilinism.

These two examples demonstrate that when change is rare likelihood often leads to similar estimates of ancestor states as those derived from applying the method of maximum parsimony. Exceptions occur because maximum likelihood is able to take time available for change into account. They also show that uncertainty of ancestral states may be relatively low when change is rare. This agrees with the results of Hillis et al. (1994) in experimental phylogenies of T7 bacteriophage.

Discrete Traits—Change Not Rare

Microhabitat Shifts in Diversification.—Ecological diversification in birds typically happens along several resource dimensions such as habitat, foraging location and prey size. The frequency and ordering of transitions along different dimensions through time are of great interest to ecologists (Diamond 1986; Richman and Price 1992; Suhonen et al. 1994). Australo-Papuan scrubwrens (*Sericornis*) are one group that have been studied to address these questions (Diamond 1986; Christidis et al. 1988). Scrubwrens are small, grey-brown birds that hunt insects in dense vegetation, mainly in wet forests. Here we focus on shifts between primarily terrestrial and primarily arboreal foraging strata (Fig. 3).

Maximum-likelihood estimation suggests that the arboreal condition is ancestral and that a terrestrial habit evolved
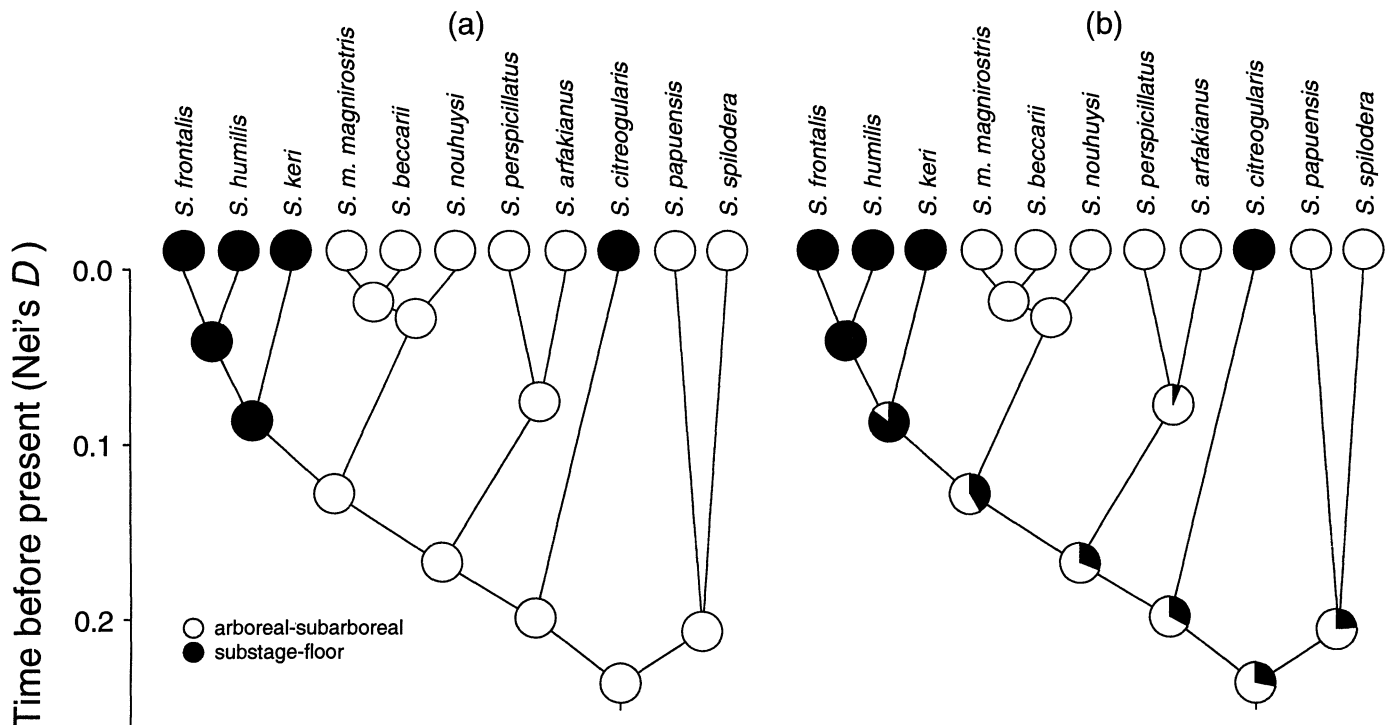

FIG. 3. Vertical foraging stratum in 11 scrubwrens (*Sericornis*). The tree is from a UPGMA analysis of Nei's (1978) genetic distances calculated on allozyme frequencies (Christidis et al. 1988). Tree structure was corroborated by a reanalysis of mtDNA restriction sites in five of these species using RESTML (Felsenstein 1993; data from Joseph and Moritz 1993). (a) Ancestor states estimated using maximum parsimony (Maddison and Maddison 1992). (b) Maximum-likelihood ancestor states assuming that all transition rates q are equal; both $\hat{q} = 5.93$. Area of pies at nodes in (b) indicate relative support for each ancestral state.
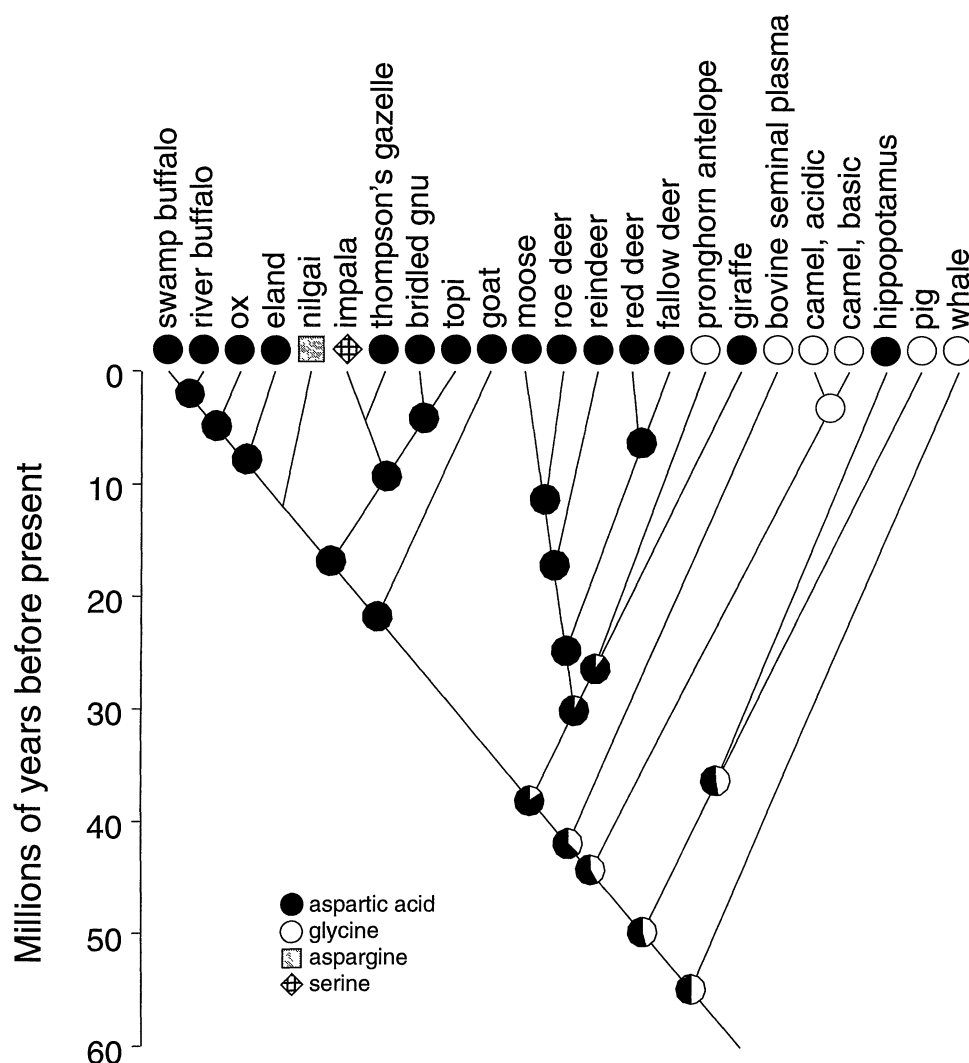
FIG. 4. Phylogeny of artiodactyls, with a whale as outgroup. The tree is modified from Jermann et al. (1995); divergence times are very approximate. Symbols indicate amino acid residue at position 38 of ribonuclease, the residue most important in determining activity of synthesized enzymes (Jermann et al. 1995). Two species (nilgai and impala) having residues other than aspartic acid or glycine were excluded from this analysis. Pies indicate relative support for the two ancestral states. Both $\hat{q} = 0.012$.

twice, once in the ancestor of *S. citreogularis* and a second time in the ancestor of the *S. frontalis* clade (Fig. 3b). Maximum parsimony yields the same estimates (Fig. 3a) but fails to indicate the high degree of uncertainty of many ancestral states indicated by maximum likelihood. Likelihood ratios for all but four recent ancestors are 3.0 or less. The certainty of an ancestor to any given pair of species having the same character state may be high if the ancestor is recent; uncertainty grows as the branch lengths increase (e.g., compare the ancestor of *S. papuensis* and *S. spilodera* to that of *S. perspicillatus* and *S. arfakianus*). Every ancestor is uncertain that has a diversity of character states among its descendants. As a result, many ancestor states are compatible with the foraging habits of descendant species.

History of Digestive Enzyme Activity.—In one of the most interesting applications of reconstruction methods, Jermann et al. (1995) estimated and then synthesized ancestral amino acid sequences of artiodactyl pancreatic ribonuclease. The enzyme is present in large quantities in the guts of ruminants and species with ruminant-like digestion, where it digests RNA derived from symbiotic microflora (Beintema et al. 1986). The synthesized enzymes corresponding to different ancestors were thermostable but exhibited considerable variation in catalytic activity that was attributable chiefly to differences in the amino acid residue at position 38 (Fig. 4). Molecules with glycine at position 38 had five times as much activity on double-stranded RNA as those with aspartic acid. Parsimony suggested that glycine was the root state and that a transition to aspartic acid, and a consequent five-fold reduction in catalytic activity, occurred after the camels split from the rest (Fig. 4).

Maximum likelihood indicates that none of the distant ancestors is well estimated. The ML state is glycine rather than aspartic acid for all these distant ancestors, but this has barely more support than the alternative state. As in the scrubwrens, uncertainty increases backward in time and with the diversity

of character states among descendants. This analysis leaves out the two rare states, but results are similar when they are included. Variations in tree topology and in the taxa included also lead to similar conclusions (Schluter 1995).

Uncertainty of distant ancestors is far less, and aspartic acid is supported as the state of these old ancestors, if the much lower transition rates between aspartic acid and glycine in other proteins and in other organisms are used for q_{01} and q_{10} instead (e.g., Dayhoff amino acid transition rates; see Benner and Gonnet 1995). However, transitions between glycine and aspartic acid at position 38 of ribonuclease are a regular feature within the class Mammalia—they occur also in rodents and kangaroos (Beintema et al. 1986). This argues against using the low estimates from other groups. A high frequency of transitions makes it difficult to estimate the character state of any ancient mammal.

Sexual Selection and Swords in Ancient Swordtails.—Diversification in sexually selected traits frequently accompanies speciation and divergence, but the mechanisms underlying the origin of preferences for such traits during mate choice remain poorly understood. Basolo (1990) used swordtails and platies (*Xiphophorus*) to test the "sensory exploitation" hypothesis (Ryan and Keddy-Hector 1992) for the origin of exaggerated male characters. The hypothesis is that female preferences for exaggerated male traits arise incidentally as by-products of natural selection on general aspects of the sensory system prior to the occurrence of the traits in males. Exaggeration of the male traits occurs later as genetic variation in such traits arises and become selected by the female preferences.

Basolo (1990) used parsimony to reconstruct ancestral states for male swords and female preference for male swords onto a preliminary phylogeny for swordtails and platies. The results indicated that the preference originated well before the first sword, as predicted by the hypothesis. Meyer et al. (1994) revised this conclusion when they presented a molecular phylogeny of the *Xiphophorus* that suggested swords were an ancient, rather than a recent, character. Wiens and Morris (1996) argue that Basolo's (1990) conclusion might still be correct: if a sword is defined strictly as a colored extension of the caudal fin then parsimony still favors swordlessness as the ancestral state in the molecular phylogeny. Basolo (1995) has recently shown that females of *Priapella olmacae*, in the sister genus to *Xiphophorus*, also prefers swords, suggesting that a preference for swords predates the origin of *Xiphophorus*. No one has attempted to assess the accuracy of the ancestor reconstructions, even assuming the molecular tree is correct.

We used the molecular phylogeny of *Xiphophorus* and *Priapella* in Meyer et al. (1994), and the strict definition of a sword favored by Basolo (1990) and Wiens and Morris (1996), to calculate likelihoods of ancestor states (Fig. 5). The results indicate that the ancestral condition is uncertain, owing to a high rate of transition between swordlessness and swords through the history of the clade. The relative likelihoods of the states swordless:sword is 1:1 in the common ancestor to the 22 *Xiphophorus* species. Swordless is the most likely state for the common ancestor to *Xiphophorus* and *Priapella*, but support for this estimate is not high (relative likelihoods for swordless:sword are 2.6:1; this diminishes to

exactly 1:1 when branch lengths are accounted for (see legend to Fig. 5). It seems unlikely that this phylogeny will allow a robust reconstruction for such a rapidly evolving trait.

These findings are tentative for two reasons. First, our analysis assumes that the rate of change is as high in the *Priapella* as in the *Xiphophorus*, whereas the rate of transition may be higher in the latter clade. Existing data are not sufficient to rule out a high rate within *Priapella*, but further data might. Second, the analysis assumes that "sword" is a single trait, whereas some evidence indicates that at least two developmentally and morphologically distinct types of sword occur in *Xiphophorus* (see Basolo 1995). If so, then the rate of transition between states for each sword trait analyzed separately (or for a single sword trait having three states instead of two) would be lower than indicated in Figure 5, leading to less uncertain ancestor states.

Continuous Traits

Specialists from Generalists.—A largely untested hypothesis of adaptive radiation, pervasive in the literature, is that ancestors of ecologically specialized species are usually generalists. One reason is that generalists may be more likely to colonize and persist in novel environments than specialists (e.g., Johnson et al. 1993). As well, competition among descendant species favors an increasingly fine partitioning of available resource types (Mayr 1942; MacArthur and Pianka 1966; Sugihara 1980). Finally, specialization may be a "dead-end" leading to extinction (Futuyma and Moreno 1988).

Phylogenetic methods could be used to test this hypothesis. One approach might be to classify each descendant species as a generalist or specialist, and then reconstruct ancestral states (generalist or specialist) using the methods for discrete traits presented earlier. Losos (1992b) used an alternative approach in Caribbean *Anolis* (Fig. 6). He took measurements of continuous morphological traits in contemporary species and reconstructed the morphology of ancestors using the method of squared change parsimony unweighted by branch lengths (Maddison 1991). Ancestral morphology was then used to classify ancestors as generalists or specialists. Below, we quantify the accuracy of this reconstruction.

Anolis is a perching lizard that hunts prey and defends territories from positions on erect stems. Species on the four large islands of the Greater Antilles (Jamaica, Hispaniola, Puerto Rico, and Cuba) are specialized in the types of perches used (Williams 1972). The majority of species fall into one of five discrete clusters called ecomorphs: crown-giant (using heavy stems in the upper canopy); twig (using canopy twigs); trunk-crown; trunk-ground; and grass-bush (Fig. 6, 7) (a sixth ecomorph, the trunk, is present only on Cuba and Hispaniola). The species in each cluster possess a common syndrome of morphology and behavior even though they have independent origins. For example, the trunk-ground ecomorph on all four islands is bulky and long-legged whereas the twig form is slim and short-legged.

We focus on the morphology of an ancestor just above the root in the tree of species from Puerto Rico (indicated in Fig. 6). The maximum-likelihood estimate of the morphology of this ancient anole is shown in Figure 7. This estimate is
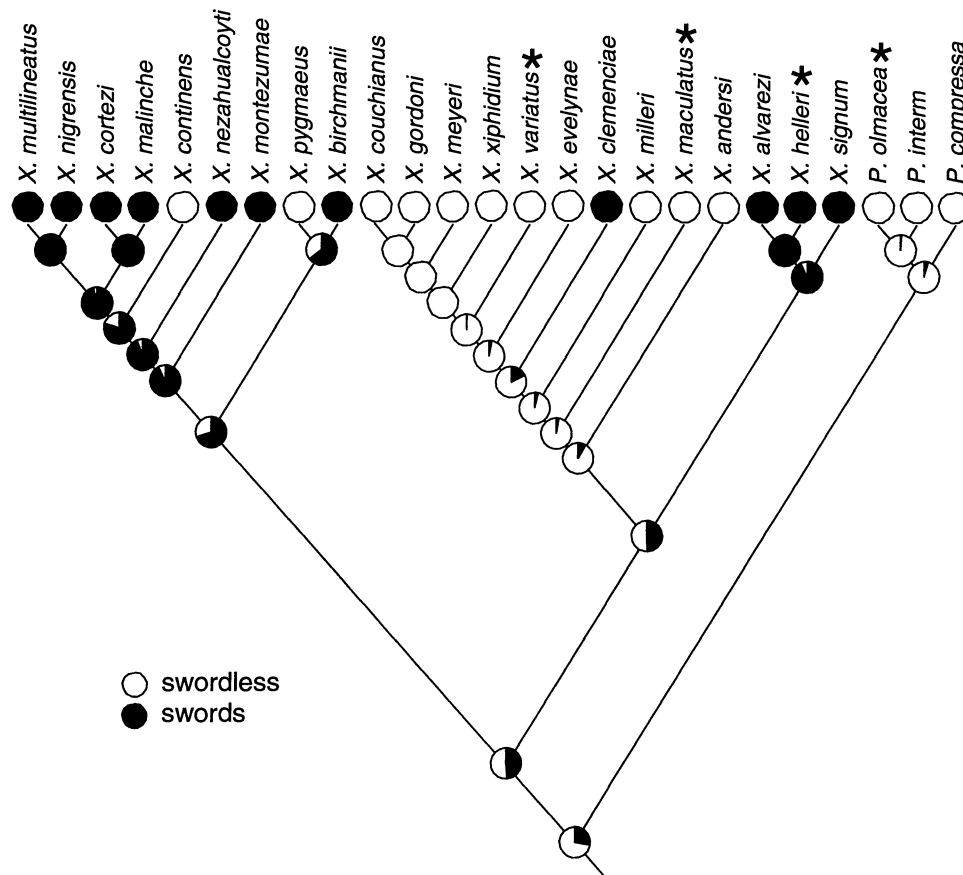
FIG. 5. Relative support for swords and swordlessness in the male ancestors of *Xiphophorus* (platies and swordtails) and *Priapella*, the sister taxon. The tree is from Figure 4 of Meyer et al. (1994), which was calculated from a maximum-parsimony majority- rule bootstrap analysis of sequences from three mtDNA regions (402 bp of the control region, 360 bp of cytochrome *b*, and 522 bp of *X-src*). All branch lengths were set to one. Both $\hat{q} = 0.15$. Results were similar when we used branch lengths obtained from DNAMLK of PHYLIP 3.5 (Felsenstein 1993) on the cytochrome *b* and *X-src* sequences, except that uncertainty in the oldest ancestor was even greater than that shown in the figure (relative likelihoods for swordlessness : sword were 1:1). We do not present the latter results because the likelihood of the DNAMLK tree, which forces tips to be contemporary, fit the sequence data e^{1780} worse than the same topology in which branch lengths are free to vary (DNAML). Hence a tree with contemporary tips cannot be obtained with these data. A female preference for males with swords has been demonstrated in all four species tested, indicated by an asterisk (Basolo 1990, 1995).

slightly different from that obtained by Losos (1992b) because it takes branch length into account. The most important feature of the reconstructed ancestor state is that it falls between the morphological states characterizing the ecomorphs. Consequently, Losos (1992b) hypothesized that this ancestor was a generalist. However, support limits on the estimated morphology of the second ancestor (equivalent to 95% confidence limits) are broad and overlap all the other ecomorphs (Fig. 7. This makes it difficult to rule out the possibility that the ancestor belonged to one of the specialist ecomorph categories of its descendants.

We interpret the *Anolis* result as follows. A maximum-likelihood estimate of an ancestor state for a continuous trait is simply a weighted average of the dimensions of the extant species at the tips of the tree (eq. [4]). Reconstructed ancestor states will tend to be intermediate for this reason. This intermediate value might correctly reveal the ancestor to have been a generalist. However, an intermediate estimate for an ancestor state may instead reflect our uncertainty over which specialist phenotype the ancestor actually possessed.

Diversification in Body Size.—A final example applies the continuous-trait method to all ancestors on a tree, to allow comparison of their relative uncertainty. We used the scrubwrens again, this time plotting estimates of the body sizes (wing lengths) of ancestral species (Fig. 8). Maximum likelihood suggests a series of ancestors of mainly intermediate size that repeatedly gave rise to both larger and smaller descendents. The variability of these estimates is, however, rather high (Fig. 8). Support limits for each node exceed the span of all species ultimately descended from it. This analysis suggests that ancestor reconstructions for continuous traits are often too variable to be of much use, except to place ancestor sizes within broad limits.

DISCUSSION

Ancestor reconstructions are estimates based on data from extant species. We have shown how the accuracy of estimates can be quantified using likelihood methods, which require a model of character state change. Our analysis of discrete traits shows that maximum-likelihood ancestor reconstructions are often similar to those obtained using maximum parsimony
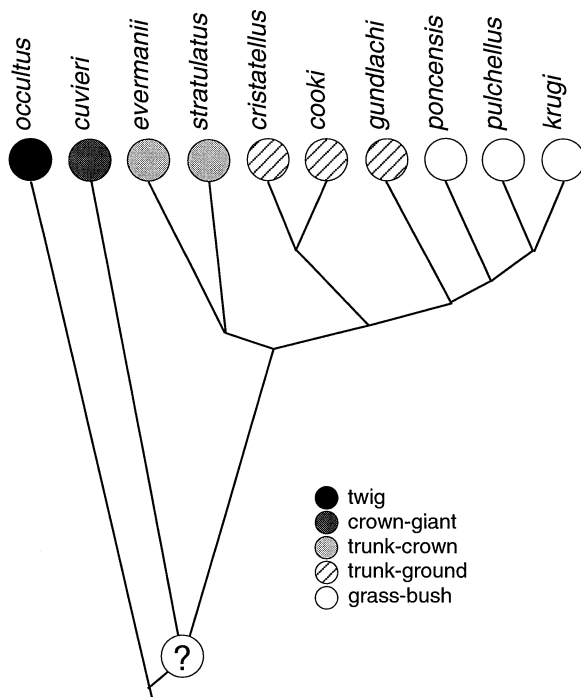
FIG. 6. Phylogeny and ecological state (ecomorphs) of Puerto Rican *Anolis* lizards. The unknown state of the second-oldest ancestor is indicated by the question mark. Tree topology and branch lengths are from Losos (1990a). A recent molecular phylogeny of the Caribbean anoles is not in full agreement with the tree shown, but the differences are only weakly supported in the DNA estimate (J. B. Losos, pers. comm. 1996).
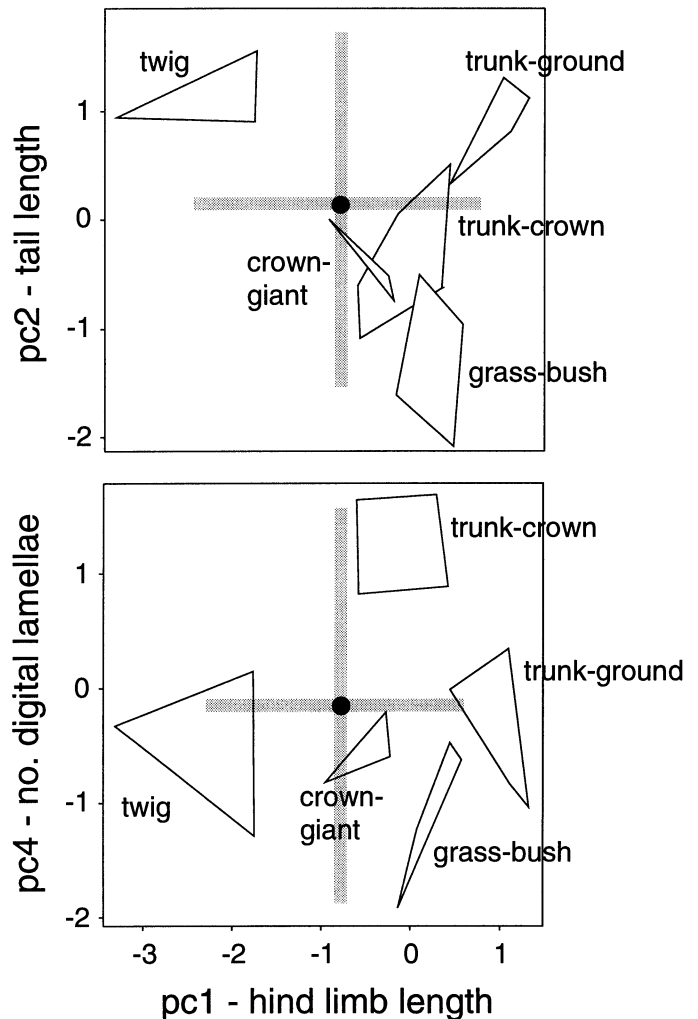

FIG. 7. Morphology of *Anolis* ecomorphs on Jamaica, Hispaniola, and Puerto Rico. Axes are principal component (PC) scores calculated as described in Losos (1992b). Axis labels indicate the traits mainly responsible for variation in PC1, PC2, and PC4; PC3 (not shown) is mainly a size factor separating the crown giants from the remaining four ecomorphs. Filled circle in the center of each plot is the maximum-likelihood estimate of morphology of the ancestor indicated in Figure 6. Shaded lines span the support limits for each estimate (equivalent to 95% confidence limits).

when character state change is estimated to be rare, and may be estimated with a high degree of certainty. However, parsimony and maximum-likelihood estimates can often differ even when change is rare, especially for reconstructions deep in the tree. In such cases, the uncertainty of the ancestor estimate is usually high. If change is estimated to have occurred frequently then ancestor reconstructions are subject to a high degree of uncertainty, especially those of more distant ancestors. This result should not be surprising. If the probability of change along branches is high then even single lineages may switch between states multiple times, making it impossible to know whether the lineage was in one state or the other even in the relatively recent past.

These results on uncertainty suggest that in many instances we cannot be confident of reconstructed ancestor states. This means that it will often be difficult to infer the timing of transitions in traits, particularly those in which change is not rare. It will also be difficult to infer the relative timing of transitions in suites of traits when change in any one of them is not rare. For example, swords appear to turn over much more rapidly than preferences for swords in the swordtail-platy group (cf. Fig. 5); gregariousness in butterfly larvae evolves rapidly compared with aposematic coloration (Sillén-Tullberg 1988). In such cases estimates of the order of transitions in the two traits is problematic because ancestor states for the fast trait will be uncertain. Measures of this uncertainty crucially affect interpretation. For example parsimony, because it underestimates the total number of state changes

on a tree (e.g., Maddison 1994), will tend to place transitions in the more rapidly evolving trait closer to the tree tips than those in the more slowly evolving trait. This effectively places an order on the timing of estimated transitions according to rates of character change. Incorporating uncertainty of distant ancestor states will reduce the risk of being misled.

Presently, ancestor states are regarded as uncertain in those cases in which the parsimony reconstruction is ambiguous (i.e., when several equally parsimonious reconstructions are found). Our examples show that ancestor states are often uncertain even when parsimony reconstructions are not ambiguous. Parsimony methods might illuminate some of this uncertainty by including in the range of acceptable alternatives reconstructions one or a few steps removed from the most parsimonious estimate. However, it is difficult to trans-
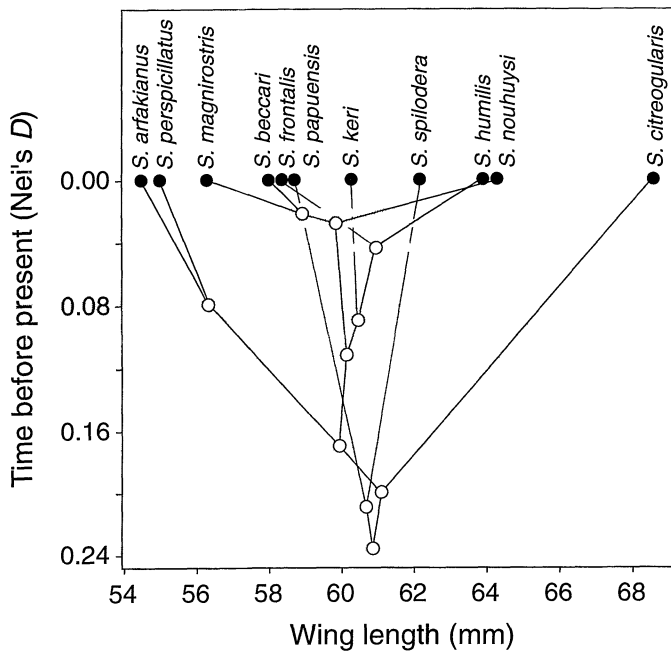
FIG. 8. Wing lengths of Australian and New Guinea scrubwrens (*Sericornis*) and of their ancestors. The tree is the same as in Figure 3. Shaded lines span the support limits for each ancestor state. Marginal distributions were computed separately for each ancestor one at a time.

late such a "parsimony interval" onto a consistent scale that allows comparison between trees or traits.

Likelihood methods quantify uncertainty with a standard probabilistic measure. They require a specific model of character evolution, and the models of choice are necessarily simple. Both the Markov model and the Brownian motion model assume that expected rates of character change are constant along every branch, and throughout time. This assumption cannot be true for phenotypic traits generally, but it may be a reasonable approximation. The need to specify an explicit model of character change in likelihood methods might seem to give parsimony an advantage. However, as we and others have shown (e.g., Frumhoff and Reeve 1994), accuracy of reconstruction is low unless the rate of change is low, a condition often violated by data. Quantifying this uncertainty requires a model of evolution.

Other models can be evaluated with simple modifications. For example, a model of punctuated evolution could be incorporated by setting all branch lengths equal to 1 (Martins and Garland 1991), as we have done in two of our examples (Figs. 2, 5). Alternatively, more complex models with more parameters could be devised. For our purposes, the most relevant issue is whether alternative models would generate states that are any less uncertain than those of the simple models. This can only be assessed with larger datasets. Often, limitations of the data will prevent rejection of simple models in favor of more complex models. When data are limited complex models can result in poor estimates of parameters and potentially misleading reconstructions, as illustrated by our example on Darwin's finches (Fig. 1).

A second weakness of the approach is that it conditions on rates of character change that are themselves estimated

with few data (species). This weakness could be remedied with realistic prior probability distributions for rate parameters. Such a procedure might be feasible for DNA sequences or amino acid compositions where previous surveys of changes between species provide realistic priors (e.g., Yang et al. 1995). Success of this approach for morphological characters is contingent on developing adequate priors. Uncertainty in estimates of ancestors is also magnified by factors we have not included in the model, such as errors in phylogeny estimation, and because only a subset (possibly nonrandom) of all the species of a clade are inevitably included.

Our results on uncertainty seem to lead to a pessimistic conclusion: that phylogenetic tests of ecological hypotheses are not possible. This does not necessarily follow. First, the problem is most serious when traits show high rates of change. Even in such cases at least some of the recent ancestors are reliable (e.g., Figs. 3, 4, 5). Second, many ecological hypotheses can be addressed without reconstruction of ancestor states. Standard comparative methods for testing correlated evolution in pairs of traits accomplish this (Felsenstein 1985; Pagel 1994). Pagel's (1994) method tests for a correlation between two discrete traits by fitting different models of trait evolution to species data and phylogenetic trees over the range of ancestor states allowed by the data. A similar approach tests whether the appearance of one trait depends on the prior appearance of another (Pagel 1994). Hypotheses of adaptive radiation and faunal buildup make explicit predictions about ancestors, and uncertainty may make them difficult to test. Yet, different hypothesis predict alternative models of character change that may nevertheless be tested by comparing their fits to data.

LITERATURE CITED

BASOLO, A. L. 1990. Female preference predates the evolution of the sword in swordtail fish. Science (Wash. DC.) 250:808–810.
———. 1995. Phylogenetic evidence for the role of a pre-existing bias in sexual selection. Proc. R. Soc. Lond., Ser. B Biol. Sci. 259:307–311.
BEINTEMA, J. J., W. M. FITCH, AND A. CARSANA. 1986. Molecular evolution of pancreatic- type ribonucleases. Mol. Biol. Evol. 3: 262–275.
BENNER, S. A., AND G. H. GONNET. 1995. Uncertainty in ancient phylogenies. Nature (Lond.) 377:109–110.
BERGER, J. O. 1985. Statistical decision theory and Bayesian analysis. Springer, Berlin.
CHRISTIDIS, L., R. SCHODDE AND P. R. BAVERSTOCK. 1988. Genetic and morphological differentiation and phylogeny in the Australo-Papuan scrubwrens (*Sericornis*, Acanthizidae). Auk 105:616–629.
DIAMOND, J. 1986. Evolution of ecological segregation in the New Guinea montane avifauna. Pp. 98–125 in J. Diamond and T. J. Case, eds. Community ecology. Harper and Rowe, New York.
EDWARDS, A. F. W. 1972. Likelihood. Cambridge Univ. Press, Cambridge.
FELSENSTEIN, J. 1973. Maximum likelihood and minimum steps methods for estimating evolutionary trees from data on discrete characters. Syst. Zool. 22:240–249.

———. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

———. 1993. PHYLIP 3.5. Distributed by author.

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.

FUTUYMA, D. J., AND G. MORENO. 1988. The evolution of ecological specialization. Annu. Rev. Ecol. Syst. 19:207–233.

GOLDMAN, N. 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. 36:182–198.

GRANT, B. R., AND P. R. GRANT. 1989. Evolutionary dynamics of a natural population: the large cactus finch of the Galápagos. Univ. of Chicago Press, Chicago.

GRANT, P. R. 1986. Ecology and evolution of Darwin's finches. Princeton Univ. Press, Princeton, NJ.

GUTTMAN, I., S. S. WILKS, AND J. S. HUNTER. 1982. Introductory engineering statistics. 3d ed. Wiley, Chichester, UK.

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

JERMANN, T. M., J. G. OPITZ, J. STACKHOUSE, AND S. A. BENNER. 1995. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. Nature (Lond.) 374:57–59.

JEFFREYS, H. 1961. The theory of probability. Oxford Univ. Press, Oxford.

JOHNSON, M. S., J. MURRAY, AND B. CLARKE. 1993. The ecological genetics and adaptive radiation of *Partula* on Moorea. Pp. 167–238 *in* D. J. Futuyma and J. Antonovics, eds. Oxford surveys in evolutionary biology. Vol. 9. Oxford Univ. Press, Oxford.

JOSEPH, L., AND C. MORITZ. 1993. Phylogeny and historical aspects of the ecology of eastern Australian scrubwrens *Sericornis* spp.—evidence from mitochondrial DNA. Mol. Ecol. 2:161–170.

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

LIEBHERR, J. K., AND A. E. HAJEK. 1990. A cladistic test of the taxon cycle and taxon pulse hypotheses. Cladistics 6:39–59.

LOSOS, J. B. 1990. Ecomorphology, peformance capability, and scaling of West Indian *Anolis* lizards: an evolutionary analysis. Ecol. Monogr. 60:369–388.

———. 1992a. A critical comparison of the taxon-cycle and character displacement models for size evolution of *Anolis* lizards in the Lesser Antilles. Copeia 1992:279–288.

———. 1992b. The evolution of convergent structure in Caribbean *Anolis* communities. Syst. Biol. 41:403–420.

———. 1996. Phylogenetic perspectives on community ecology. Ecology 77:1344–1354.

MACARTHUR, R. H., AND E. R. PIANKA. 1966. On optimal use of a patchy environment. Evolution 100:603–609.

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. 1991. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. Syst. Zool. 40:304–314.

———. 1995. Calculating the probability distribution of ancestral states reconstructed by parsimony on phylogenetic trees. Syst. Biol. 44:474–481.

MADDISON, W. P., AND D. R. MADDISON. 1992. MacClade. Vers. 3. Sinauer, Sunderland, MA.

MARTINS, E. P. 1996. Conducting phylogenetic comparative studies when the phylogeny is not known. Evolution 50:12–22.

MARTINS, E. P., AND T. GARLAND JR. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. Evolution 45:534–557.

MARTINS, E. P., AND T. F. HANSEN. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149:646–667.

MAYR, E. 1942. Systematics and the origin of species. Columbia Univ. Press, New York.

MEYER, A., J. MORRISSEY, AND M. SCHARTL. 1994. Recurrent origin of a sexually selected trait in *Xiphophorus* fishes inferred from a molecular phylogeny. Nature (Lond.) 368:539–542.

NEI, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590.

PAGEL, M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc. R. Soc. Lond. Ser. B Biol. Sci. 255:37–45.

RICHMAN, A. D., AND T. PRICE. 1992. Evolution of ecological differences in the Old World leaf warblers. Nature (Lond.) 355:817–821.

RICKLEFS, R. E., AND G. W. COX. 1972. Taxon cycles in the West Indian avifauna. Am. Nat. 106:195–219.

RONQVIST, F. 1994. Evolution of parasitism among closely related species: phylogenetic relationships and the origin of inquilinism in gall wasps (Hymenoptera: Cynipidae). Evolution 48:241–266.

RUMMEL, J. D., AND J. ROUGHGARDEN. 1985. A theory of faunal build-up for competition communities. Evolution 39:1009–1033.

ROUGHGARDEN, J., AND S. PACALA. 1989. Taxon cycle among Anolis lizard populations: review of the evidence. Pp. 403–442 *in* D. Otte and J. Endler, eds. Speciation and its consequences. Sinauer, Sunderland, MA.

RYAN, M. J., AND A. KEDDY-HECTOR. 1992. Directional patterns of female mate choice and the role of sensory biases. Am. Nat. 139(Suppl.):S4–S35.

SCHLUTER, D. 1986. Morphological and phylogenetic relations among the Darwin's finches. Evolution 38:921–930.

———. 1988. Character displacement and the adaptive divergence of finches on islands and continents. Am. Nat. 131:799–824.

———. 1995. Uncertainty in ancient phylogenies. Nature (Lond.) 377:108–109.

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

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

SIMPSON, G. G. 1944. Tempo and mode in evolution. Columbia Univ. Press, New York.

STERN, D. L., AND P. R. GRANT. 1996. A phylogenetic reanalysis of allozyme variation among populations of Galápagos finches. Zool. J. Linn. Soc. 118:119–134.

SUGIHARA, G. 1980. Minimal community structure: an explanation of species abundance patterns. Am. Nat. 116:770–787.

SUHONEN, J., R. V. ALATALO, AND L. GUSTAFSSON. 1994. Evolution of foraging ecology in Fennoscandian tits (*Parus* spp.). Proc. R. Soc. Lond. Ser. B Biol. Sci. 258:127–131.

WIENS, J. J., AND M. R. MORRIS. 1996. Character definitions, sexual selection and the evolution of swordtails. Am. Nat. 147:866–869.

WILLIAMS, E. E. 1972. The origin of faunas. Evolution of lizard congeners in a complex island fauna: a trial analysis. Evol. Biol. 6:47–89.

WILSON, E. O. 1961. The nature of the taxon cycle in the Melanesian ant fauna. Am. Nat. 95:169–193.

YANG, S. Y., AND J. L. PATTON. 1981. Genic variability and differentiation in the Galápagos finches. Auk 98:230–242.

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 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.

Corresponding Editor: E. Martins

APPENDIX

Here we derive the marginal distributions of the rate parameter β and the ancestral state parameters μ. This provides the basis for the results in equation (5). A general reference for the techniques and results presented below is Guttman et al. (1982).

From its definition, $Q(\mu)$ (eq. 4) is a non-negative quadratic form

in the parameters (μ) and the data (trait values for the species at the tips of the phylogeny). Regarded as a function of the parameters only, Q is a quadratic function. The maximum-likelihood estimate for μ will be denoted by $\hat{\mu}$: Q is minimized at $\hat{\mu}$. Since terms in Q of degree higher than 2 all vanish, we can take the Taylor series of Q to obtain

$$Q(\underset{\sim}{\mu}) = Q(\hat{\underset{\sim}{\mu}}) + \frac{1}{2} \sum_{i,i'} \frac{\partial^2 Q(\hat{\underset{\sim}{\mu}})}{\partial \mu_i \partial \mu_{i'}} (\mu_i - \hat{\mu}_i)(\mu_{i'} - \hat{\mu}_{i'}). \quad (A1)$$

This relationship may then be used in the likelihood function (eq. 3).

It is convenient to adopt a Bayesian approach to characterizing the uncertainty in parameter estimates. This approach regards the data as fixed and the parameters as unknown, in contrast to the frequentist viewpoint that the data are random and the parameters are fixed. It is necessary to adopt a prior density for the unknown parameters, which characterizes our initial information about them. Such priors may reflect characteristics of related groups. In the present cases, we assign a flat prior to the μ parameters, that is, a constant density from $-\infty$ to ∞. However, β must be positive (hence a flat prior is inappropriate), but $ln\,\beta$ ranges from $-\infty$ to ∞. Therefore we assign a flat prior to $ln\,\beta$, which is equivalent to assigning a prior proportional to $1/\beta$ to β. These choices are motivated in Jeffries (1961). With these assumptions the prior density is given by

$$P(\underset{\sim}{\mu}, \beta)d\mu_1 d\mu_2 \ldots d\mu_N d\beta = d\mu_1 \ldots d\mu_N \frac{d\beta}{\beta}. \quad (A2)$$

We show below that this choice of prior leads to results of the same form as frequentist results. According to Bayesian analysis, the posterior density (which describes our knowledge of the parameters taking account of the data) is given by the product of the likelihood and the prior density:

$$\Pi(\underset{\sim}{\mu}, \beta) = L(\underset{\sim}{\mu}, \beta) \cdot P(\underset{\sim}{\mu}, \beta). \quad (A3)$$

Marginal Distribution of β

The posterior density for β may be obtained by integrating over all possible values of the ancestral state parameters (μ_i). We call this the marginal density for β. It is given by

$$D_\mu(\beta)d\beta$$

$$= \int_{\mu_1=-\infty}^{\infty} \ldots \int_{\mu_N=-\infty}^{\infty} L(\underset{\sim}{\mu}, \beta)P(\underset{\sim}{\mu}, \beta) \, d\mu_1 \ldots d\mu_N \frac{d\beta}{\beta}. \quad (A4)$$

Two intermediate results are required for evaluation of $D_\mu(\beta)$.

The first step is to convert the sum in equation (A1) to a sum of squares. This is achieved by using the eigenvectors of the matrix of second derivatives of Q (here denoted by $\mathbf{A}$), which produces the principal axes decomposition. Let $\mathbf{v}_j$, $j = 1, \ldots, N$, denote the normalized eigenvectors of $\mathbf{A}$, and let κ_j be the corresponding eigenvalues. Then

$$\mathbf{A}\mathbf{v}_j = \kappa_j \mathbf{v}_j. \quad (A5)$$

Let the matrix $\mathbf{U}$ have these eigenvectors as columns. Since $\mathbf{A}$ is symmetric, it follows that $\mathbf{U}$ is orthogonal, that is, $\mathbf{U}^T\mathbf{U}$ is the identity matrix, and $\mathbf{U}^T = \mathbf{U}^{-1}$. It follows from equation (A5) that

$$\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{D}, \quad (A6)$$

where $\mathbf{D}$ is a diagonal matrix with $\kappa_1, \ldots, \kappa_N$ on the diagonal. Given any matrix of values $\underset{\sim}{\mu}$, let the vector $\mathbf{w}$ be given by

$$\mathbf{w} = \mathbf{U}^T(\underset{\sim}{\mu} - \hat{\underset{\sim}{\mu}}). \quad (A7)$$

It follows that

$$\sum_{i,i'} \frac{\partial^2 Q(\hat{\underset{\sim}{\mu}})}{\partial \mu_i \partial \mu_{i'}} (\mu_i - \hat{\mu}_i)(\mu_{i'} - \hat{\mu}_{i'})$$

$$= (\underset{\sim}{\mu} - \hat{\underset{\sim}{\mu}})^T \mathbf{A}(\underset{\sim}{\mu} - \hat{\underset{\sim}{\mu}}) = \mathbf{w}^T \mathbf{D} \mathbf{w}. \quad (A8)$$

The second equality in equation (A8) follows from the substitution $\underset{\sim}{\mu} - \hat{\underset{\sim}{\mu}} = \mathbf{U}\mathbf{w}$ and (A6). Now the integration over $\underset{\sim}{\mu}$ takes the form

$$D_\mu(\beta)d\beta$$

$$= \int_{w_1=-\infty}^{\infty} \ldots$$

$$\int_{w_N=-\infty}^{\infty} \frac{1}{\beta^N} \exp\left(-\frac{1}{2}Q[\hat{\mu}] - \frac{1}{4\beta} \sum_{i=1}^N \kappa_i w_i^2\right) dw_1 \ldots dw_N \frac{d\beta}{\beta}. \quad (A9)$$

The second step in the calculation of the marginal density of β is to evaluate the integral of a single squared term of the form just obtained. This is given by the formula

$$\int_{-\infty}^{\infty} \exp\left(-\frac{\kappa w^2}{4\beta}\right) dw = \sqrt{\frac{4\pi\beta}{\kappa}}. \quad (A10)$$

This result may be applied to each integration over w. Provided that none of the κ_j are zero, the result is

$$D_\mu(\beta)d\beta \propto \exp\left(-\frac{Q[\hat{\mu}]}{2\beta}\right)\frac{1}{\beta^{1+N/2}}d\beta. \quad (A11)$$

This density must be normalized so that its integral over β from 0 to ∞ is unity. Hence factors in equation (A9) or (A10) that do not depend upon β may be ignored. If we set $u = Q(\hat{\mu})/\beta$ and introduce u as variable of integration in (A11), the resulting density is

$$D_\mu(\beta)d\beta \propto \exp(-u/2)u^{N/2}du/u. \quad (A12)$$

Hence u has a χ^2 distribution with N degrees of freedom (Jeffries 1961).

A corresponding frequentist analysis would lead to a similar result but a different interpretation: $Q(\hat{\mu})/N$ is an unbiased estimator for the "true" rate parameter β, and the sampling distribution for $u = Q(\hat{\mu})/\beta$ is χ^2 with N degrees of freedom. This second interpretation sees the observed statistic u as one outcome of a sampling process that could be repeated many times on new random samples of data (i.e., on new trait values at the tips of the same tree). The Bayesian interpretation is that the marginal distribution subjectively describes the various plausible values of the unknown parameter β, given the single set of data and a prior distribution for the parameter.

Marginal Distribution of Character Parameters

A similar analysis leads to the marginal distributions for the ancestral states, describing the uncertainty in characters in a subset of p ancestral nodes. Denote the subset in question by $I = \{i_1, \ldots i_p\}$, and let the remaining nodes belong to set $M = \{m_1, \ldots m_{N-p}\}$. We shall denote indices in I by i and j, and we shall denote indices in M by m and n. We wish to compute the joint marginal distribution at nodes in I. Denote the corresponding $\{\mu_i, i \in I\}$ by μ^*. The marginal density for μ^* is defined by

$$D_M(\mu^*)d\mu_{i_1} \ldots d\mu_{i_p}$$

$$= \int_{\beta=0}^{\infty} \int_{\mu_{m_1}=-\infty}^{\infty} \ldots$$

$$\int_{\mu_{m_{N-p}}=-\infty}^{\infty} L(\underset{\sim}{\mu}, \beta)P(\underset{\sim}{\mu}, \beta) \, d\mu_{m_1} \ldots d\mu_{m_{N-p}} d\beta. \quad (A13)$$

To evaluate D_M, we again write $Q(\mu)$ as a sum of squares. For any fixed value of μ^*, the maximum-likelihood estimates of the remaining μ_m, $m \in M$ will be denoted by $g_m(\mu^*)$, and the vector of all such functions $g_m(\mu^*)$ will be denoted by $\mathbf{g}(\mu^*)$. Thus

$$\frac{\partial Q}{\partial \mu_m}(\mathbf{g}[\mu^*], \mu^*) \equiv 0, \qquad m \in M. \quad (A14)$$

It follows from differentiation of equation (A14) with respect to μ_i that

$$\sum_n \frac{\partial^2 Q}{\partial \mu_m \partial \mu_n}(\mathbf{g}[\mu^*], \mu^*)\frac{\partial g_n(\mu^*)}{\partial \mu_i} + \frac{\partial^2 Q(\mathbf{g}[\mu^*], \mu^*)}{\partial \mu_m \partial \mu_i} = 0. \quad (A15)$$

For each value of i, the system (A15) can be solved for $\partial g_m / \partial \mu_i$, $m \in M$. Let

$$R(\mu^*) = Q(\mathbf{g}[\mu^*], \mu^*). \qquad (A16)$$

It follows from equation (A14) that

$$\frac{\partial R}{\partial \mu_i}(\mu^*) \equiv \frac{\partial Q}{\partial \mu_i}(\mathbf{g}[\mu^*], \mu^*). \qquad (A17)$$

A differentiation of equation (A17) with respect to μ_j produces

$$\frac{\partial^2 R(\mu^*)}{\partial \mu_i \mu_j} = \sum_m \frac{\partial^2 Q}{\partial \mu_i \partial \mu_m}(\mathbf{g}[\mu^*], \mu^*)\frac{\partial g_m(\mu^*)}{\partial \mu_j}$$
$$+ \frac{\partial^2 Q}{\partial \mu_i \partial \mu_j}(\mathbf{g}[\mu^*], \mu^*). \qquad (A18)$$

In view of equation (A15), equation (A18) can be rewritten as

$$\frac{\partial^2 R(\mu^*)}{\partial \mu_i \mu_j} = -\sum_{m,n} \frac{\partial^2 Q}{\partial \mu_m \partial \mu_n}(\mathbf{g}[\mu^*], \mu^*)\frac{\partial g_m}{\partial \mu_i}\frac{\partial g_n}{\partial \mu_j}$$
$$+ \frac{\partial^2 Q}{\partial \mu_i \partial \mu_j}(\mathbf{g}[\mu^*], \mu^*). \qquad (A19)$$

The first term in equation (A19) is always negative: it corresponds to a decrease in the information that the data imply about μ^*, since the parameters μ_m must be estimated at the same time as μ^*. A Taylor series for Q takes the form

$$Q(\mathbf{g}[\mu^*], \mu^*)$$
$$= Q(\mathbf{g}[\hat{\mu}^*], \hat{\mu}^*) + \frac{1}{2}\sum_{i,j} \frac{\partial^2 R}{\partial \mu_i \partial \mu_j}(\hat{\mu}^*)(\mu_i - \hat{\mu}_i)(\mu_j - \hat{\mu}_j)$$
$$+ \frac{1}{2}\sum_{m,n} \frac{\partial^2 Q}{\partial \mu_m \partial \mu_n}(\mathbf{g}[\hat{\mu}^*], \hat{\mu}^*)(\mu_m - g_m[\mu^*])$$
$$\times (\mu_n - g_n[\mu^*]). \qquad (A20)$$

When equation (A18) is substituted into equation (A13), the second sum in equation (A20) may be reduced to a sum of squares and integrated as in (A5–A11) to produce a factor of $\beta^{(N-p)/2}$. Thus the marginal density $D_M(\mu^*)$ takes the form

$$D_M(\mu^*)d\mu_{i_1} \ldots d\mu_{i_p}$$
$$\propto \int_0^\infty \exp\left(-\frac{B}{2\beta}\right)\frac{1}{\beta^{1+(N+p)/2}} \, d\beta \, d\mu_{i_1} \ldots d\mu_{i_p}. \qquad (A21)$$

Here

$$B = Q(\mathbf{g}[\hat{\mu}^*], \hat{\mu}^*) + \frac{1}{2}\sum_{i,j} \frac{\partial^2 R}{\partial \mu_i \partial \mu_j}(\hat{\mu}^*)(\mu_i - \hat{\mu}_i)(\mu_j - \hat{\mu}_j). \qquad (A22)$$

To integrate over the distribution of β the following result is required (Jeffries 1961):

$$\int_0^\infty \exp(-B/\beta)\frac{d\beta}{\beta^{1+m}} = \frac{\Gamma(m)}{B^m}. \qquad (A23)$$

This result follows from a change of variables in the definition of the gamma function. Now equation (A21) implies that

$$D_M(\mu^*) \propto \frac{1}{B^{(N+p)/2}}d\mu^*. \qquad (A24)$$

The p-variate t-distribution with N degrees of freedom has the density

$$f(\mu^*) \propto \left(1 + \frac{1}{N}\sum_{i,j=1}^p V_{i,j}^{-1}[\mu_i - \hat{\mu}_i][\mu_j - \hat{\mu}_j]\right)^{(N+p)/2}, \qquad (A25)$$

where $NV/(N-2)$ is the covariance matrix of the distribution (Berger 1985, p. 561). The marginal density D_M as given in equation (A24) has this form if

$$V^{-1} = \left(\frac{\partial^2 R}{\partial \mu_i \partial \mu_j}\right)\frac{N}{2Q(\mathbf{g}[\hat{\mu}^*], \hat{\mu}^*)}. \qquad (A26)$$

If we are considering only one ancestor at a time (i.e., $p = 1$) we may set

$$C = \sqrt{\frac{2Q(\mathbf{g}[\hat{\mu}^*], \hat{\mu}^*)}{Nd^2 Q(\mathbf{g}[\hat{\mu}^*], \hat{\mu}^*)/d\mu^{*2}}}. \qquad (A27)$$

Then $t = (\mu^* - \hat{\mu}^*)/C$ has a univariate t-distribution with N degrees of freedom.

As in the case of the marginal distribution of β, the frequentist results would be formally the same as the Bayesian results. In a frequentist analysis the maximum-likelihood estimates for $\mu_1, \ldots \mu_N$ and β are obtained, and the sampling distribution of t is t with N degrees of freedom.

Our estimates of ancestor states for discrete traits were conditional on the maximum-likelihood estimates of transition rates (eq. 2), whereas here we integrated over the posterior distribution of β when computing marginal distributions of ancestral states for continuous traits (eq. A21). Conditioning on the unbiased estimate of β (i.e., $\hat{\beta}$) modifies the analysis as follows.

The maximum-likelihood estimate for β is $Q(\hat{\mu})/2N$, which is obtained by setting the derivative of

$$L_\mu(\beta) \propto \exp\left(\frac{Q[\hat{\mu}]}{2\beta}\right)\frac{1}{\beta^N} \qquad (A28)$$

with respect to β equal to zero and solving for $\hat{\beta}$. Equation (A28) is obtained from equation (A11) after removing the factor $1/\beta$, which corresponds to the prior and hence isn't part of the likelihood. This estimate is biased because there are $2N$ squared terms in $(\hat{\mu})$, whereas only N parameters have been estimated. The corresponding unbiased estimate of β is $\hat{\beta} = Q(\hat{\mu})/N$. When this result is used in equation (A21), the result is

$$D_M(\mu^*) \propto \exp\left(\frac{-NB}{2Q[\hat{\mu}]}\right)d\mu^*. \qquad (A29)$$

But

$$\exp\left(\frac{-NB}{2Q[\hat{\mu}]}\right) \propto \exp\left(-\frac{[\mu^* - \hat{\mu}^*]^2}{2C^2}\right). \qquad (A30)$$

This implies that μ^* has a normal distribution with mean $\hat{\mu}^*$ and variance C^2. In this case $t = (\mu^* - \hat{\mu}^*)/C$ has a standard normal distribution. The t-distribution derived for this quantity in equations (A24–A26) corresponds to the correction for the normal distribution introduced by Student for small sample sizes.