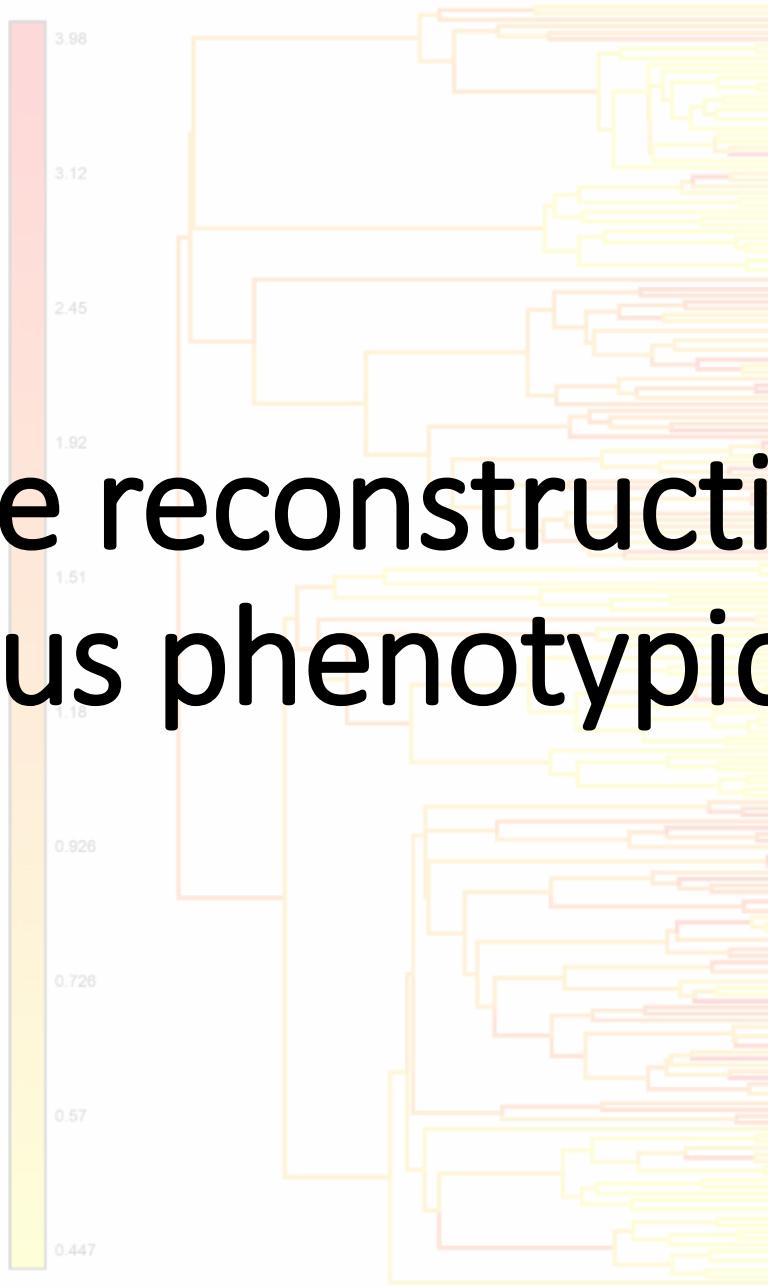


a) generating process



b) fitted Γ model



c) fitted hidden-rates model



Ancestral state reconstruction of discrete
& continuous phenotypic characters

But perhaps the most direct question we can ask about ancestral species is simply, what were they like? The endeavor of answering this question using a phylogenetic tree is called **ancestral state reconstruction** (reviewed in Felsenstein 2004; Harmon 2019), and **for better or for worse**, the estimation of ancestral traits has long been, and continues to be, an important goal in phylogenetic comparative biology.

microhabitat evolution affected tail spine evolution as well as the converse, along with one in which the two characters evolved independently from each other. Use both likelihood ratio tests and AIC to compare among your different fitted models. Both species can be found exclusively in forested, forest-fringe, or open habitats. Many butterflies are numerous, more generalized species are found in two or all three habitat types. Use *Estpolynk* to fit a polymorphic trait evolution model to the evolution of habitat type in *Mycelisina* assuming both ordered character evolution (that is, forest → forest fringe → open) or unordered polymorphic trait evolution. The two files you need (*Mycelisina.phylogeny.nex* and *Mycelisina_habitat.csv*) are modified from Halali et al. (2020) and can be downloaded from the book website.

Reconstructing ancestral states

8.1 Introduction

Many phylogenetic comparative methods purport to give us some insight about the evolutionary past. For instance, in chapters 2, 3, and 7, we learned how to ask if the traits under study (continuous or discrete) tended to precipitate in certain ways over time.

8

1. Examine the statistical properties of maximum likelihood estimation of ancestral states for a continuous trait using numerical simulation: both when our assumed model is correct as well as when it is wrong.
2. Examine the statistical properties of maximum likelihood estimation of ancestral states for a categorical trait using maximum likelihood ratio tests and AIC to compare among different fitted models. Use 95 percent confidence intervals around these estimates.

¹We prefer to use the term *estimation* over *reconstruction*, because the former emphasizes its statistical (and thus probabilistic) nature. Nevertheless, reconstruction is widely used, so we employ both words more or less interchangeably here.

²Although it's possible—and can be very useful—to include data from fossils.

What is ancestral state reconstruction?

- Ancestral state reconstruction (aka. *ancestral character estimation* & variants thereof) is the endeavor to estimate the values of a phenotype from species data related by a phylogenetic tree.
- Ancestral state reconstruction has long been one of the most popular phylogenetic methods.
- This is pretty easy to comprehend as it purports to offer evolutionary biologists (in many ways, the *historians of life*) a **window on the past**.

Discrete vs. continuous characters

- Ancestral state reconstruction can be undertaken by both discretely and continuously valued phenotypic traits.
- A continuously-valued trait might be body mass, limb lengths, thermal tolerances, genome size, etc.
- Today, we'll start with **discrete characters**. These could include a presence/absence feature, a meristic (counted) trait, a qualitative attribute, etc.

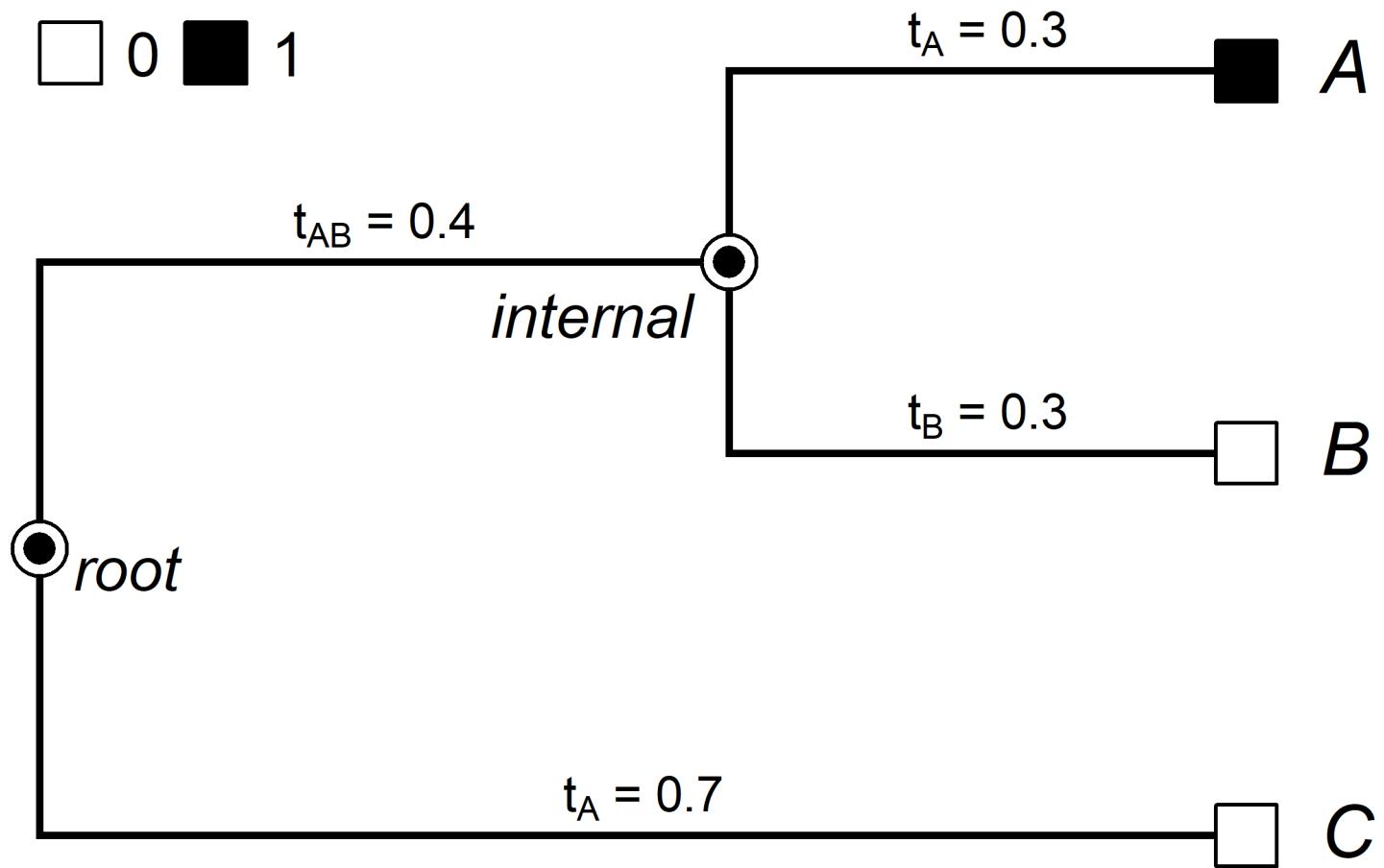
Ancestral state reconstruction

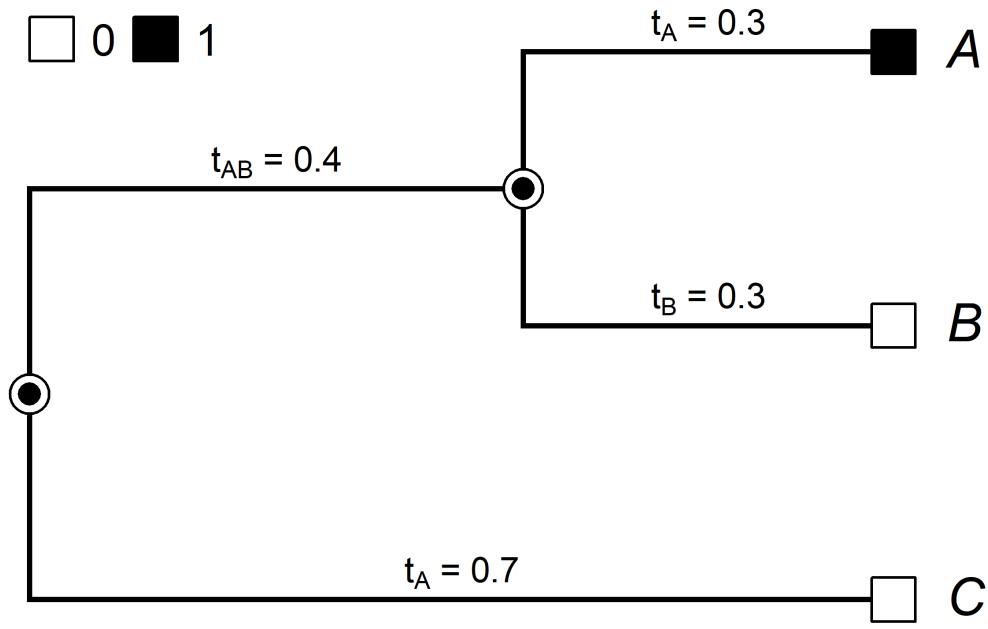
- Here, I'm going to focus on ancestral state reconstruction based on explicit mathematical models for how traits evolve (not, say, parsimony).
- This is not because parsimony can't give us good estimates of ancestral states: sometimes it does.
- This is because absent an explicit model we have no way to assess the strength of evidence in support of our reconstruction nor compare it to alternative scenarios.

The Mk model

- Overwhelmingly, the predominant model used to approximate discrete character evolution on phylogenetic trees (and thus to reconstruct ancestral states) is one called the Mk (or, sometimes, the ‘extended Mk’) model.
- The Mk model is a continuous-time, k -state, **Markov** process in which the instantaneous transition rates between character levels are given in the $k \times k$ transition matrix, **Q**.
- To compute the probability of starting some time interval t in one condition & ending in the same or other, we calculate $\exp(\mathbf{Q} \times t)$.

Computing the probability





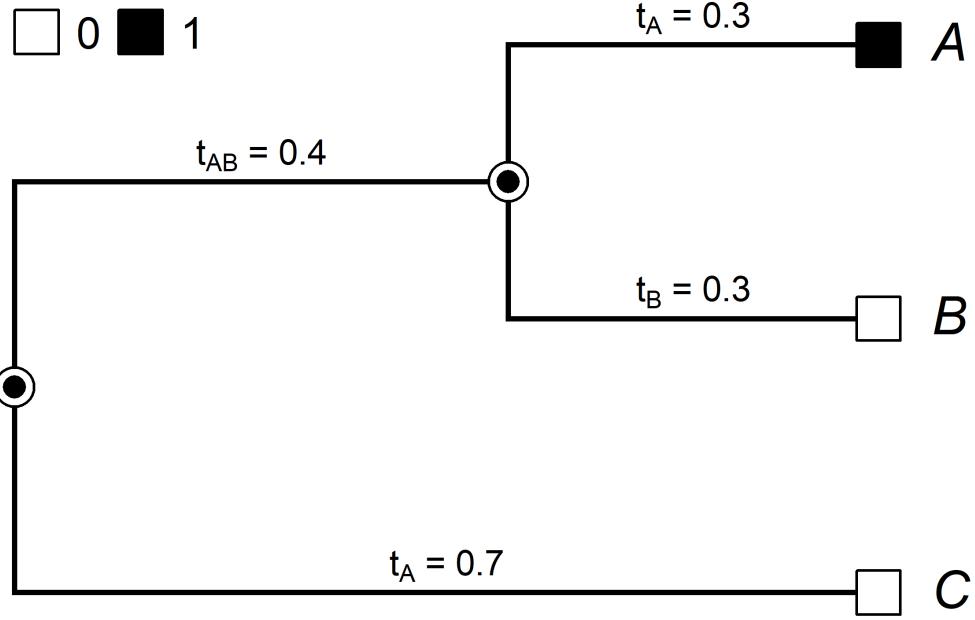
$$\mathbf{Q} = \begin{bmatrix} -q_{0,1} & q_{0,1} \\ q_{1,0} & -q_{1,0} \end{bmatrix} = \begin{bmatrix} -0.2 & 0.2 \\ 0.2 & -0.2 \end{bmatrix}$$

$$\mathbf{P}_t = \exp(\mathbf{Q}t)$$

$$\mathbf{P}_{t=0.4} = \exp(\mathbf{Q} \times 0.4) = \begin{bmatrix} 0.926 & 0.074 \\ 0.074 & 0.926 \end{bmatrix}$$

$$\mathbf{P}_{t=0.3} = \exp(\mathbf{Q} \times 0.3) = \begin{bmatrix} 0.943 & 0.057 \\ 0.057 & 0.943 \end{bmatrix}$$

$$\mathbf{P}_{t=0.7} = \exp(\mathbf{Q} \times 0.7) = \begin{bmatrix} 0.878 & 0.122 \\ 0.122 & 0.878 \end{bmatrix}$$

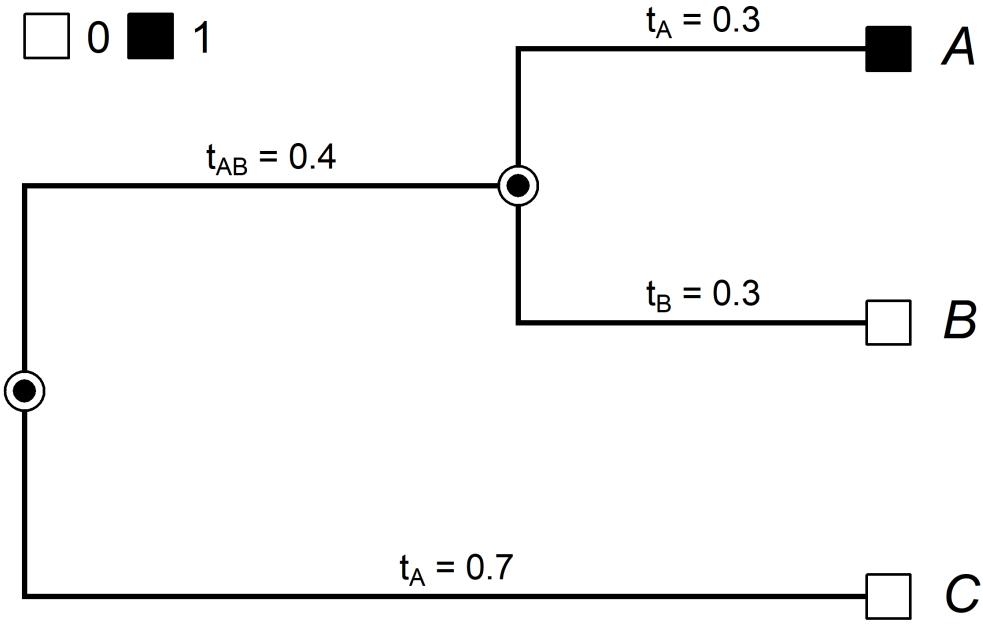


$$\mathbf{P}_{t=0.4} = \exp(\mathbf{Q} \times 0.4) = \begin{bmatrix} 0.926 & 0.074 \\ 0.074 & 0.926 \end{bmatrix}$$

$$\mathbf{P}_{t=0.3} = \exp(\mathbf{Q} \times 0.3) = \begin{bmatrix} 0.943 & 0.057 \\ 0.057 & 0.943 \end{bmatrix}$$

$$\mathbf{P}_{t=0.7} = \exp(\mathbf{Q} \times 0.7) = \begin{bmatrix} 0.878 & 0.122 \\ 0.122 & 0.878 \end{bmatrix}$$

$$P(1, 0, 0) =$$



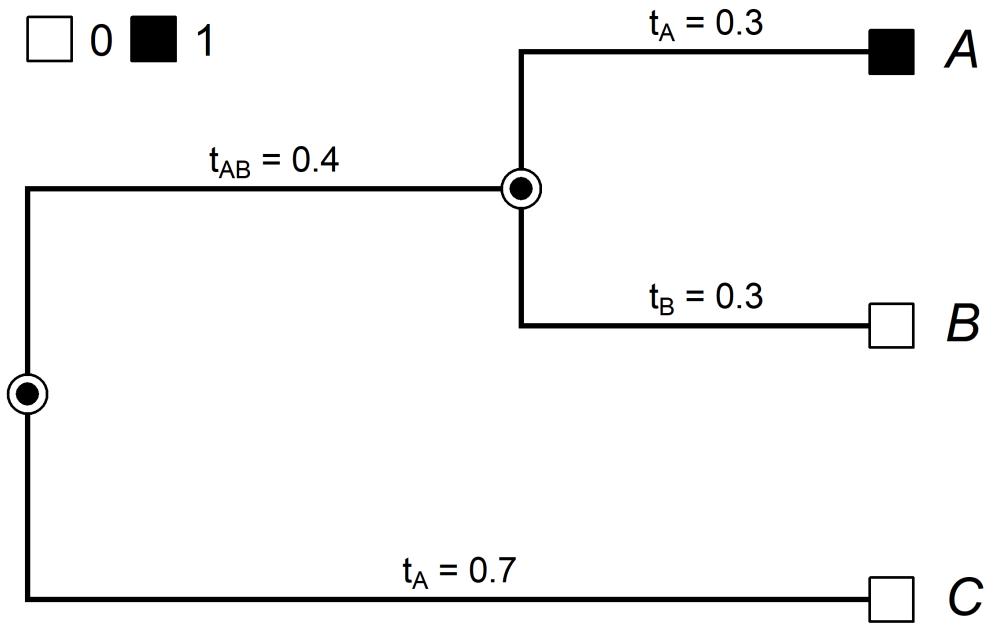
$$\mathbf{P}_{t=0.4} = \exp(\mathbf{Q} \times 0.4) = \begin{bmatrix} 0.926 & 0.074 \\ 0.074 & 0.926 \end{bmatrix}$$

$$\mathbf{P}_{t=0.3} = \exp(\mathbf{Q} \times 0.3) = \begin{bmatrix} 0.943 & 0.057 \\ 0.057 & 0.943 \end{bmatrix}$$

$$\mathbf{P}_{t=0.7} = \exp(\mathbf{Q} \times 0.7) = \begin{bmatrix} 0.878 & 0.122 \\ 0.122 & 0.878 \end{bmatrix}$$

$$P(1, 0, 0) = \pi_0 \times P(0|0, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|0, t_C)$$

0 1



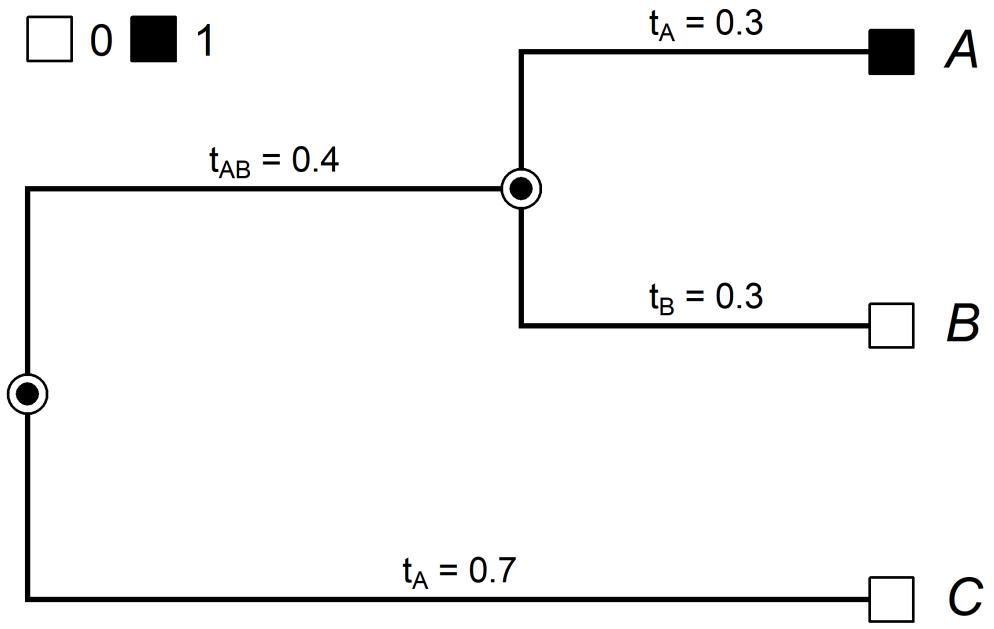
$$\mathbf{P}_{t=0.4} = \exp(\mathbf{Q} \times 0.4) = \begin{bmatrix} 0.926 & 0.074 \\ 0.074 & 0.926 \end{bmatrix}$$

$$\mathbf{P}_{t=0.3} = \exp(\mathbf{Q} \times 0.3) = \begin{bmatrix} 0.943 & 0.057 \\ 0.057 & 0.943 \end{bmatrix}$$

$$\mathbf{P}_{t=0.7} = \exp(\mathbf{Q} \times 0.7) = \begin{bmatrix} 0.878 & 0.122 \\ 0.122 & 0.878 \end{bmatrix}$$

$$\begin{aligned} P(1, 0, 0) = & \pi_0 \times P(0|0, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|0, t_C) \\ & + \pi_0 \times P(1|0, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|0, t_C) \\ & + \pi_1 \times P(0|1, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|0, t_C) \\ & + \pi_1 \times P(1|1, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|1, t_C) \end{aligned}$$

0 1



$$\mathbf{P}_{t=0.4} = \exp(\mathbf{Q} \times 0.4) = \begin{bmatrix} 0.926 & 0.074 \\ 0.074 & 0.926 \end{bmatrix}$$

$$\mathbf{P}_{t=0.3} = \exp(\mathbf{Q} \times 0.3) = \begin{bmatrix} 0.943 & 0.057 \\ 0.057 & 0.943 \end{bmatrix}$$

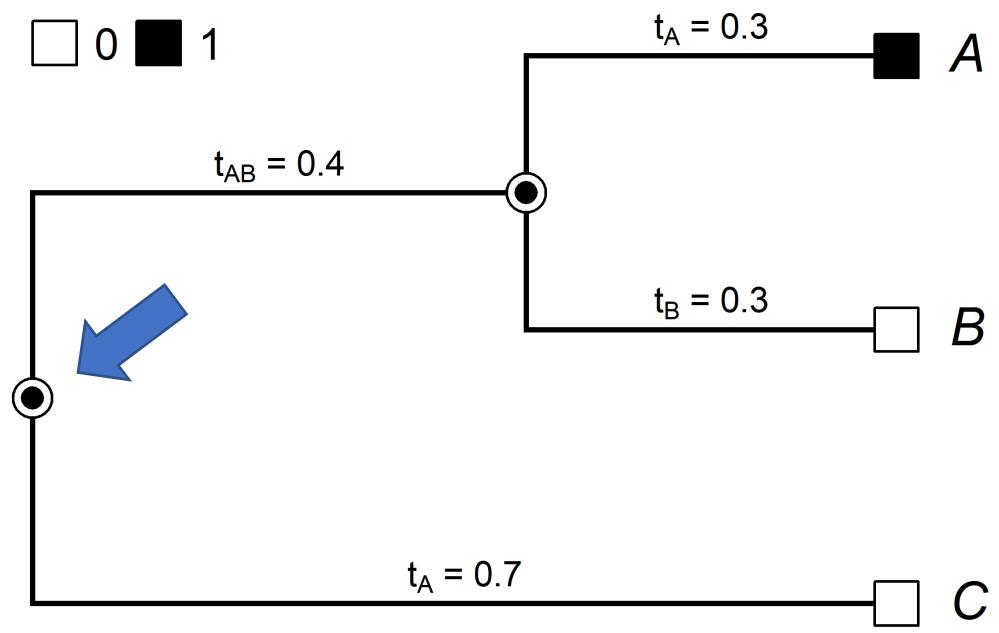
$$\mathbf{P}_{t=0.7} = \exp(\mathbf{Q} \times 0.7) = \begin{bmatrix} 0.878 & 0.122 \\ 0.122 & 0.878 \end{bmatrix}$$

$$\begin{aligned} P(1, 0, 0) &= \pi_0 \times P(0|0, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|0, t_C) \\ &\quad + \pi_0 \times P(1|0, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|0, t_C) \\ &\quad + \pi_1 \times P(0|1, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|0, t_C) \\ &\quad + \pi_1 \times P(1|1, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|1, t_C) \\ &= 0.0267 \end{aligned}$$

Marginal Reconstruction

- What is known as *marginal ancestral state reconstruction* under the Mk model involves going to each node, successively fixing each state, and then recomputing the probability of our data conditioned on that state.
- These values are normally rescaled to sum to 1.0 – which simply involves dividing their value by the total likelihood.
(Consequently, they're often called *marginal scaled likelihoods*.)
- Yang (2006) also pointed out that marginal scaled likelihoods are a type of *empirical Bayes posterior probability*.

0 1



$t_A = 0.3$

A

$t_{AB} = 0.4$

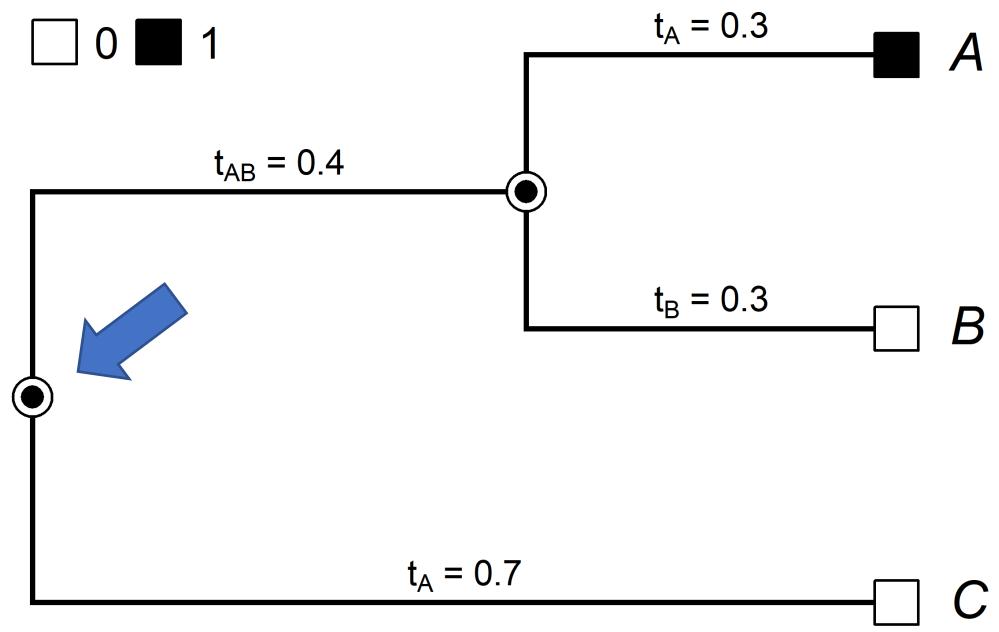
$t_B = 0.3$

B

$t_A = 0.7$

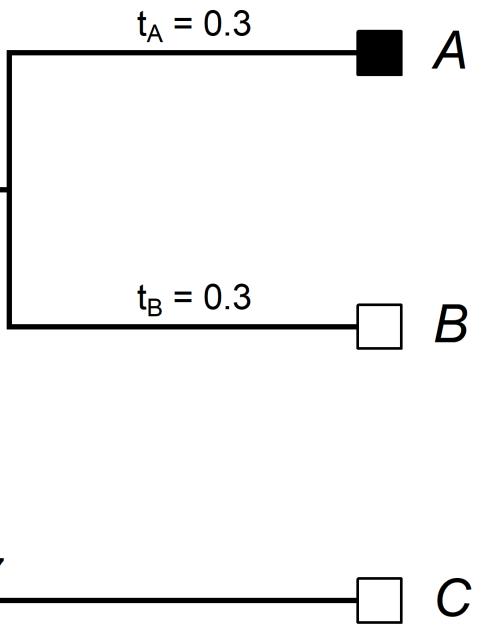
C

0 1



$$\begin{aligned} P(\text{root} = 0) &= \pi_0 \times P(0|0, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|0, t_C) \\ &\quad + \pi_0 \times P(1|0, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|0, t_C) \\ &= 0.5 \times 0.0434 + 0.5 \times 0.0035 \\ &= 0.0234 \end{aligned}$$

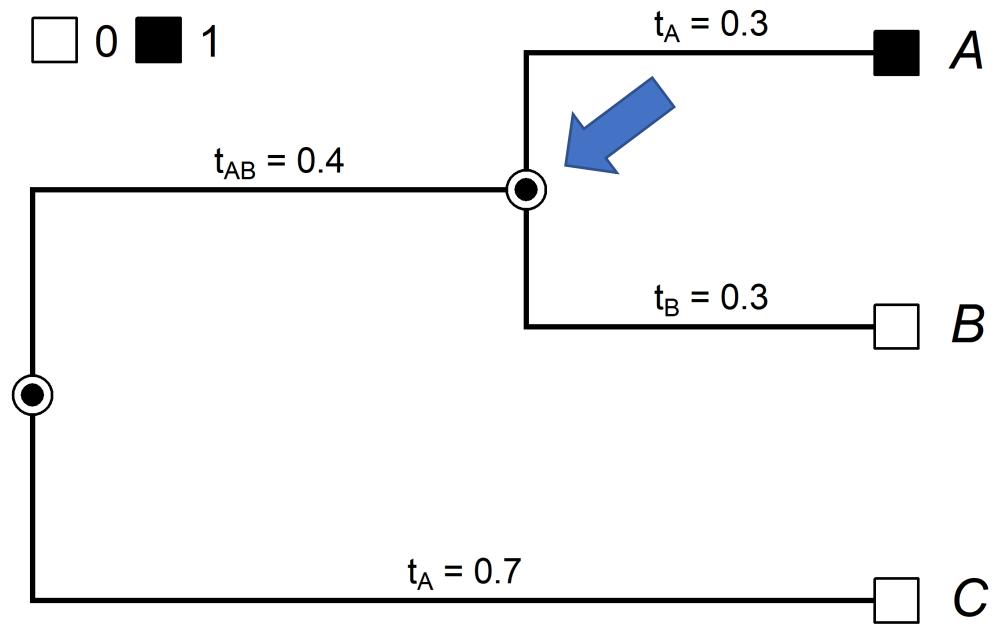
0 1



$$\begin{aligned} P(\text{root} = 0) &= \pi_0 \times P(0|0, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|0, t_C) \\ &\quad + \pi_0 \times P(1|0, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|0, t_C) \\ &= 0.5 \times 0.0434 + 0.5 \times 0.0035 \\ &= 0.0234 \end{aligned}$$

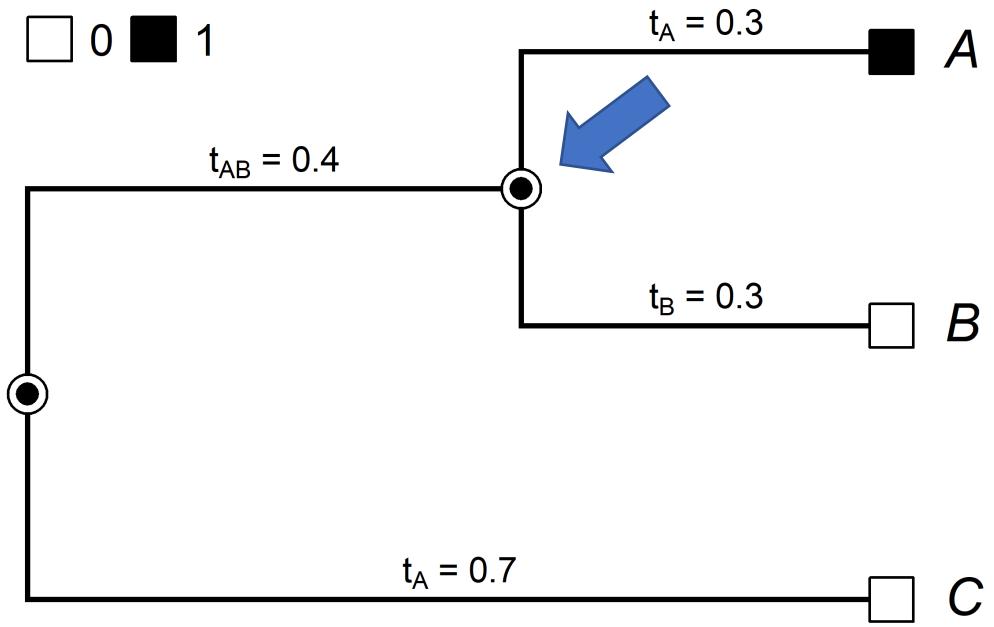
$$\begin{aligned} P(\text{root} = 1) &= \pi_1 \times P(0|1, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|1, t_C) \\ &\quad + \pi_1 \times P(1|1, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|1, t_C) \\ &= 0.5 \times 0.0005 + 0.5 \times 0.0060 \\ &= 0.0033 \end{aligned}$$

0 1



$$\begin{aligned} P(internal = 0) &= \pi_0 \times P(0|0, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|0, t_C) \\ &\quad + \pi_1 \times P(0|1, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|1, t_C) \\ &= 0.5 \times 0.0434 + 0.5 \times 0.0005 \\ &= 0.0219 \end{aligned}$$

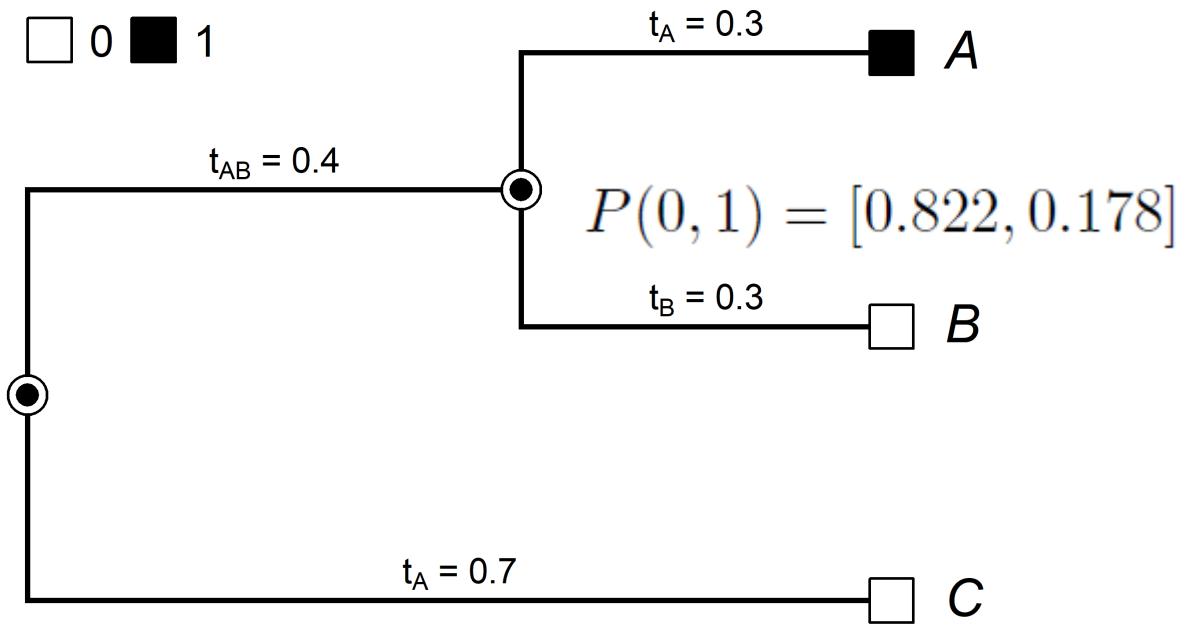
0 1



$$\begin{aligned} P(internal = 0) &= \pi_0 \times P(0|0, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|0, t_C) \\ &\quad + \pi_1 \times P(0|1, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|1, t_C) \\ &= 0.5 \times 0.0434 + 0.5 \times 0.0005 \\ &= 0.0219 \end{aligned}$$

$$\begin{aligned} P(internal = 1) &= \pi_0 \times P(1|0, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|0, t_C) \\ &\quad + \pi_1 \times P(1|1, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|1, t_C) \\ &= 0.5 \times 0.0035 + 0.5 \times 0.0060 \\ &= 0.0047 \end{aligned}$$

0 1

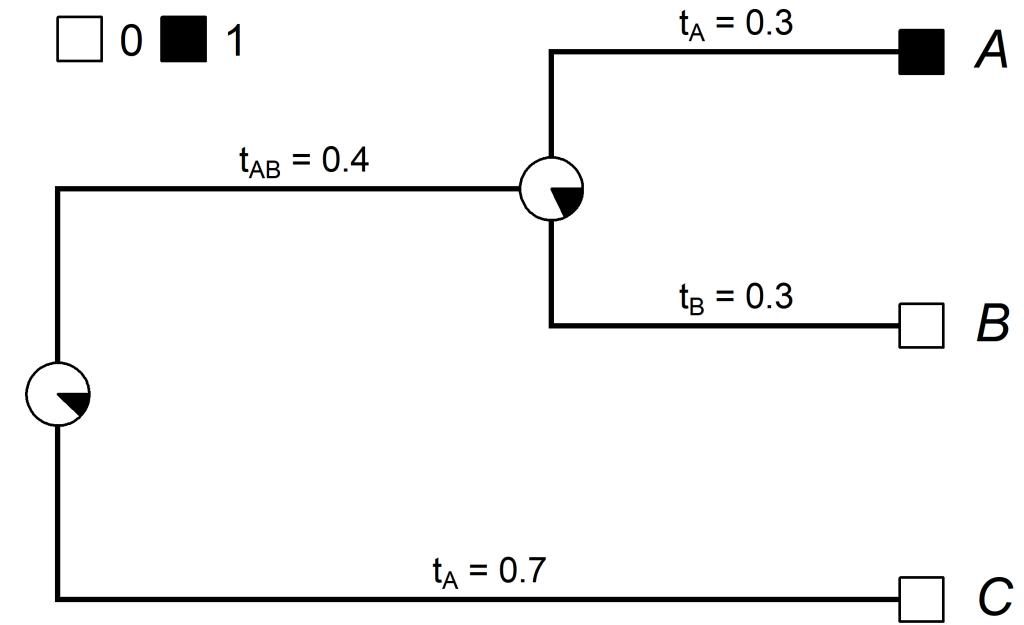


$$P(0, 1) = [0.822, 0.178]$$

$$\begin{aligned} P(internal = 0) &= \pi_0 \times P(0|0, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|0, t_C) \\ &\quad + \pi_1 \times P(0|1, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|1, t_C) \\ &= 0.5 \times 0.0434 + 0.5 \times 0.0005 \\ &= 0.0219 \end{aligned}$$

$$\begin{aligned} P(internal = 1) &= \pi_0 \times P(1|0, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|0, t_C) \\ &\quad + \pi_1 \times P(1|1, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|1, t_C) \\ &= 0.5 \times 0.0035 + 0.5 \times 0.0060 \\ &= 0.0047 \end{aligned}$$

0 1

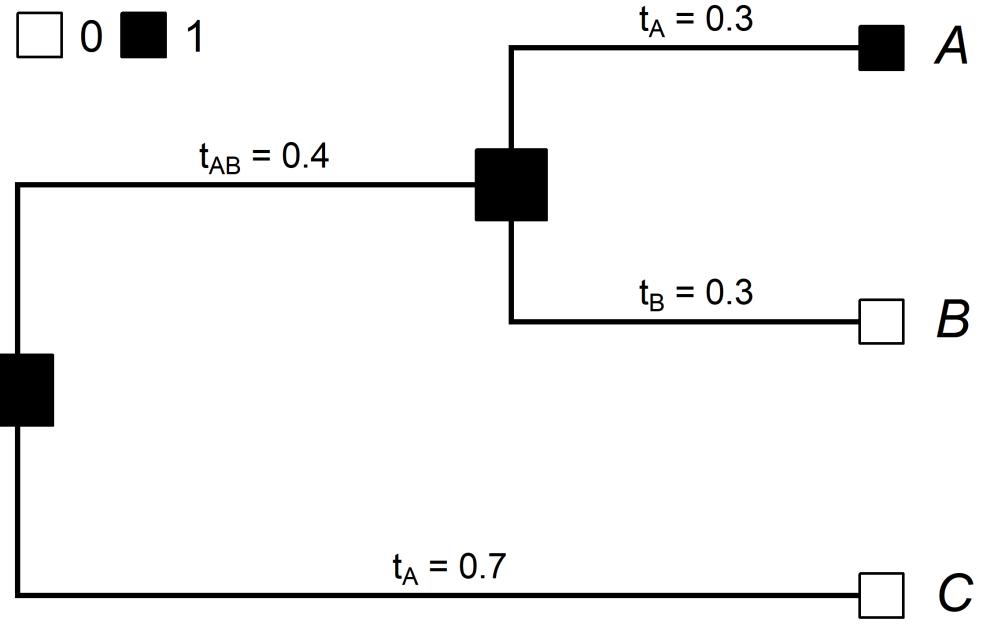


$$\begin{aligned} P(internal = 0) &= \pi_0 \times P(0|0, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|0, t_C) \\ &\quad + \pi_1 \times P(0|1, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|1, t_C) \\ &= 0.5 \times 0.0434 + 0.5 \times 0.0005 \\ &= 0.0219 \end{aligned}$$

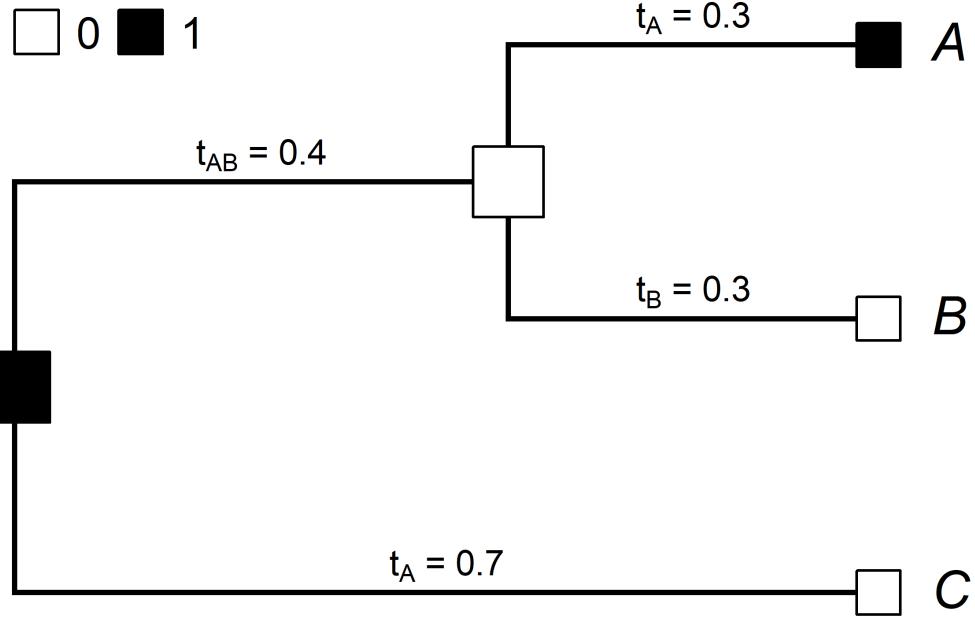
$$\begin{aligned} P(internal = 1) &= \pi_0 \times P(1|0, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|0, t_C) \\ &\quad + \pi_1 \times P(1|1, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|1, t_C) \\ &= 0.5 \times 0.0035 + 0.5 \times 0.0060 \\ &= 0.0047 \end{aligned}$$

Joint Reconstruction

- *Joint reconstruction* has a very different aim.
- It involves finding the set of states across all nodes with the highest probability under our fitted model.

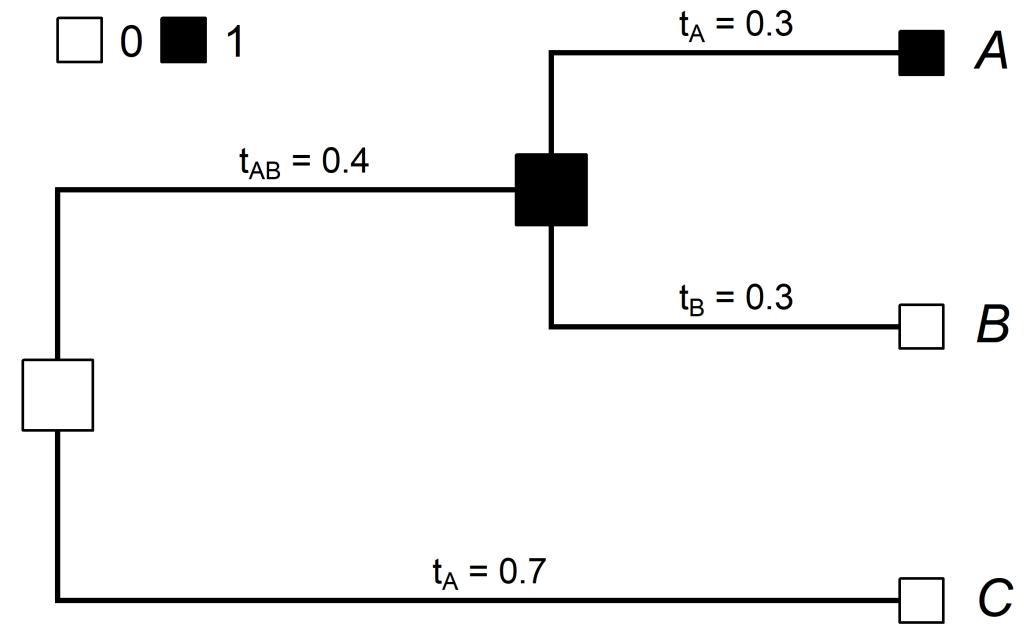


$$P([1, 1]) = \pi_1 \times P(1|1, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|1, t_C) = 0.0030$$



$$P([1, 1]) = \pi_1 \times P(1|1, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|1, t_C) = 0.0030$$

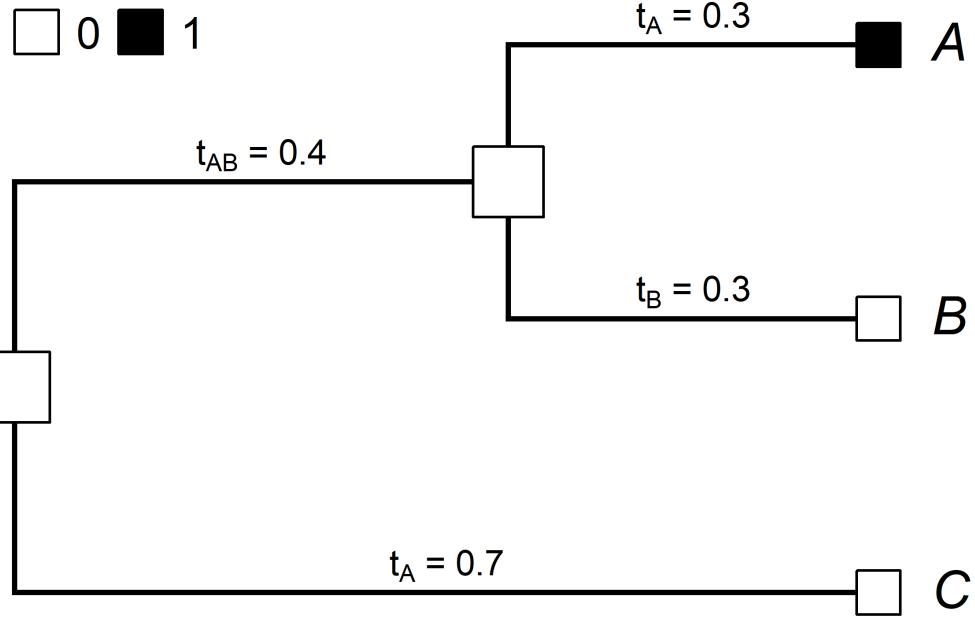
$$P([1, 0]) = \pi_1 \times P(0|1, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|1, t_C) = 0.0002$$



$$P([1, 1]) = \pi_1 \times P(1|1, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|1, t_C) = 0.0030$$

$$P([1, 0]) = \pi_1 \times P(0|1, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|1, t_C) = 0.0002$$

$$P([0, 1]) = \pi_0 \times P(1|0, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|0, t_C) = 0.0017$$



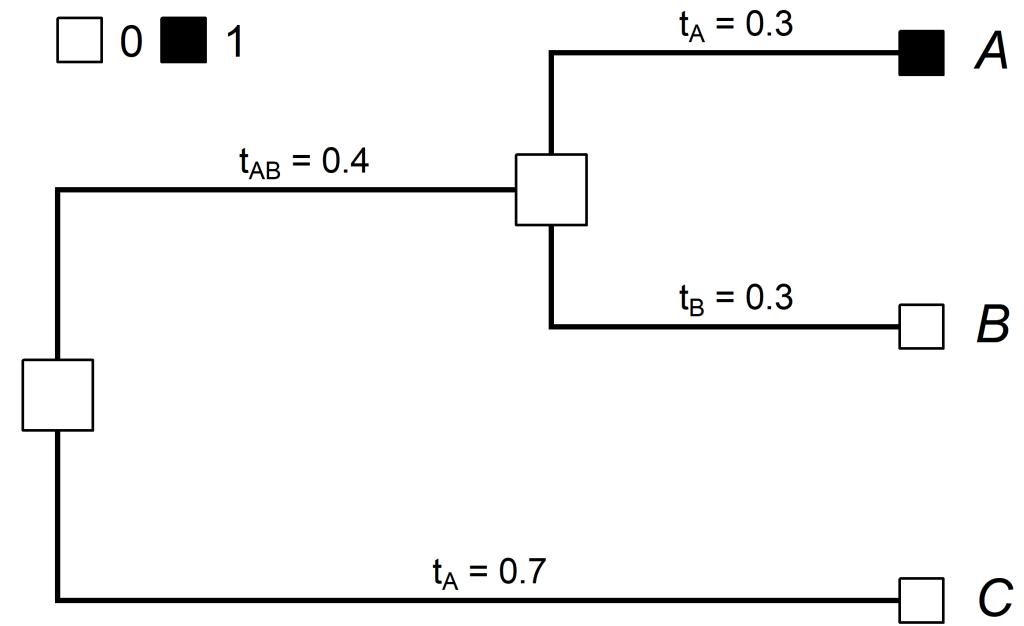
$$P([1, 1]) = \pi_1 \times P(1|1, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|1, t_C) = 0.0030$$

$$P([1, 0]) = \pi_1 \times P(0|1, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|1, t_C) = 0.0002$$

$$P([0, 1]) = \pi_0 \times P(1|0, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|0, t_C) = 0.0017$$

$$P([0, 0]) = \pi_0 \times P(0|0, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|0, t_C) = 0.0217$$

0 1



$$P([1, 1]) = \pi_1 \times P(1|1, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|1, t_C) = 0.0030$$

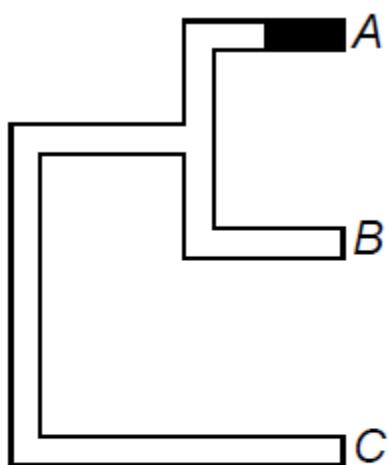
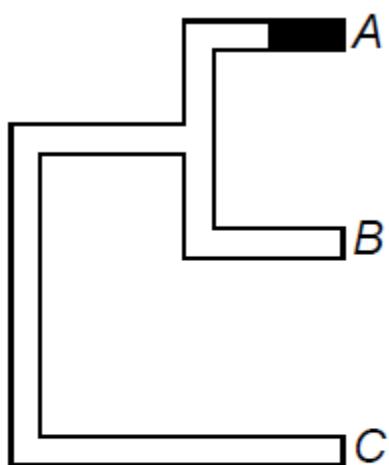
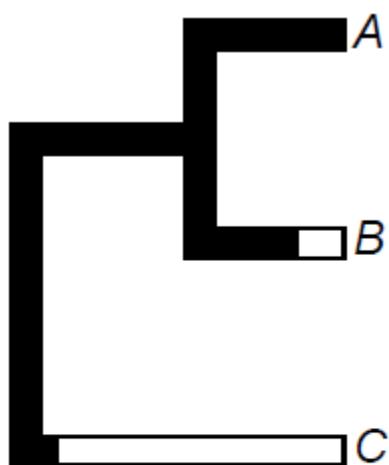
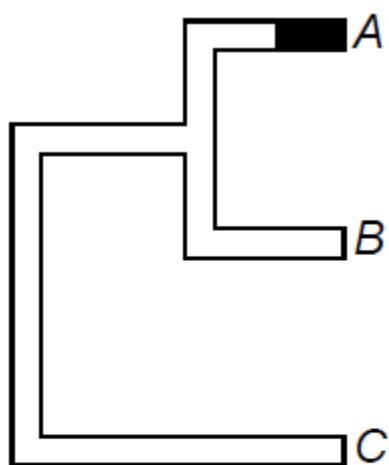
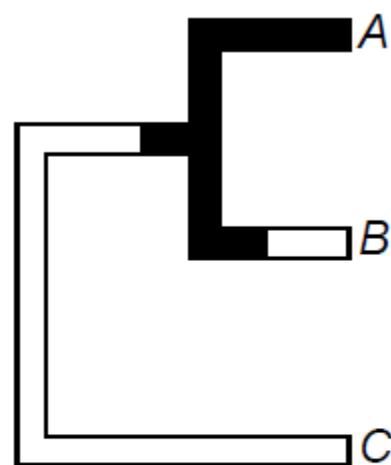
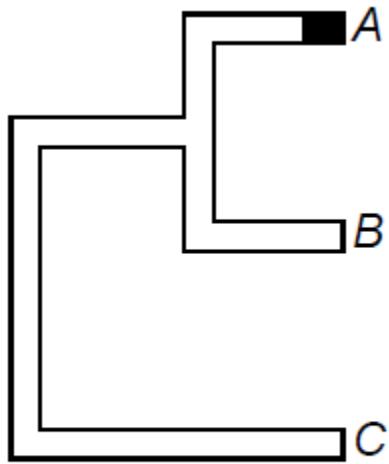
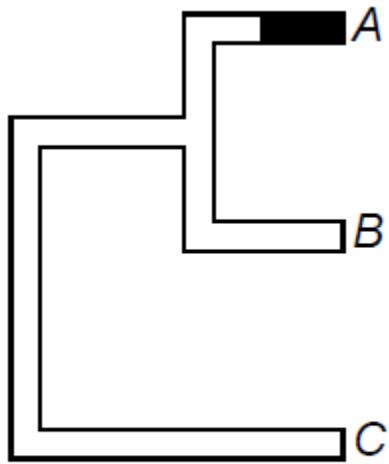
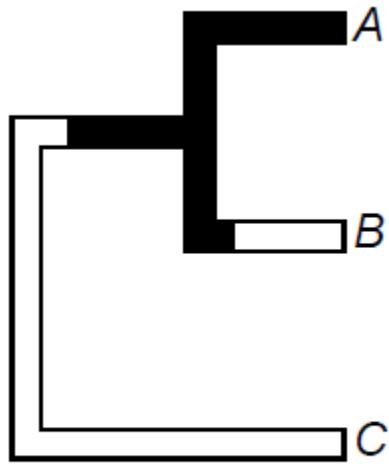
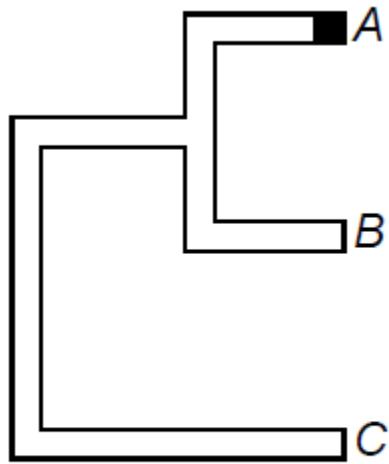
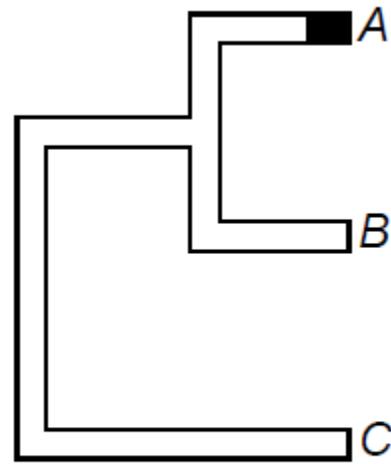
$$P([1, 0]) = \pi_1 \times P(0|1, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|1, t_C) = 0.0002$$

$$P([0, 1]) = \pi_0 \times P(1|0, t_{AB}) \times P(1|1, t_A) \times P(0|1, t_B) \times P(0|0, t_C) = 0.0017$$

$$P([0, 0]) = \pi_0 \times P(0|0, t_{AB}) \times P(1|0, t_A) \times P(0|0, t_B) \times P(0|0, t_C) = 0.0217$$

Stochastic Character Mapping

- Another closely related approach for ancestral state reconstruction is a procedure called stochastic character mapping.
- In stochastic character mapping we sample complete character histories of our discrete trait.
- Each stochastic history is consistent with our tip data and sampled in proportion to its probability under our model.

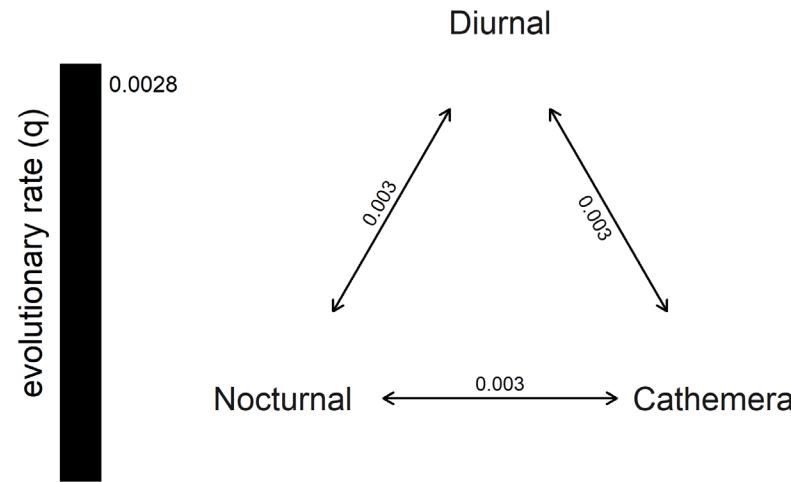


□ 0 ■ 1

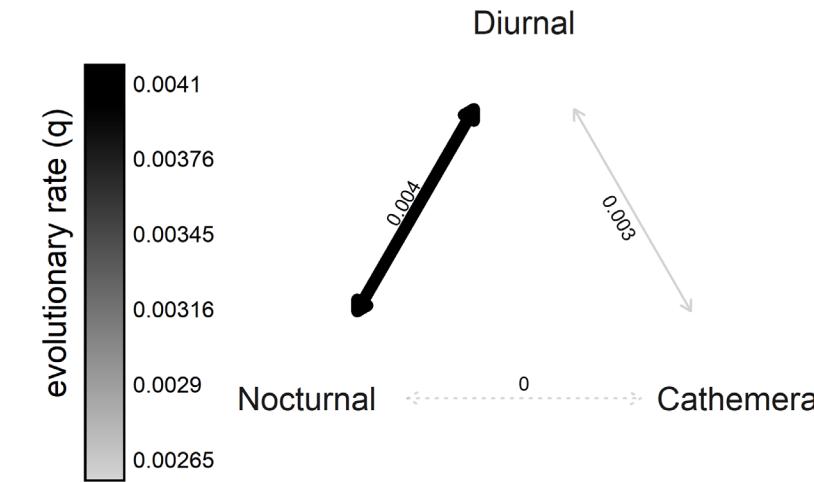
Empirical example I: Marginal reconstruction of diel activity pattern in primates

- To illustrate marginal reconstruction, let's take a dataset of diel activity pattern, coded as ‘nocturnal,’ ‘diurnal,’ and ‘cathemeral’ (active randomly during the day & night), for 90 species of living primates.
- Phylogeny & trait data from Kirk & Kay (2004).

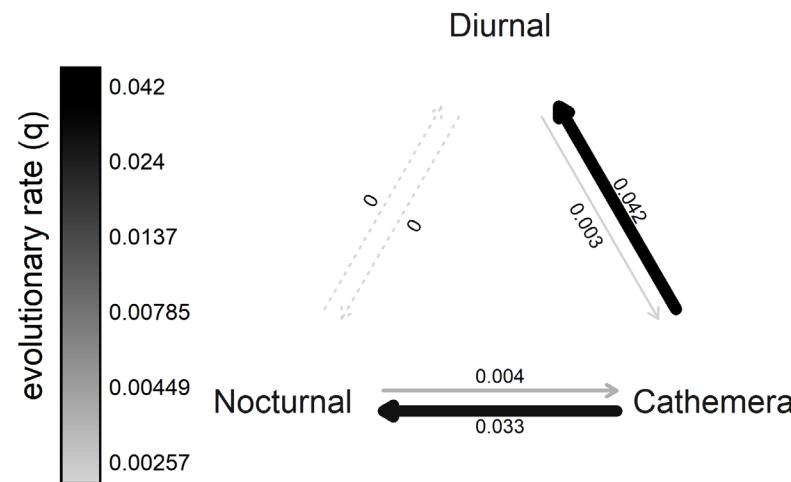
a) AIC = 62.7; model weight = 0.28



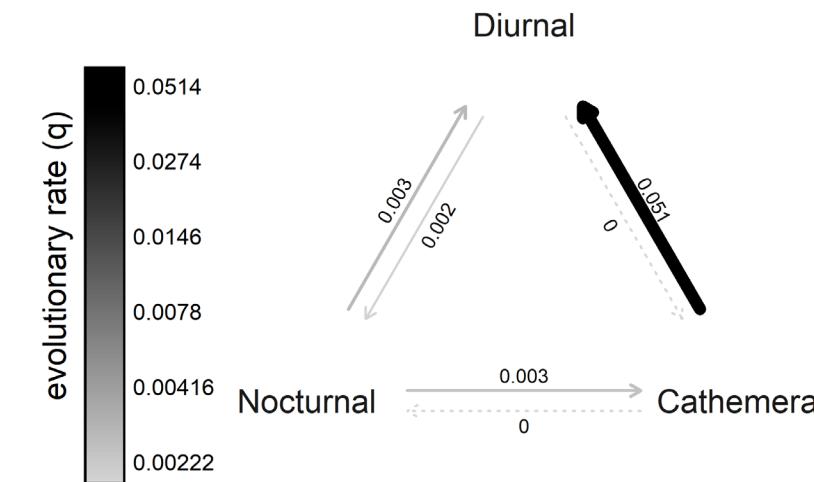
b) AIC = 63.08; model weight = 0.23

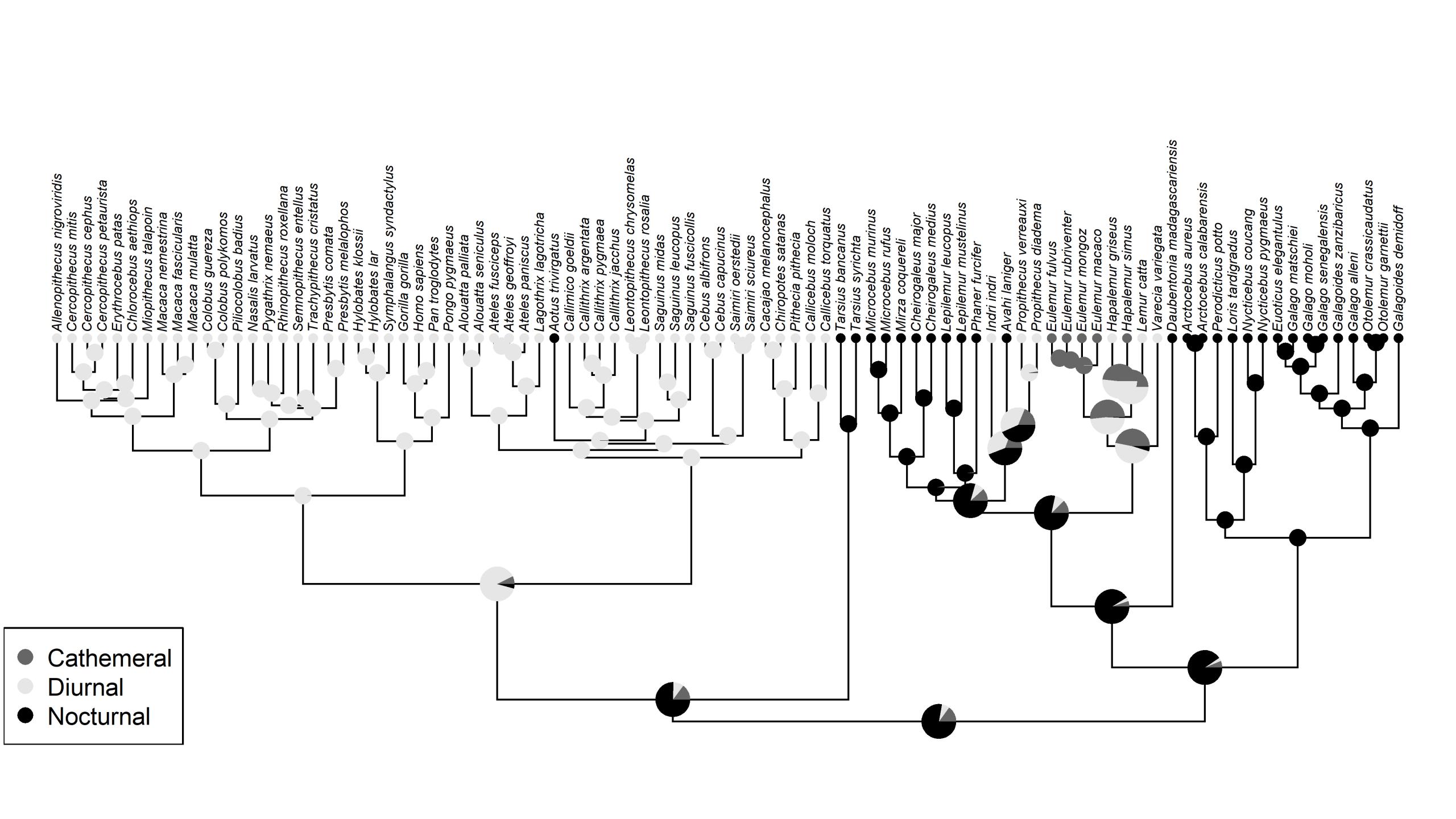


c) AIC = 63.15; model weight = 0.22



d) AIC = 62.75; model weight = 0.27





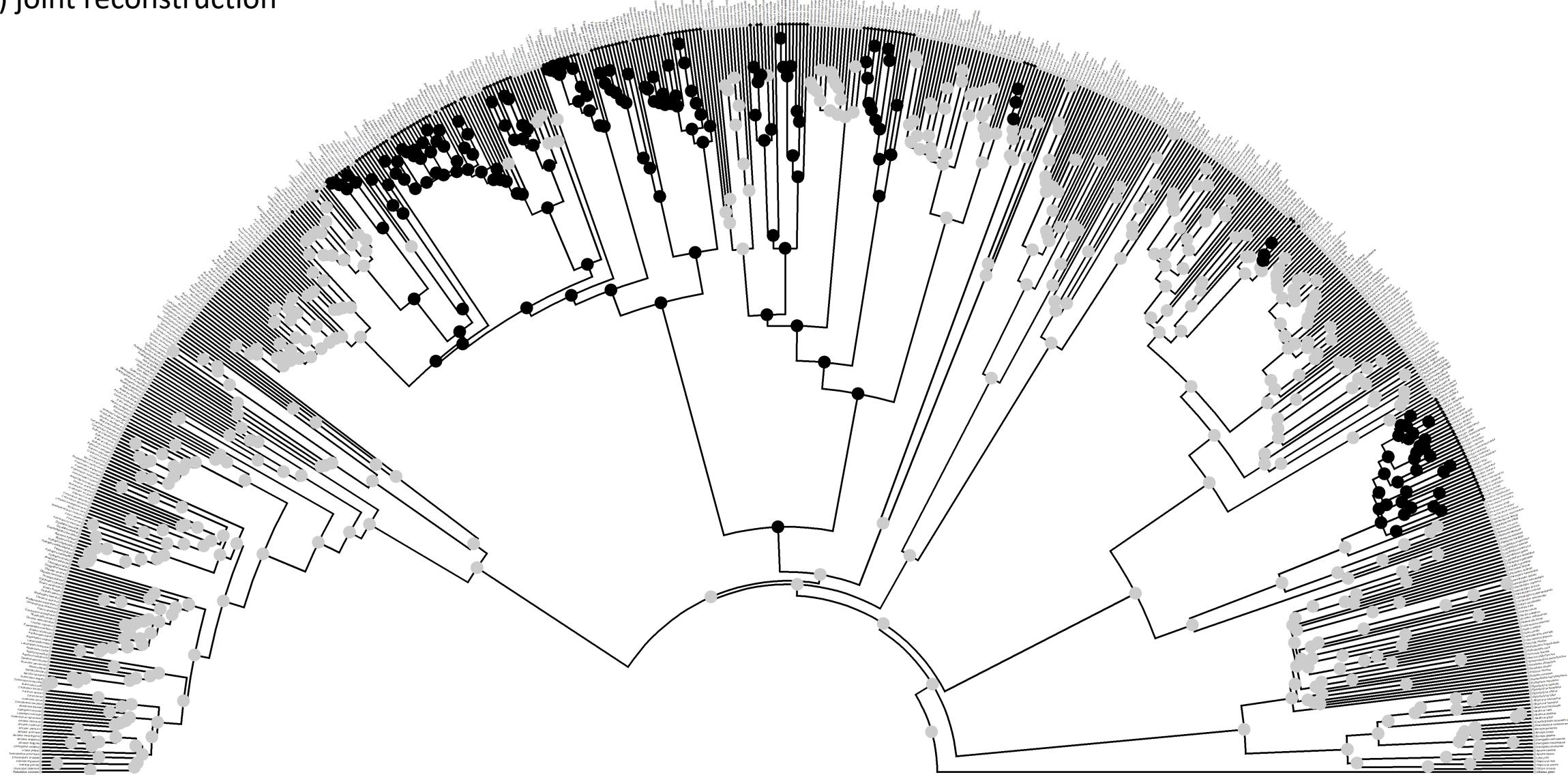
Empirical example II: Joint reconstruction of tail spine presence/absence in lizards

- To illustrate joint reconstruction, I used a dataset & phylogeny of the presence or absence of ‘tail spines’ in 658 species of lizards.
- The phylogeny & data for this example come from Ramm et al. (2019).

Table 1: Estimated transition rates, log-likelihood, number of parameters, AIC, and model weights for two different discrete character evolution models for the evolution of the presence or absence of tail spines in lizards. See main text for more details.

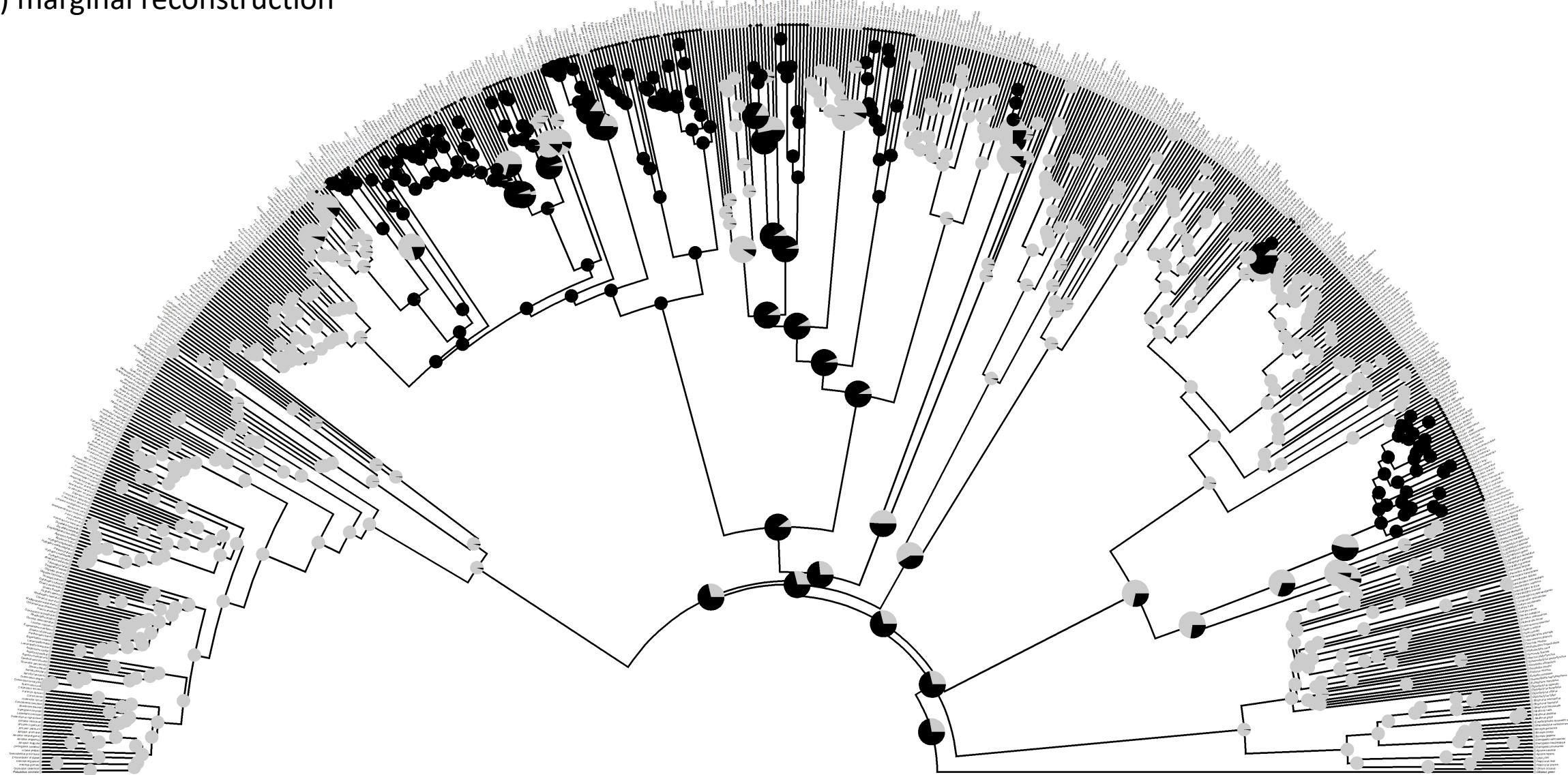
	$q_{0,1}$	$q_{1,0}$	log(L)	d.f.	AIC	weight
ER model	0.00152	0.00152	-123.5618	1	249.1235	0.0688
ARD model	0.00059	0.01112	-119.9564	2	243.9129	0.9312

a) joint reconstruction



● non-spiny
● spiny

b) marginal reconstruction



non-spiny
spiny

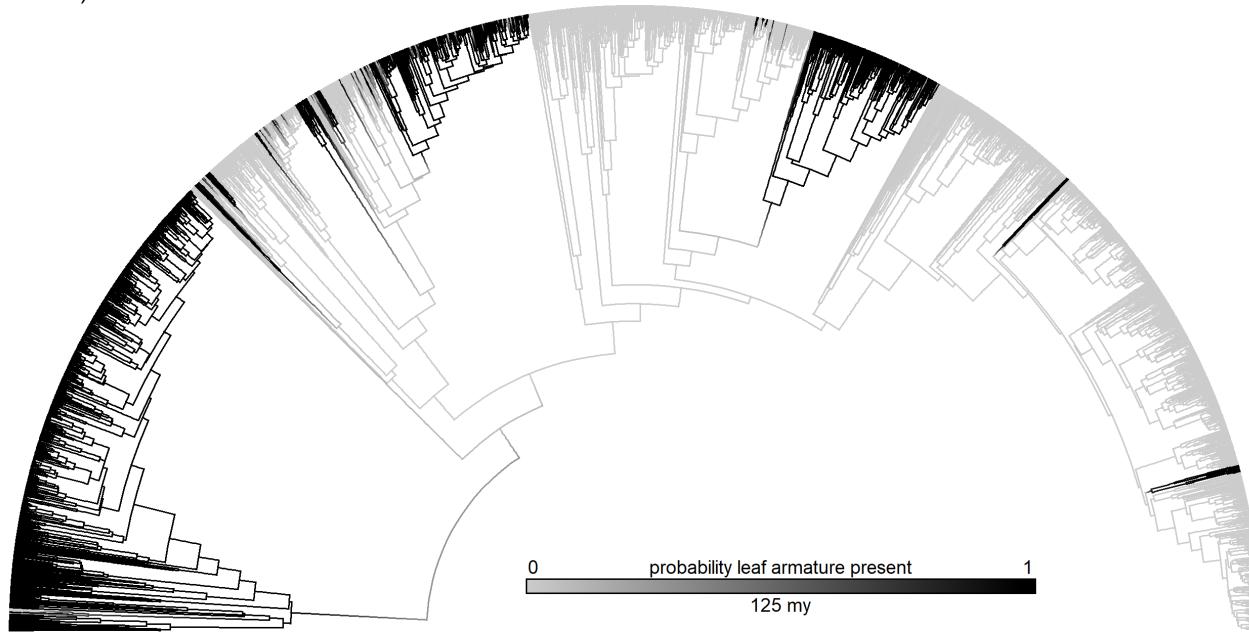
Empirical example III: Stochastic character mapping palm leaf armature

- To demonstrate stochastic character mapping I used a dataset for presence/absence of leaf armature in a phylogeny of 2,539 species palms.
- The phylogeny & dataset for this example come from Onstein et al. (2022).

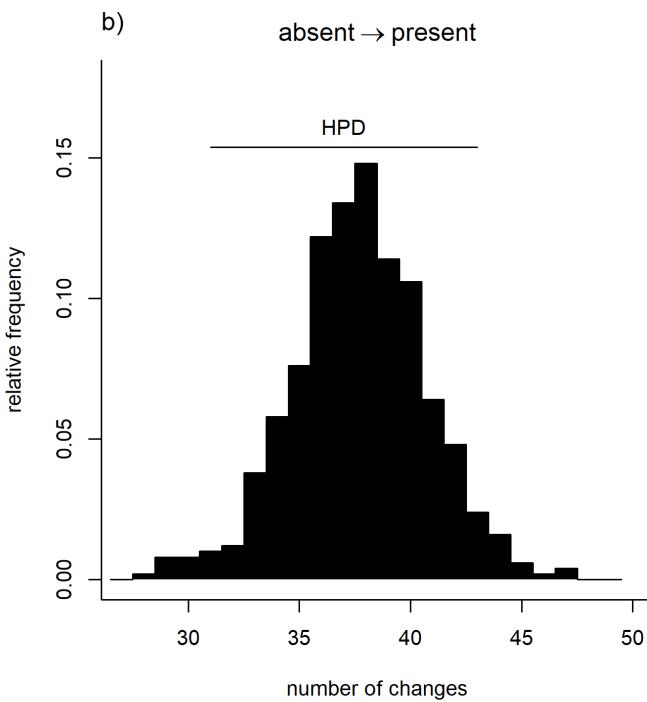
Table 2: Estimated transition rates, log-likelihood, number of parameters, AIC, and model weights for four different discrete character evolution models for the evolution of the presence or absence of leaf armature in palms. See main text for more details.

	$q_{0,1}$	$q_{1,0}$	log(L)	d.f.	AIC	weight
ER model	0.00392	0.00392	-431.5580	1	865.1161	0.57424
absent → present	0.01162	0.00000	-617.2943	1	1236.5886	0.00000
present → absent	0.00000	0.01279	-485.6286	1	973.2573	0.00000
ARD model	0.00338	0.00471	-430.8572	2	865.7145	0.42576

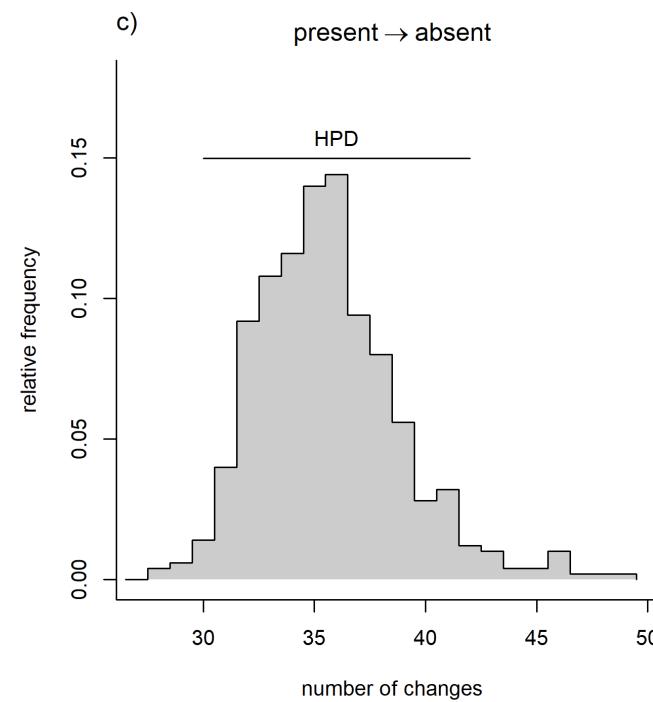
a)



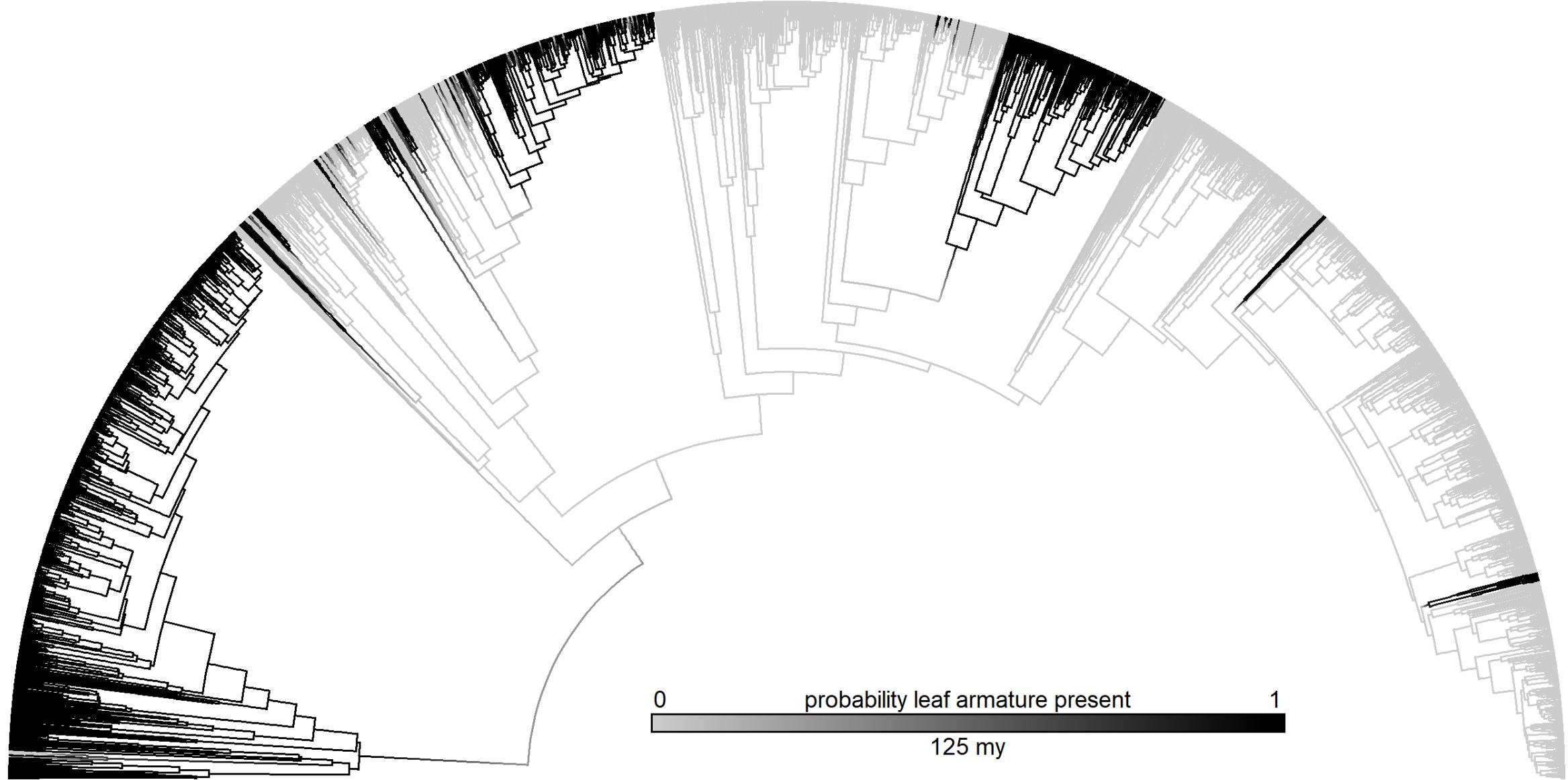
b)



c)



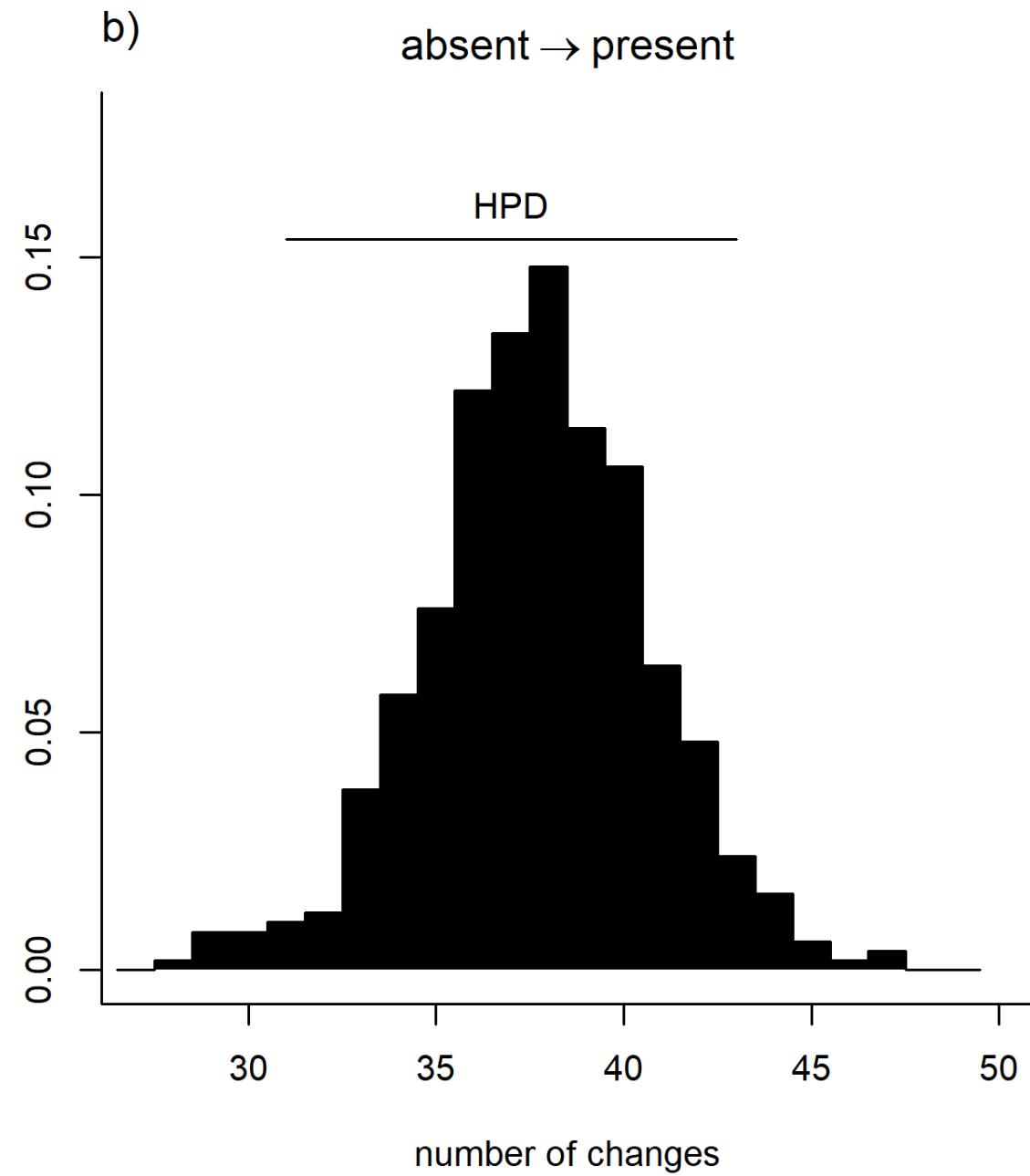
a)



b)

absent → present

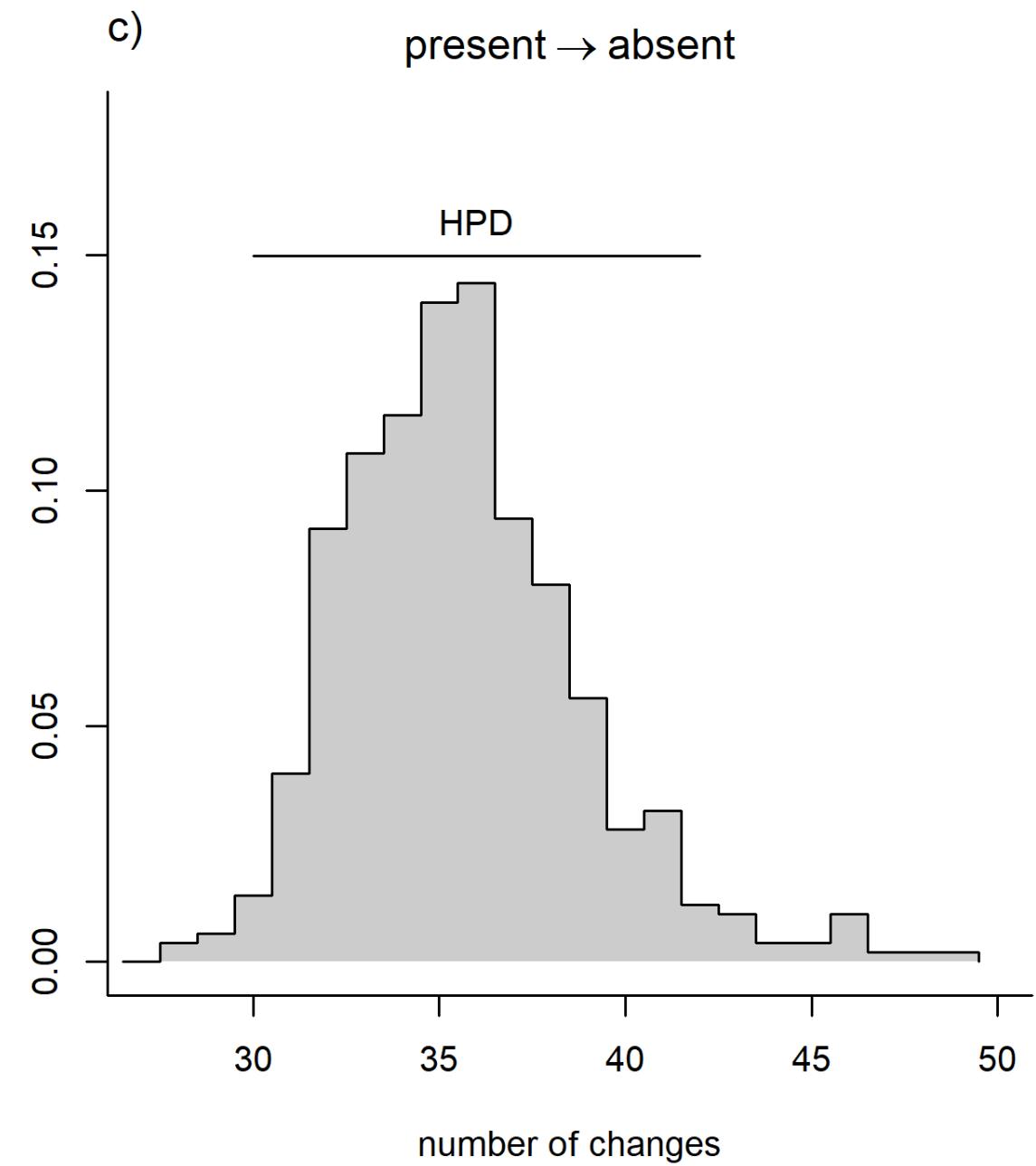
relative frequency



c)

present → absent

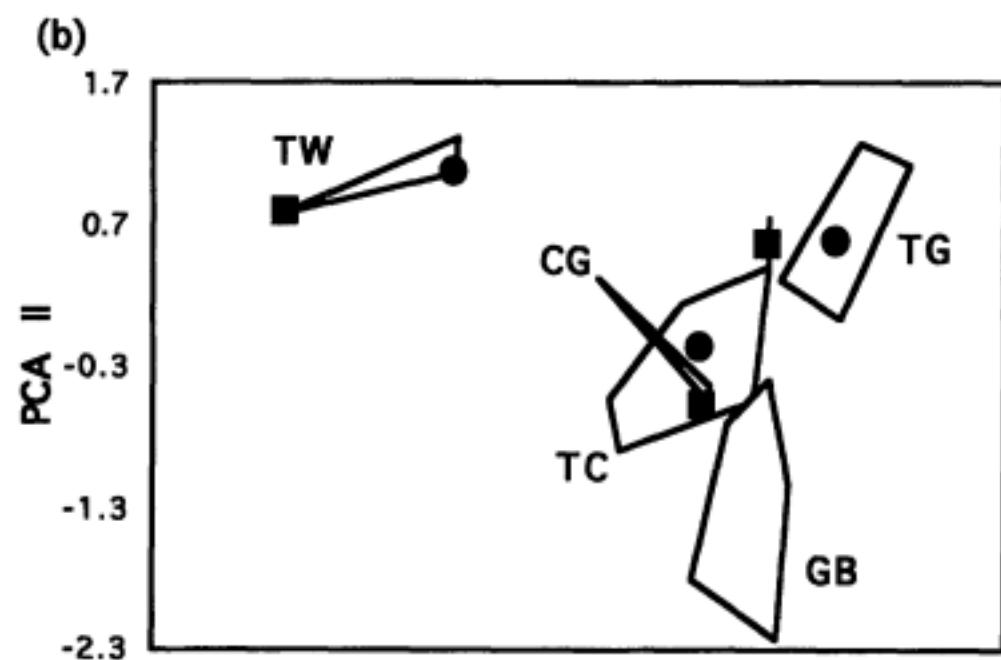
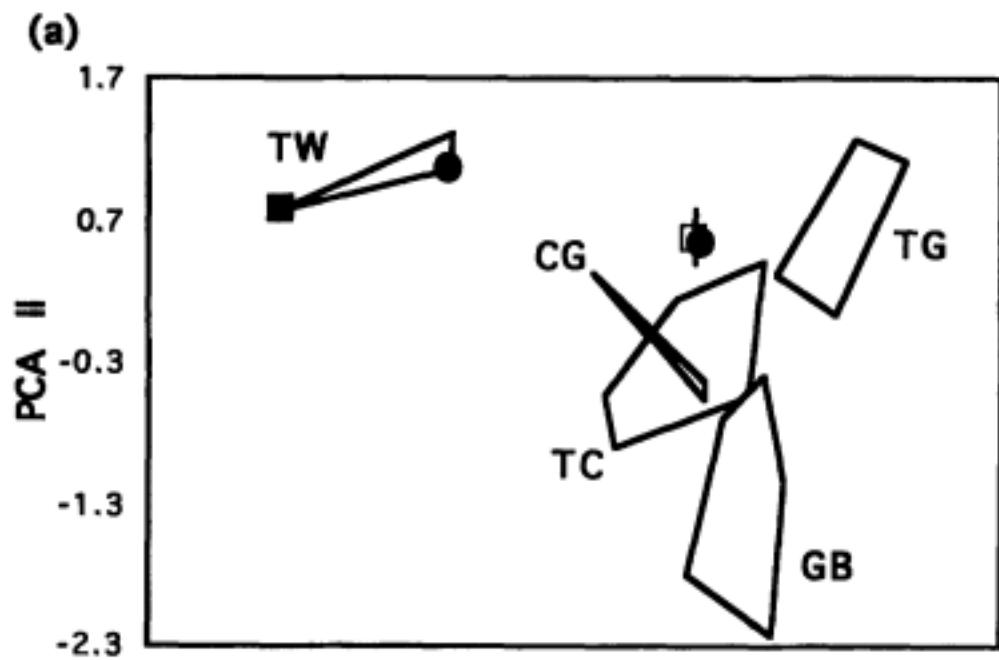
relative frequency

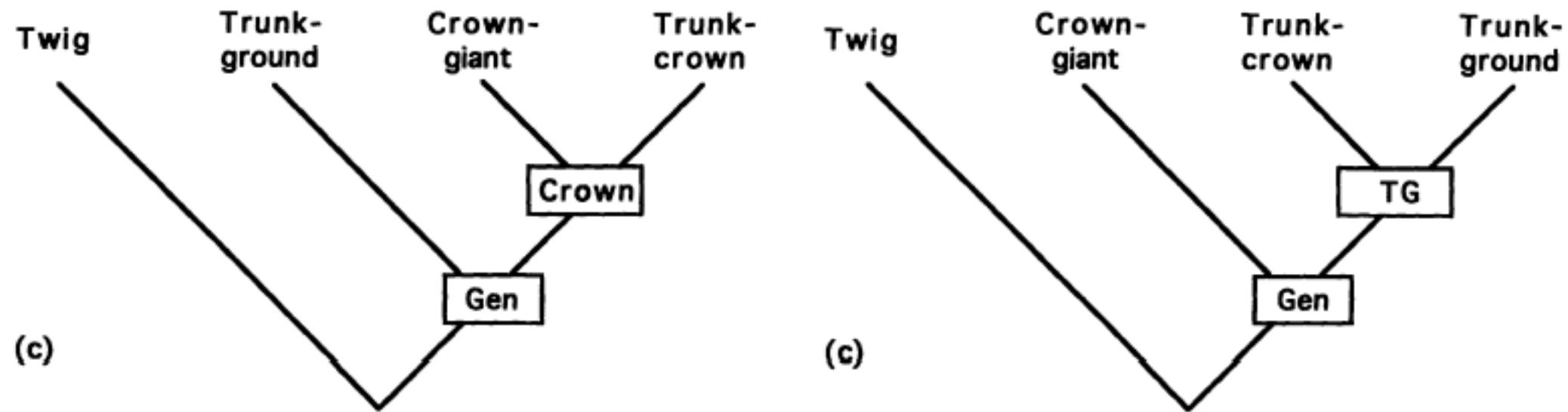
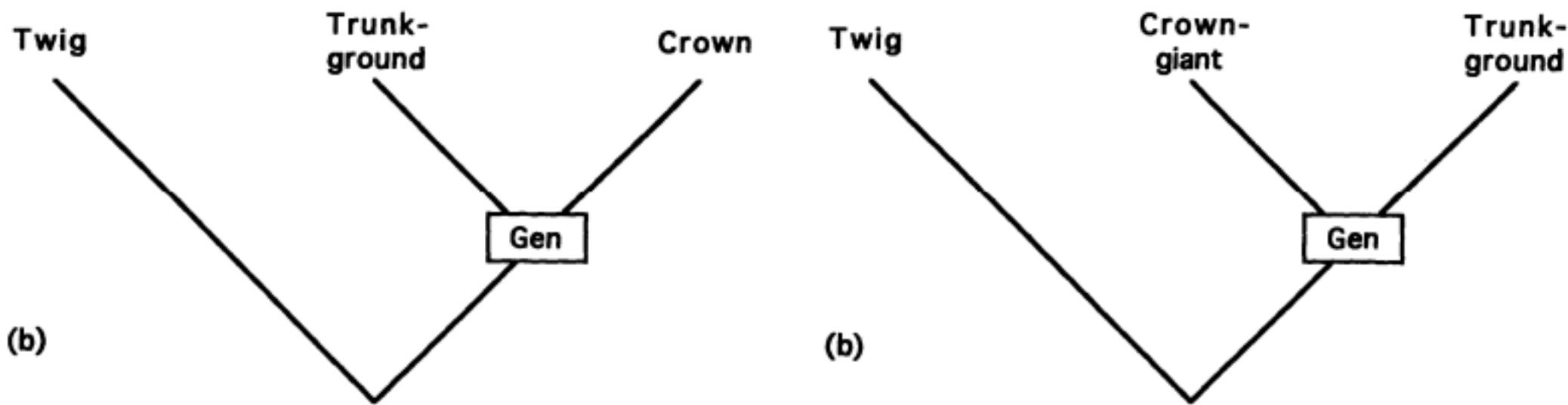
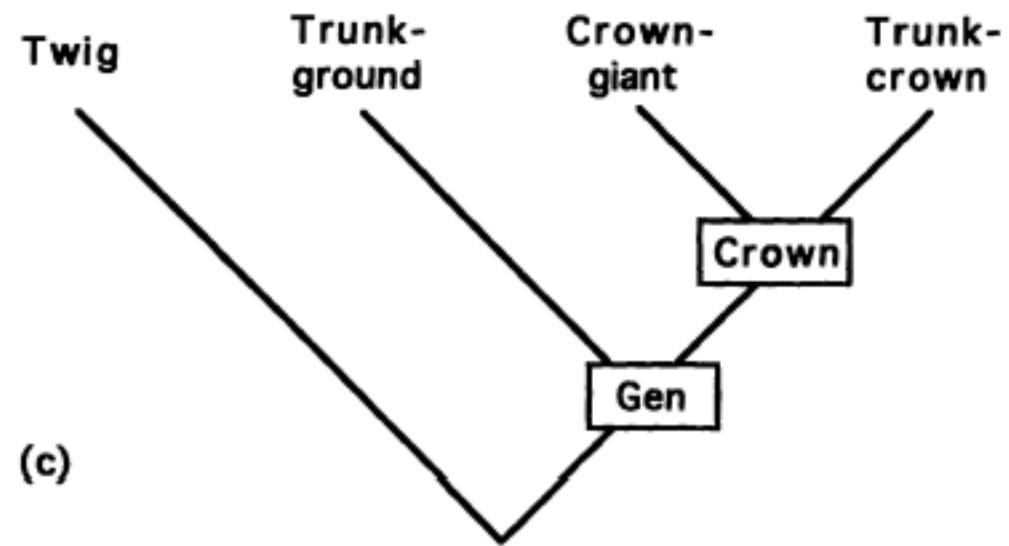
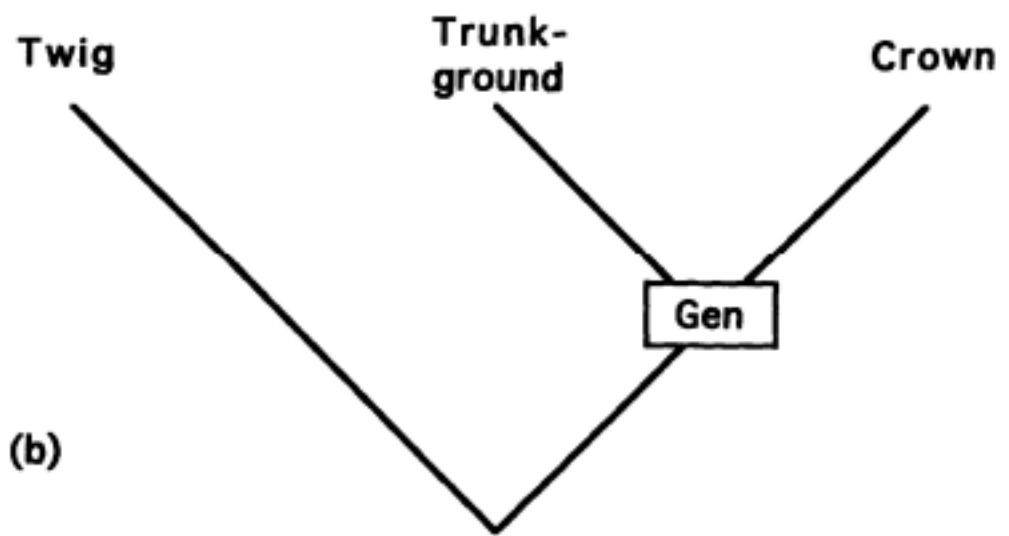


Ancestral character estimation for
continuous characters

First a story...







EVOLUTION

INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

PUBLISHED BY

THE SOCIETY FOR THE STUDY OF EVOLUTION

Vol. 51

December 1997

No. 6

Evolution, 51(6), 1997, pp. 1699–1711

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*

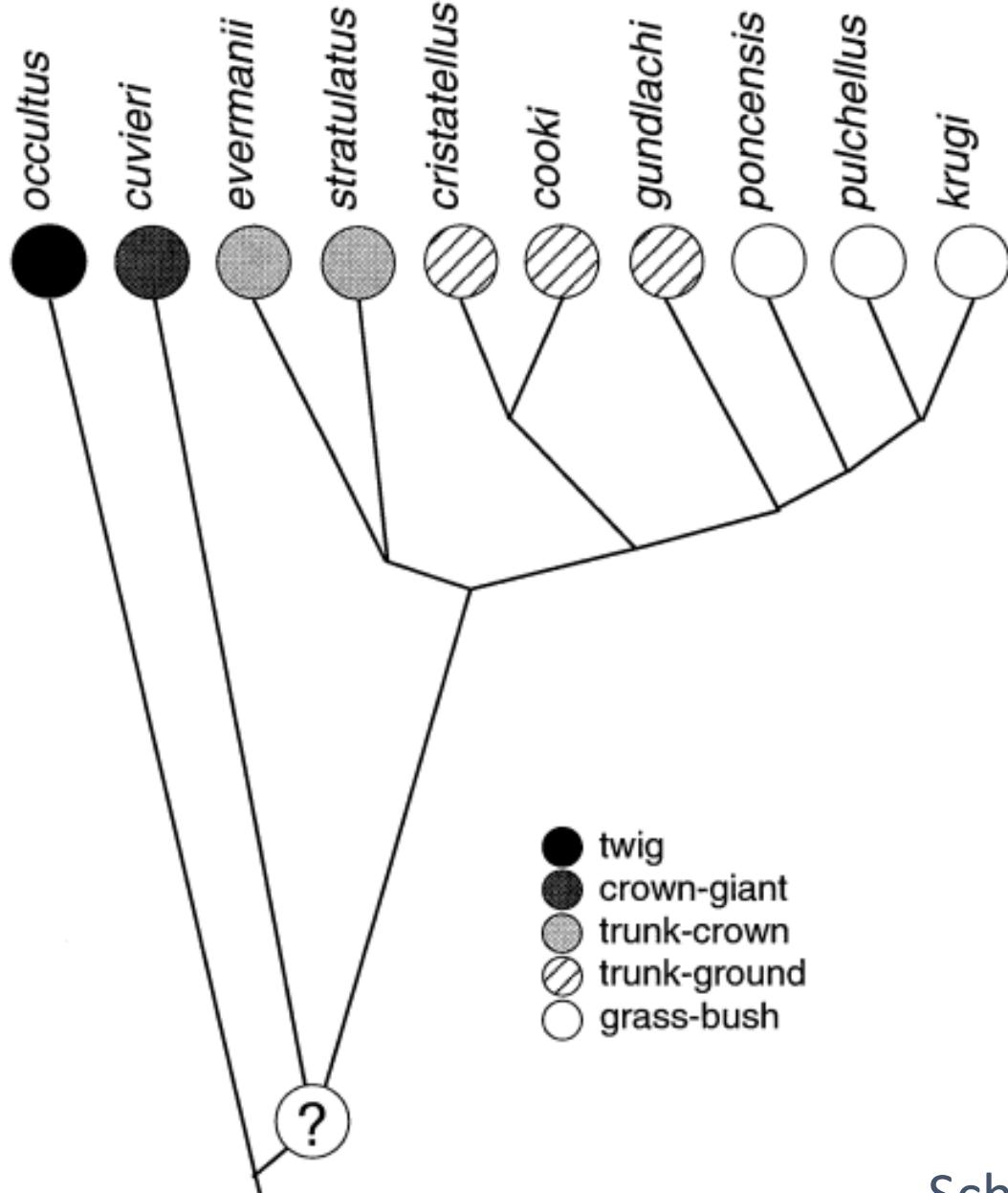
²*E-mail:* schluter@zoology.ubc.ca

³*Biology Department O-116, University of California—San Diego, La Jolla, California 92093*

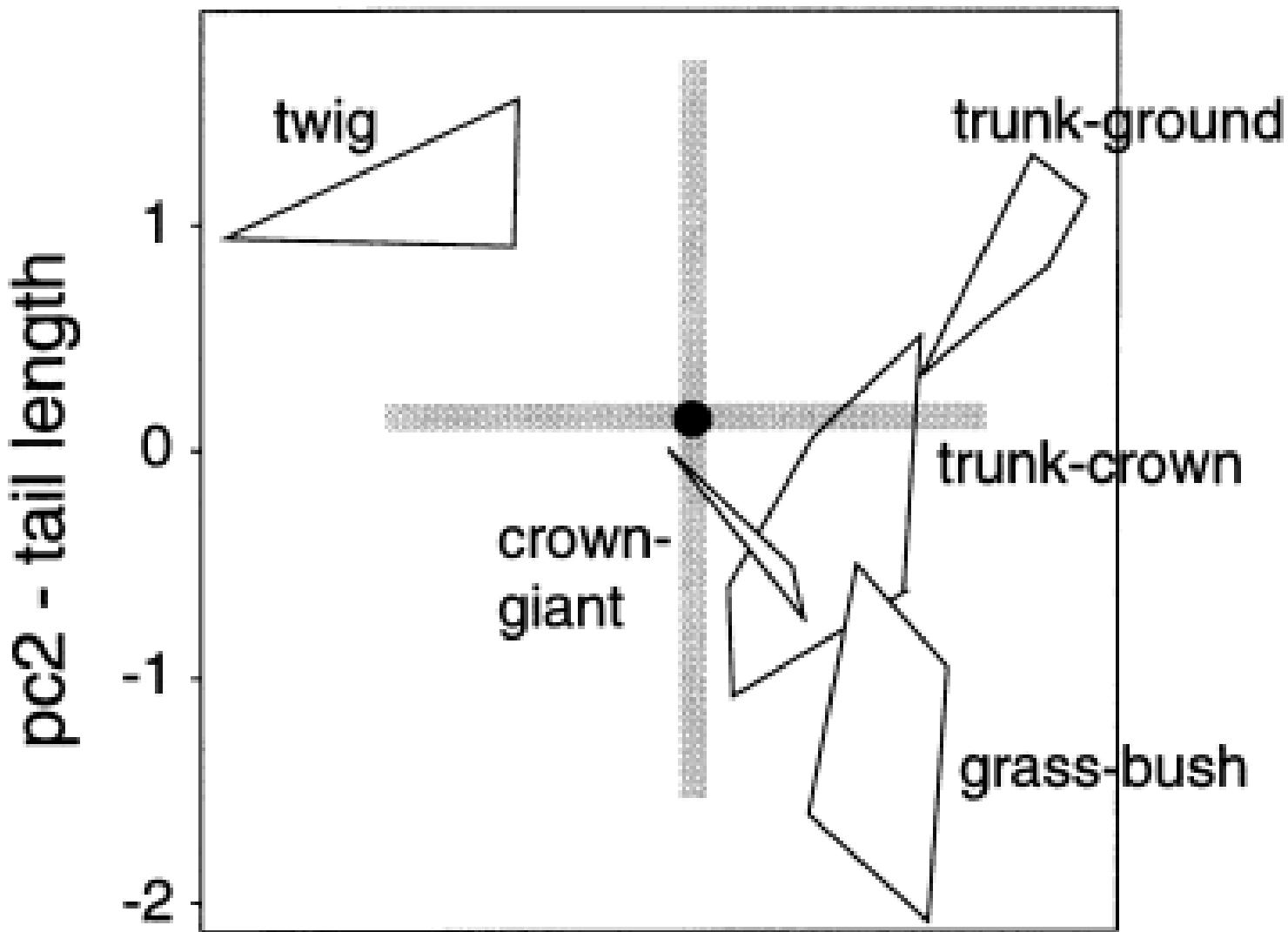
E-mail: tprice@ucsd.edu

⁴*E-mail:* mooers@bio.uva.nl

⁵*E-mail:* ludwig@zoology.ubc.ca



Schluter et al. 1997



Schluter et al. 1997

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.

How do we reconstruct
ancestral states for
continuous characters?

Ancestral state estimation requires that we specify a particular **model for trait evolution**

**Most common model for continuous
character ancestral states is
Brownian motion**

When characters evolve under Brownian motion
on a tree, then the species' ending values
will follow a **multivariate normal distribution**

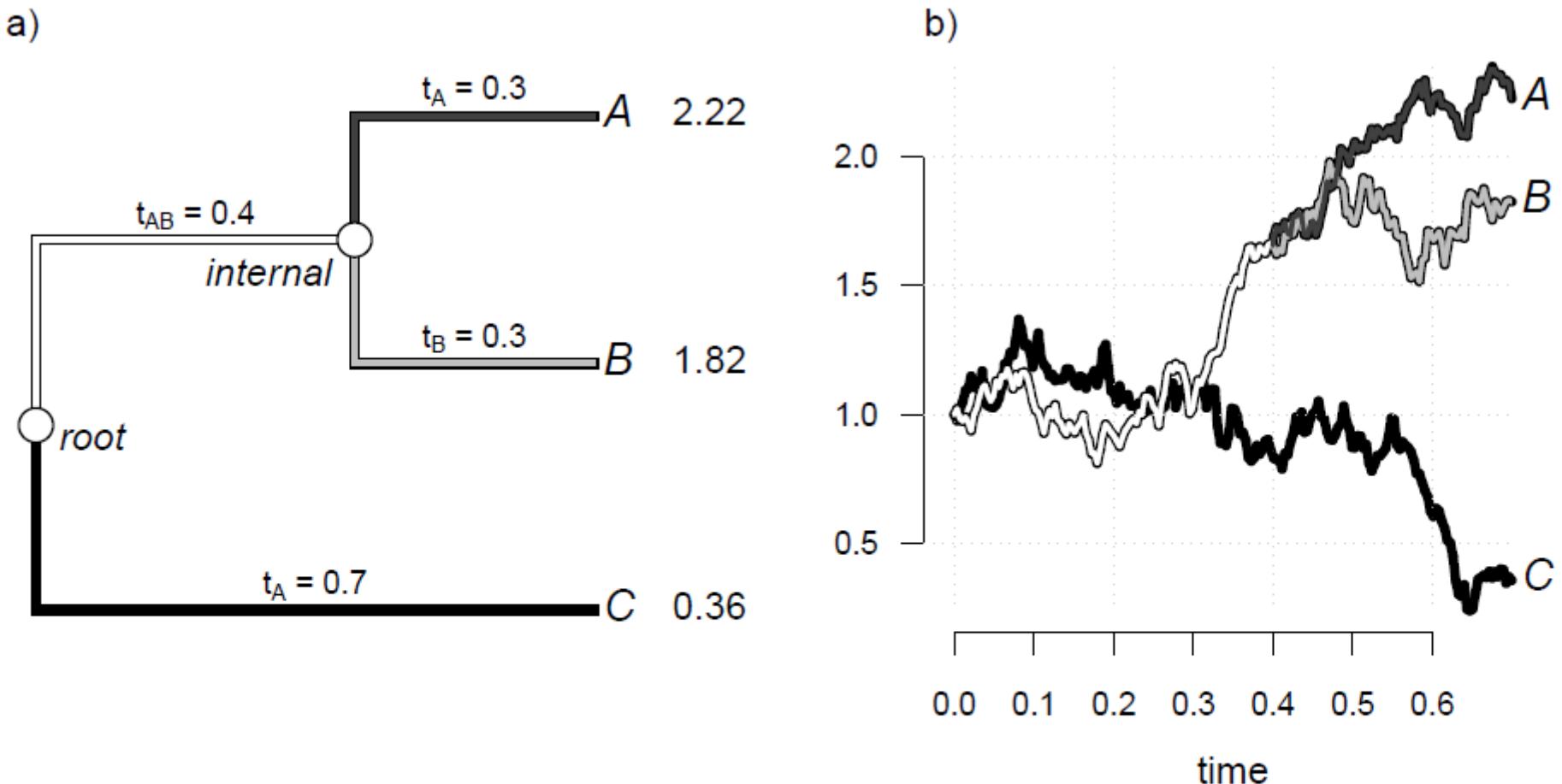


Figure 9: (a) Three-taxon phylogenetic tree of Figure 1, but in which each edge of the tree has been plotted with a different color. (b) Single realization of Brownian motion evolution on the tree of figure panel (a). See main text for more details.

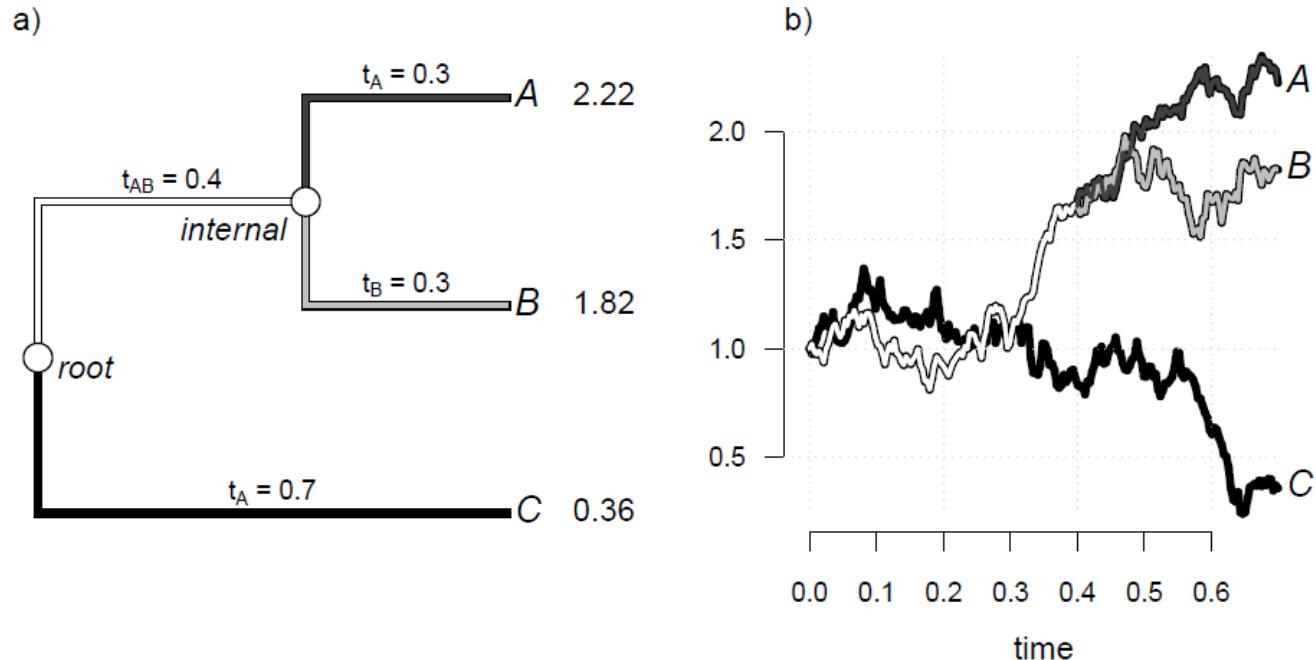


Figure 9: (a) Three-taxon phylogenetic tree of Figure 1, but in which each edge of the tree has been plotted with a different color. (b) Single realization of Brownian motion evolution on the tree of figure panel (a). See main text for more details.

$$\mathbf{C} = \mathbf{B} \begin{bmatrix} A & B & C \\ A & t_A + t_{AB} & t_{AB} & 0.0 \\ B & t_{AB} & t_B + t_{AB} & 0.0 \\ C & 0.0 & 0.0 & t_C \end{bmatrix} = \mathbf{B} \begin{bmatrix} A & B & C \\ 0.7 & 0.4 & 0.0 \\ 0.4 & 0.7 & 0.0 \\ 0.0 & 0.0 & 0.7 \end{bmatrix}$$

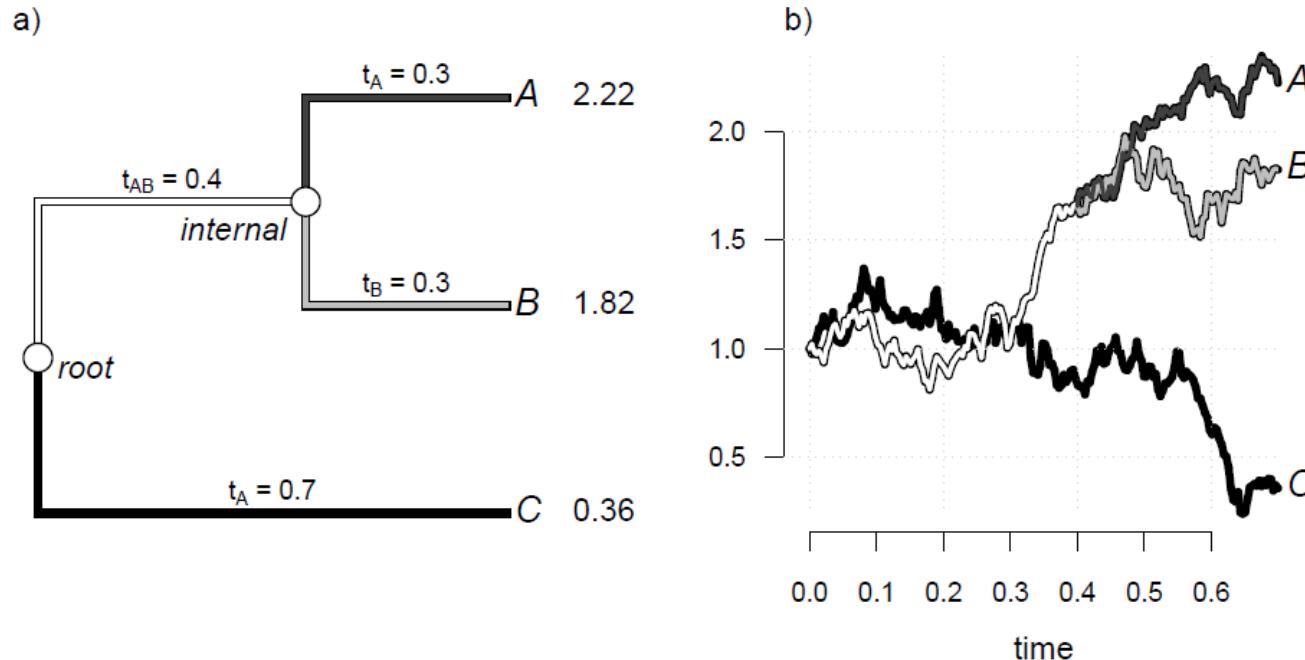


Figure 9: (a) Three-taxon phylogenetic tree of Figure 1, but in which each edge of the tree has been plotted with a different color. (b) Single realization of Brownian motion evolution on the tree of figure panel (a). See main text for more details.

$$C = \begin{matrix} & \begin{matrix} A & B & C & internal \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ internal \end{matrix} & \left[\begin{matrix} t_A + t_{AB} & t_{AB} & 0.0 & t_{AB} \\ t_{AB} & t_B + t_{AB} & 0.0 & t_{AB} \\ 0.0 & 0.0 & t_C & 0.0 \\ t_{AB} & t_{AB} & 0.0 & t_{AB} \end{matrix} \right] \end{matrix} = \begin{matrix} & \begin{matrix} A & B & C & internal \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ internal \end{matrix} & \left[\begin{matrix} 0.7 & 0.4 & 0.0 & 0.4 \\ 0.4 & 0.7 & 0.0 & 0.4 \\ 0.0 & 0.0 & 0.7 & 0.0 \\ 0.4 & 0.4 & 0.0 & 0.4 \end{matrix} \right] \end{matrix}$$

$$P(\mathbf{x}) = \frac{\exp(-\frac{1}{2}[\mathbf{x} - \mathbf{x}_0]'(\sigma^2 \mathbf{C})^{-1}[\mathbf{x} - \mathbf{x}_0])}{\sqrt{(2\pi)^N \times \det(\sigma^2 \mathbf{C})}}$$

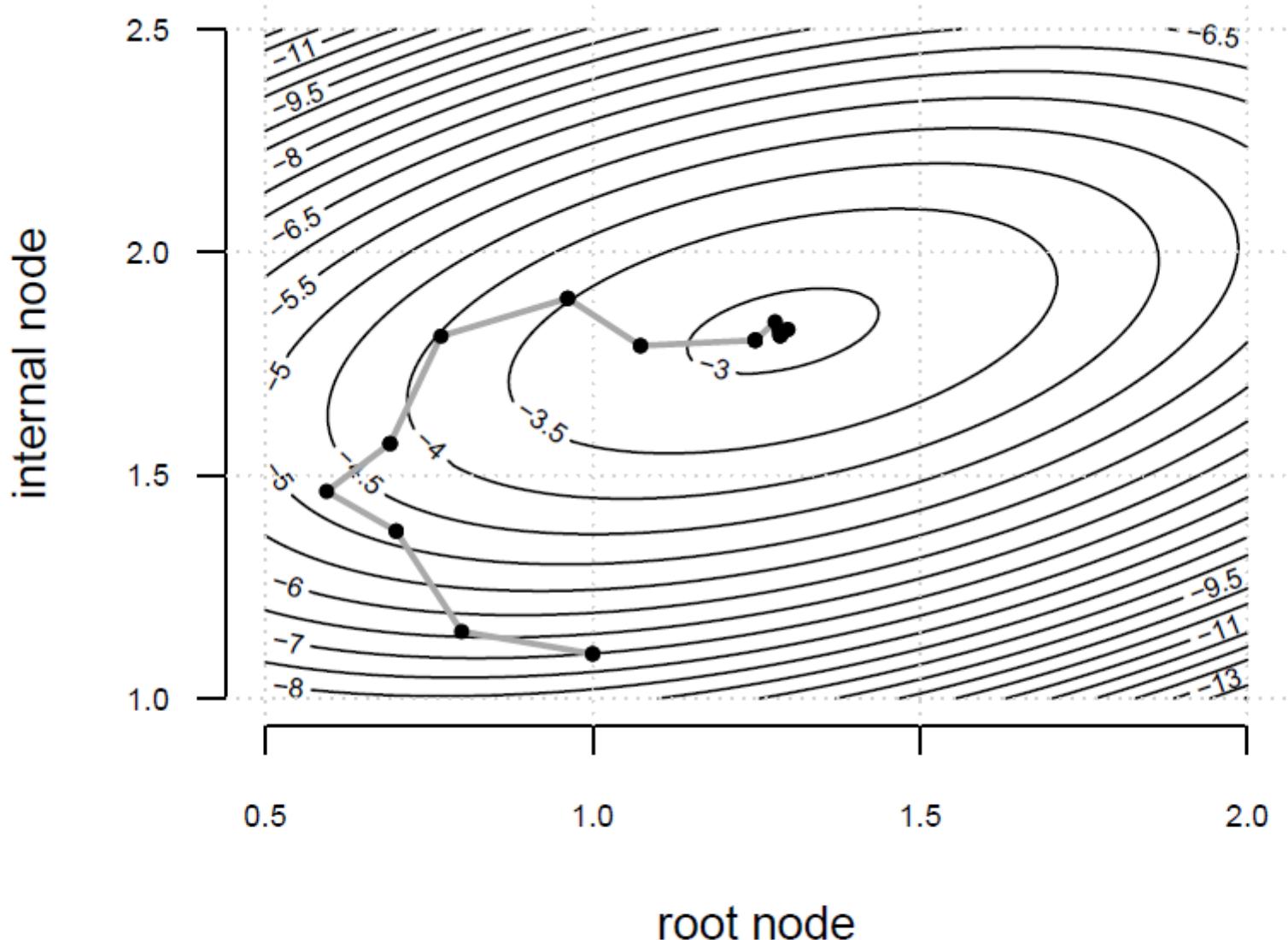


Figure 10: Log-likelihood surface for the numerical values of the root and internal nodes of the tree and ata of Figure 9. See main text for more details.

**There are a number of fast pruning algorithms,
implemented in R, that give the MLEs for ancestral
character states under Brownian motion**

These are both joint and
marginal ancestral state
estimates*

*which are equal

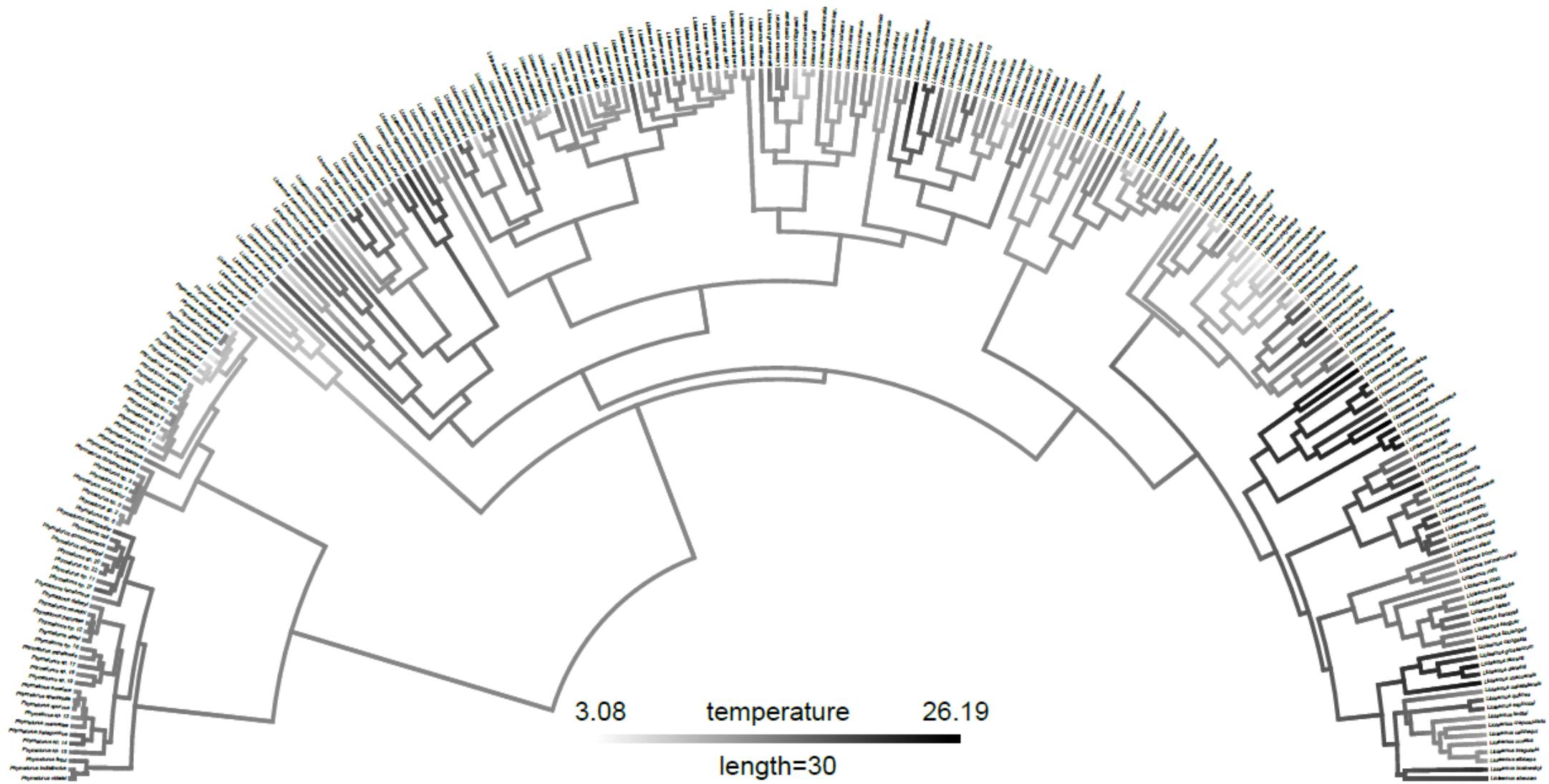
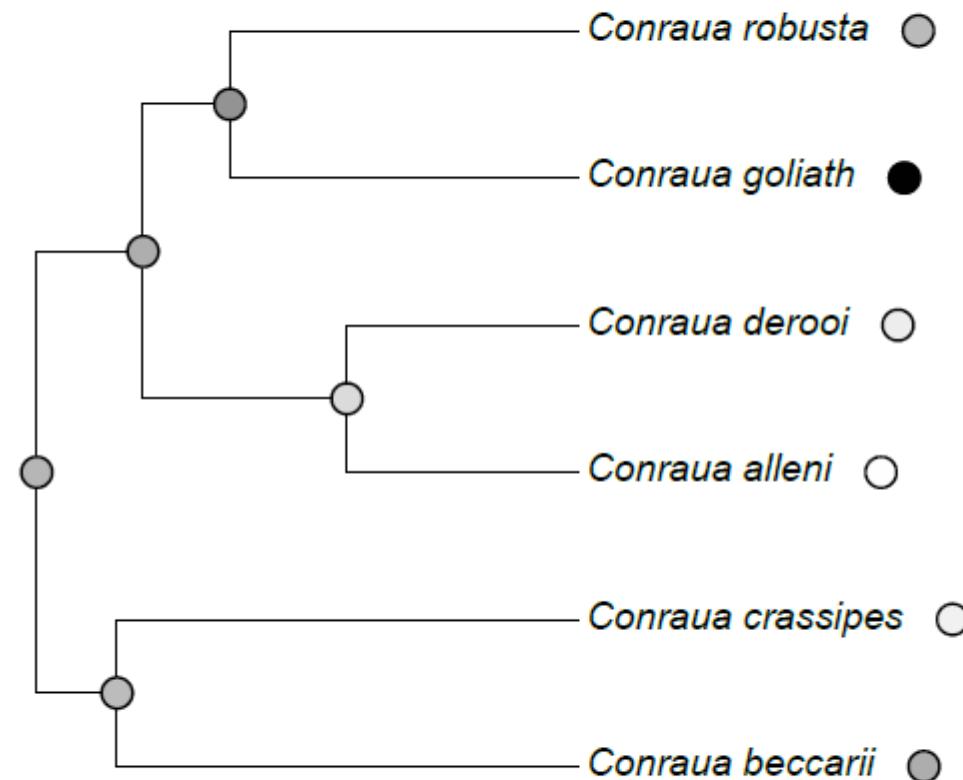


Figure 11: A phylogenetic tree of the Maximum Likelihood ancestral states (along edges) and observed values (at the tips) of maximum environmental temperature among lizards of the South American family Liolaemidae. See main text for more details.

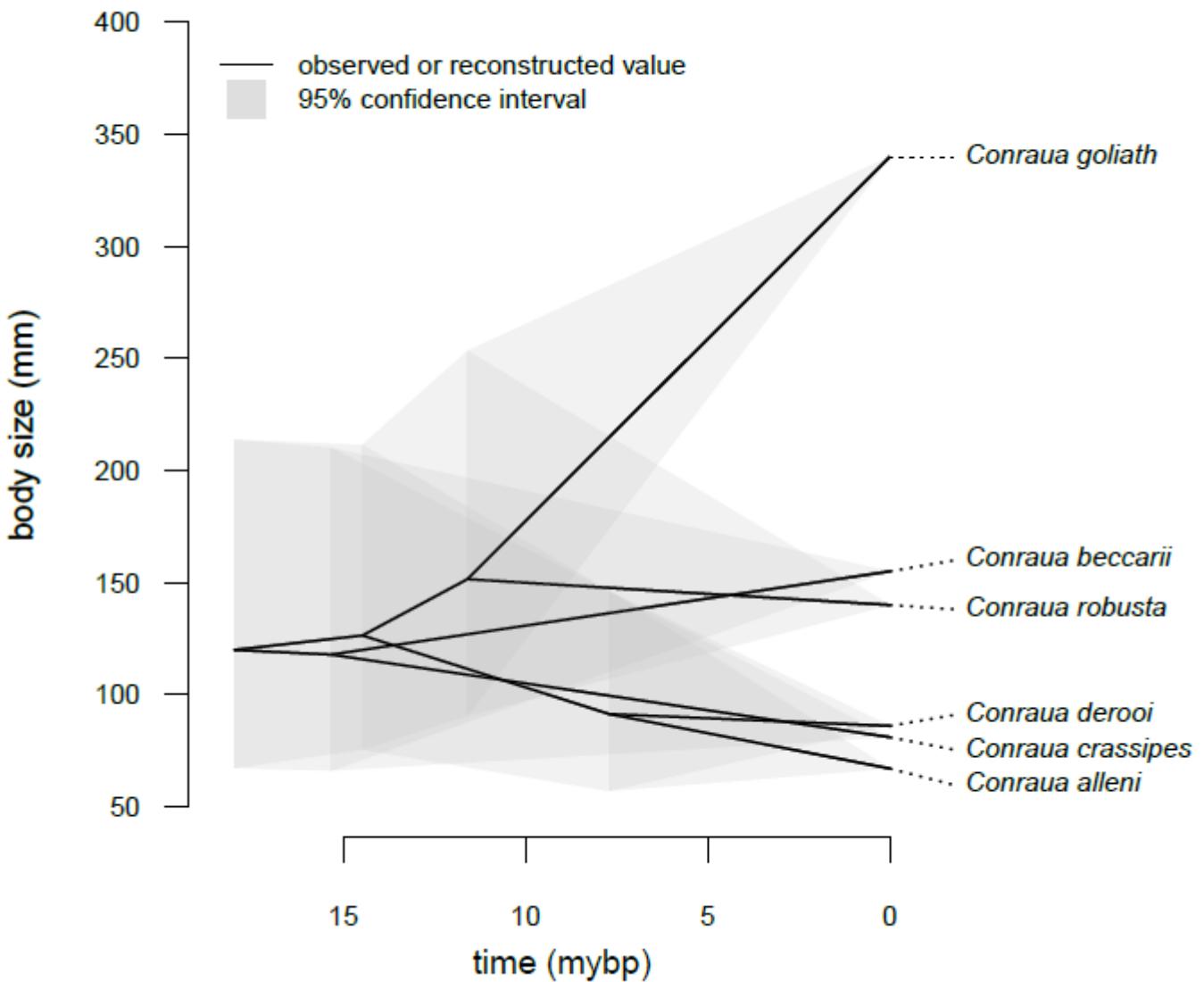
**We can also calculate the uncertainty
of these estimates under our model**

a)

67 body size (mm) 340

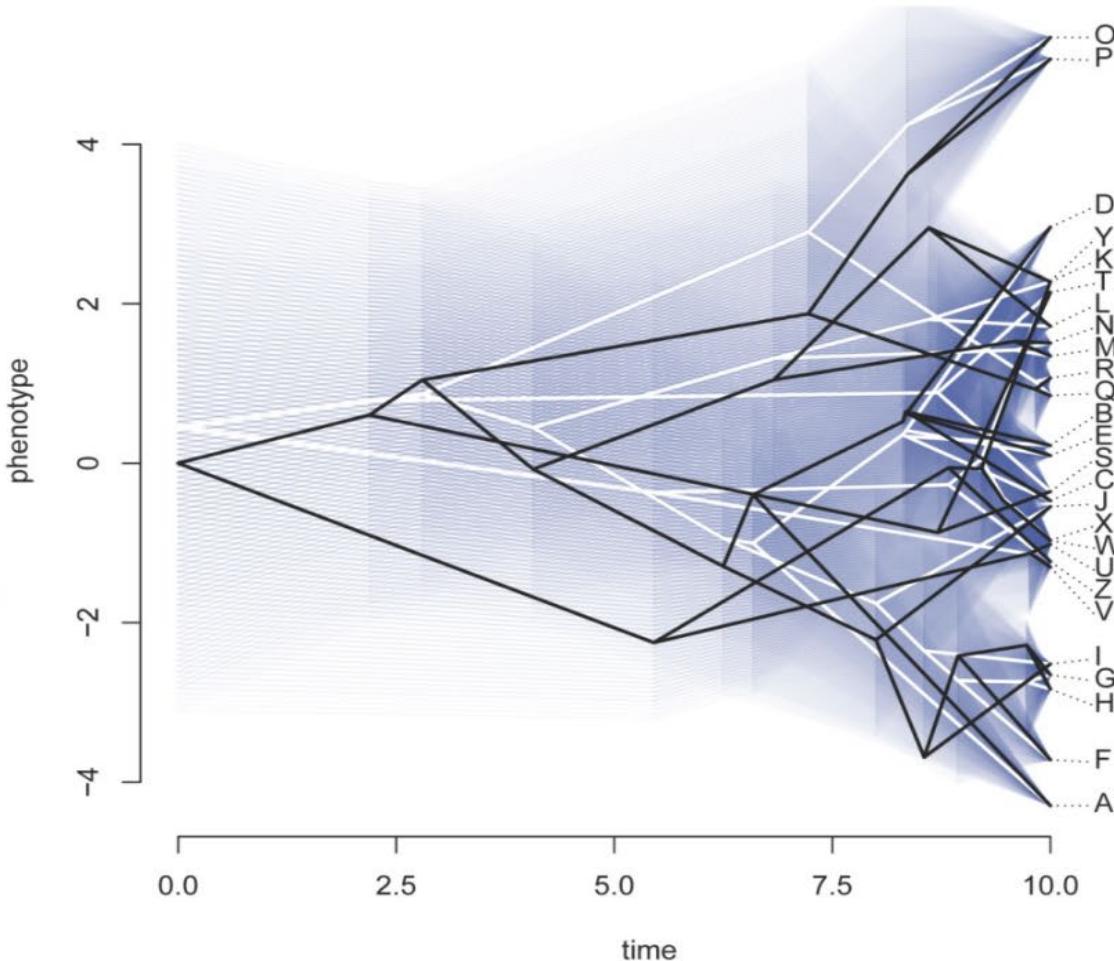


b)

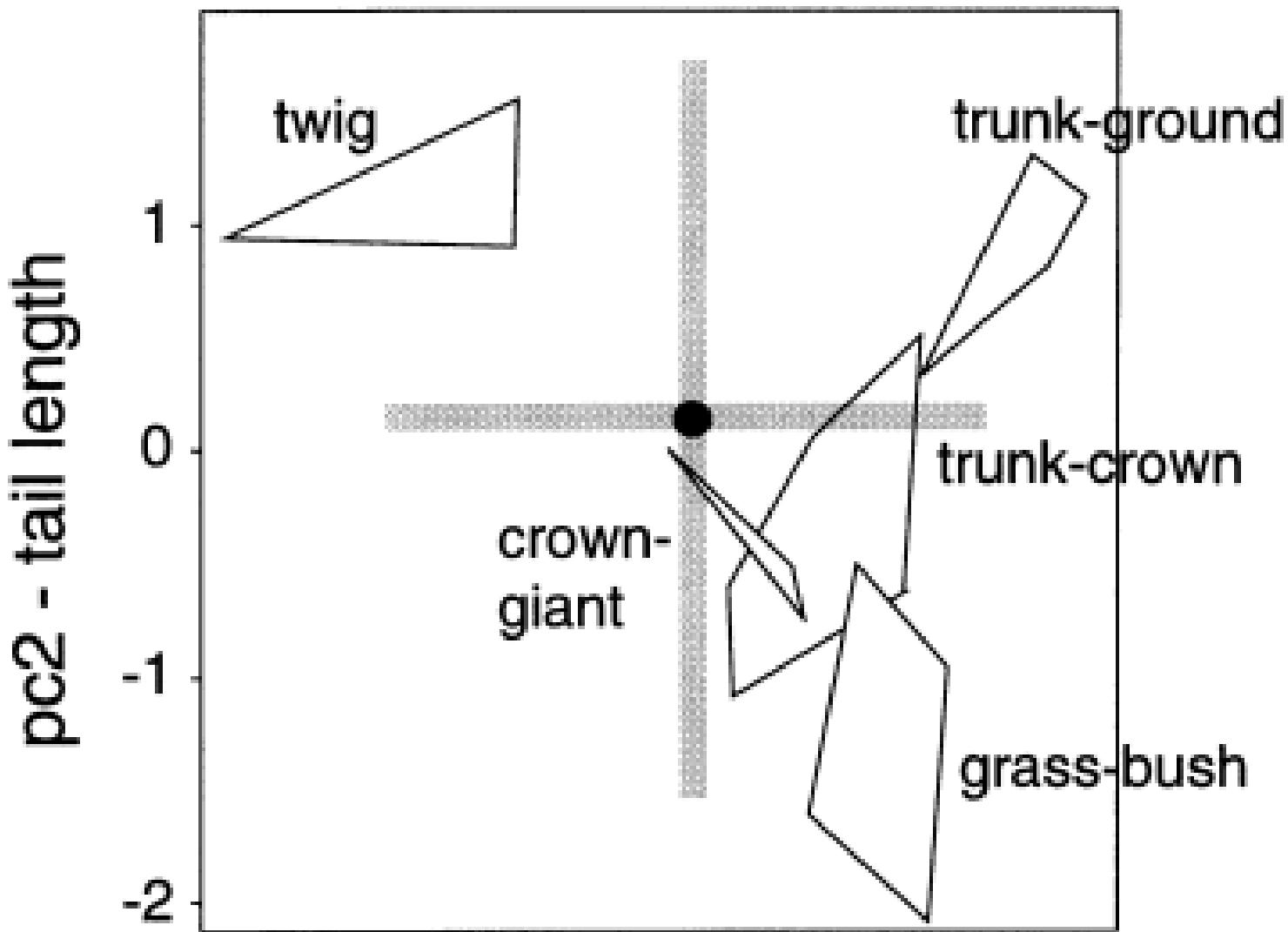


Problem 1:

- The variance on ancestral character estimates is *large*.
- For example, in the figure at right, the 95% CI for the root almost includes all observed values for the tip taxa.
- Saying that the uncertainty is large is not the same as saying ancestral state estimates are *wrong*, however.



We should always
estimate uncertainty for
ancestral states



Schluter et al. 1997

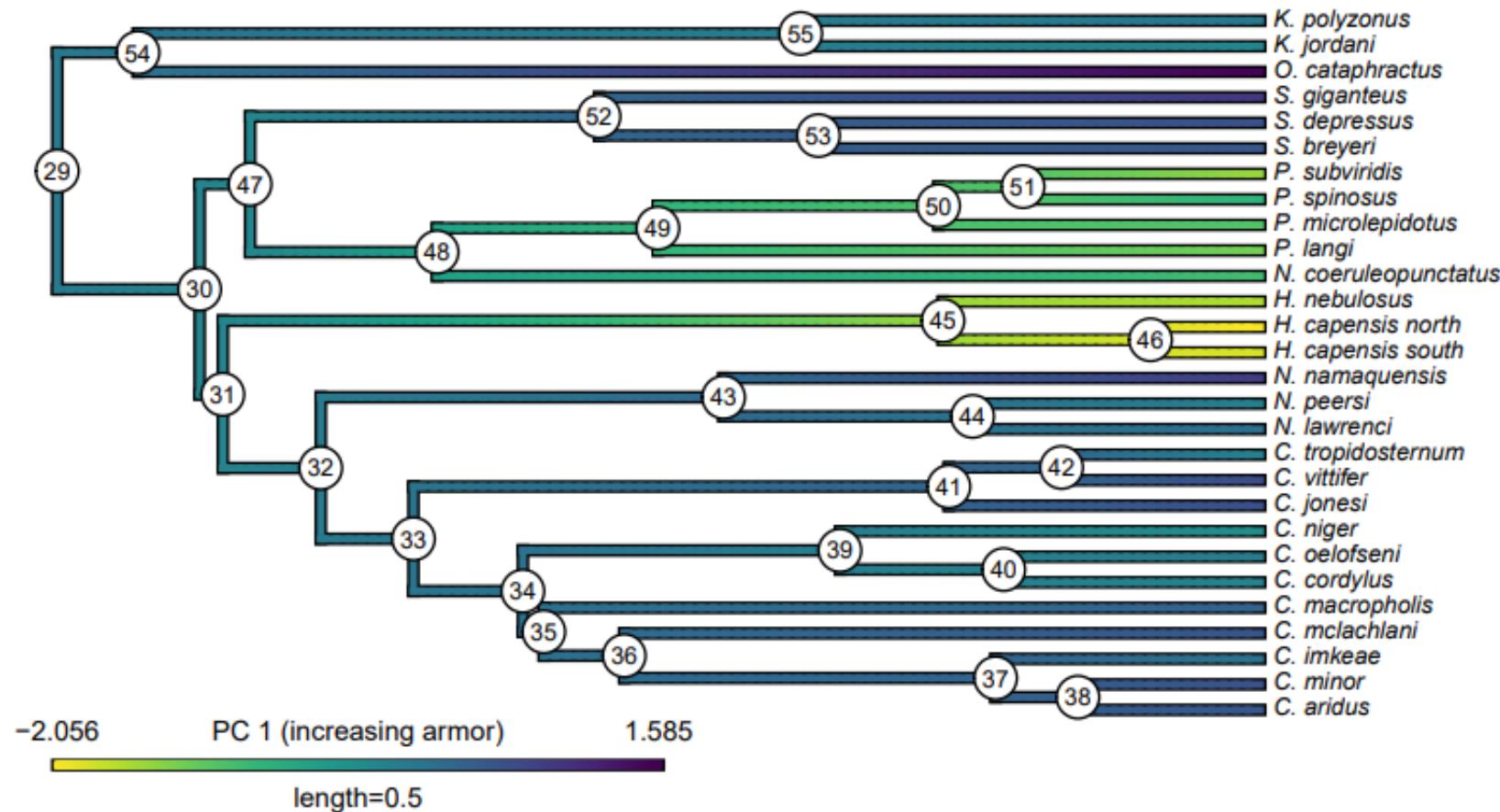


Figure 11 Reconstructed ancestral values from Bayesian MCMC projected onto the nodes and edges of the tree. Numerical values at internal nodes are node indices from our input phylogeny. Data consist of PC 1 from a phylogenetic principal components analysis of cordylid morphological traits, and separate highly armored (high values) from lightly armored (low values) lizards (Broeckhoven et al., 2016). See main text for more details.

Full-size DOI: 10.7717/peerj.16505/fig-11

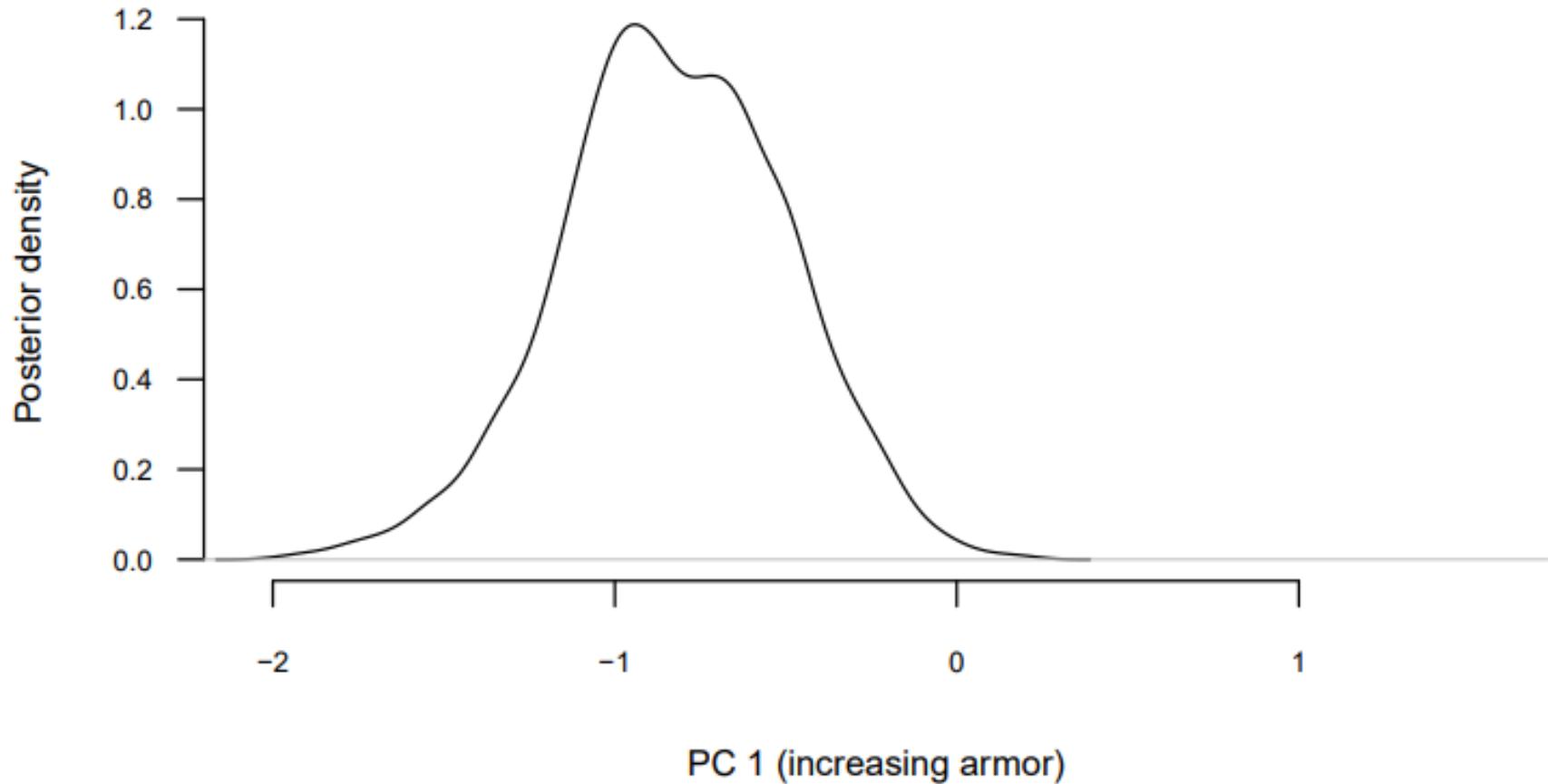


Figure 12 Posterior probability density at node 49 of Fig. 11 from Bayesian MCMC ancestral state reconstruction of PC 1 from a morphological analysis on a phylogenetic tree of cordylid lizards. Node 49 corresponds to the common ancestor of the *Pseudocordylus*: a relatively lightly armored cordylid clade. See main text for more details.

Full-size DOI: 10.7717/peerj.16505/fig-12

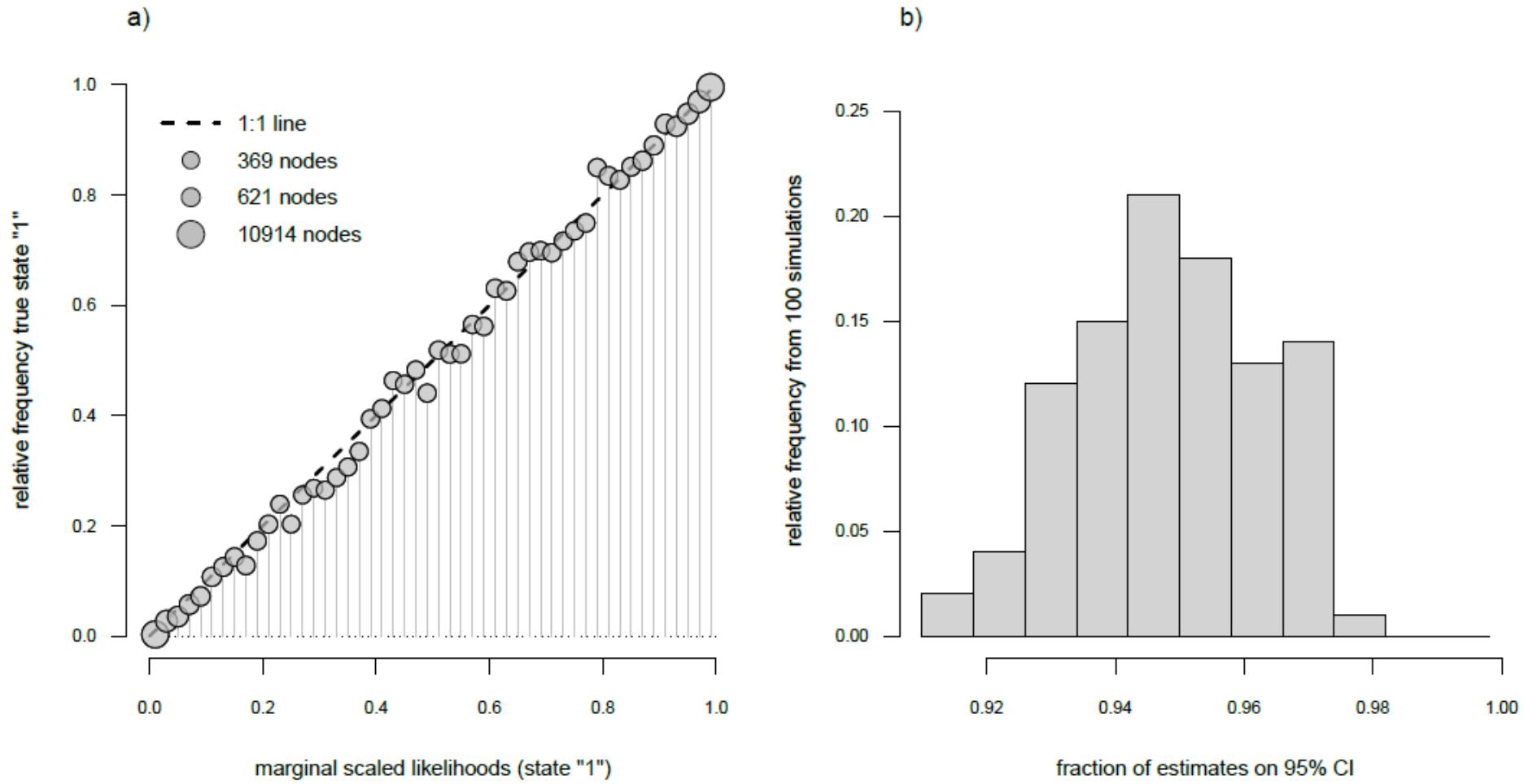
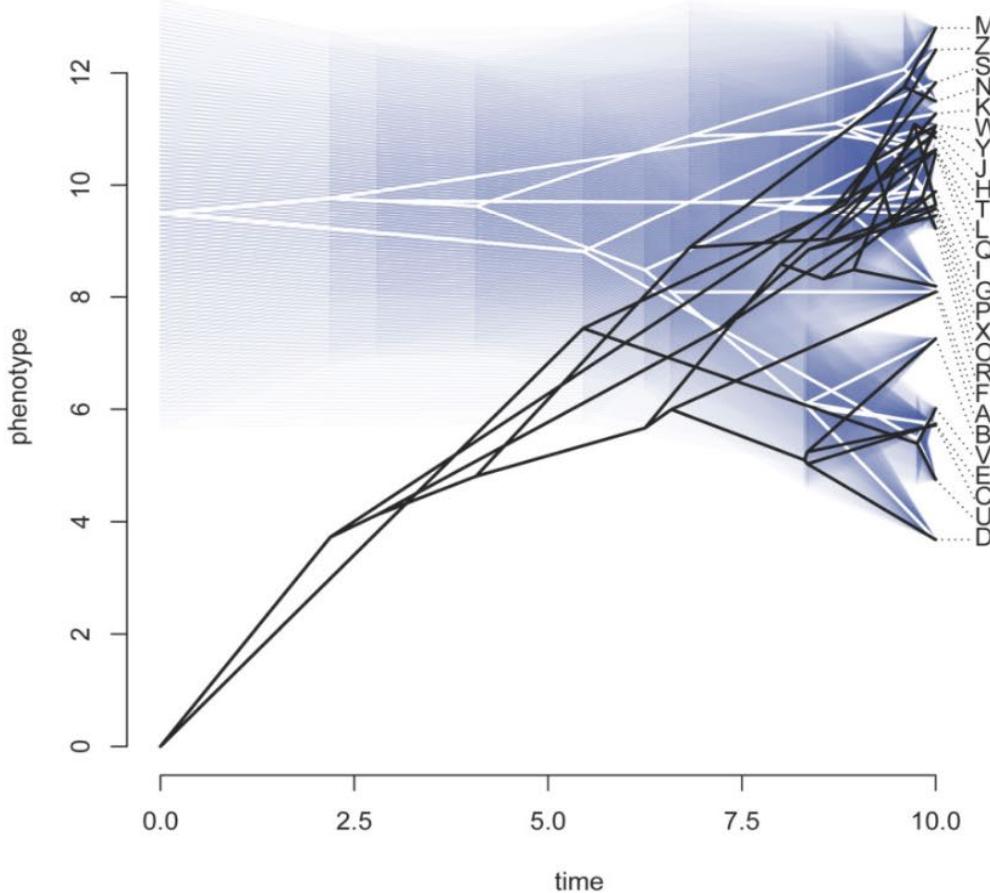


Figure 13: Accuracy of ancestral state reconstruction of discrete (a) and continuous (b) characters when the model for estimation is correct. (a) Node marginal scaled likelihoods (of state “1”) compared to the relative frequency that each node was in that condition. If the scaled likelihoods are an accurate measure of the true probability of that each node was in each character state, then these values should form a 1:1 line. Point diameters have been scaled by the natural logarithm of the sample size (number of nodes) for each bin. (b) Distribution of the relative frequency (from 100 simulations) in which the true ancestral value fell on the 95% confidence interval of each node estimate, averaged across all nodes by simulation. See main text for additional details.

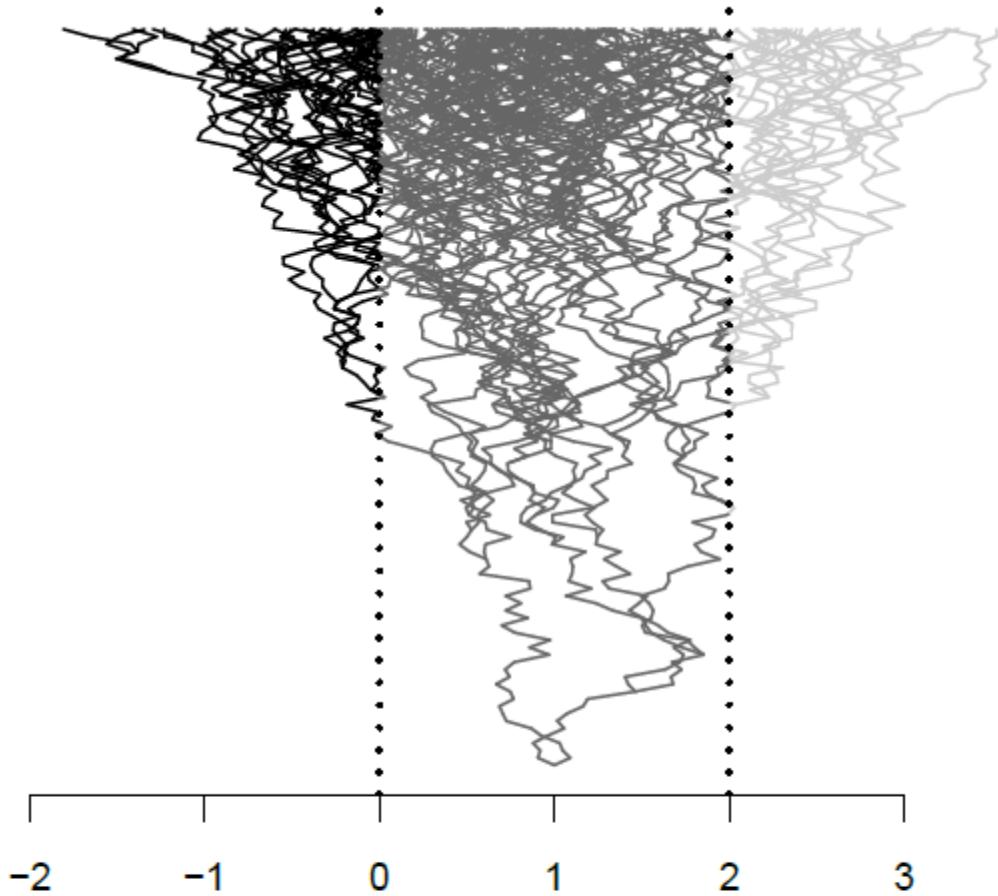
If the assumed model is
wrong, then the
estimates are also going
to be wrong

Problem 2:

- If the model is incorrect, ancestral character estimation is *really* bad.
- For instance, the data at right were simulated with a trend
- This means it is very important that we keep in mind that any hypothesis tests about ancestral character values *dependent* intrinsically on the validity of our fitted model.



a)



b)

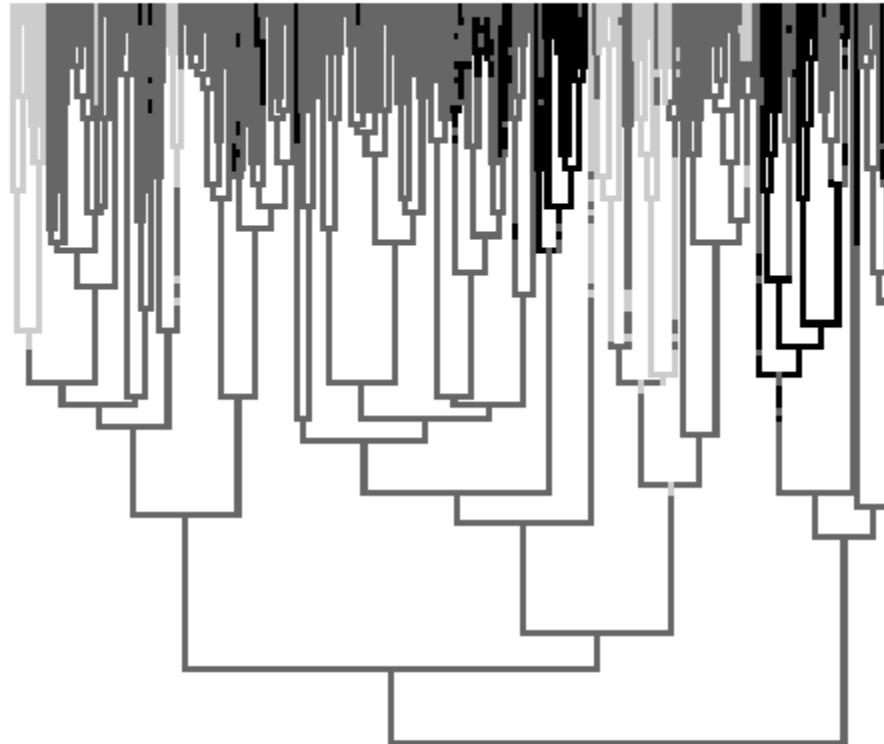


Figure 16: Illustration of evolution under the threshold model. (a) The evolution of liabilities: the unobserved continuous character whose condition determines the state of the discrete trait. The thresholds between discrete character levels in the threshold trait are shown using the vertical dotted lines. (b) The realized evolution of the discrete character across the branches and nodes of the phylogeny. See main text for more details.

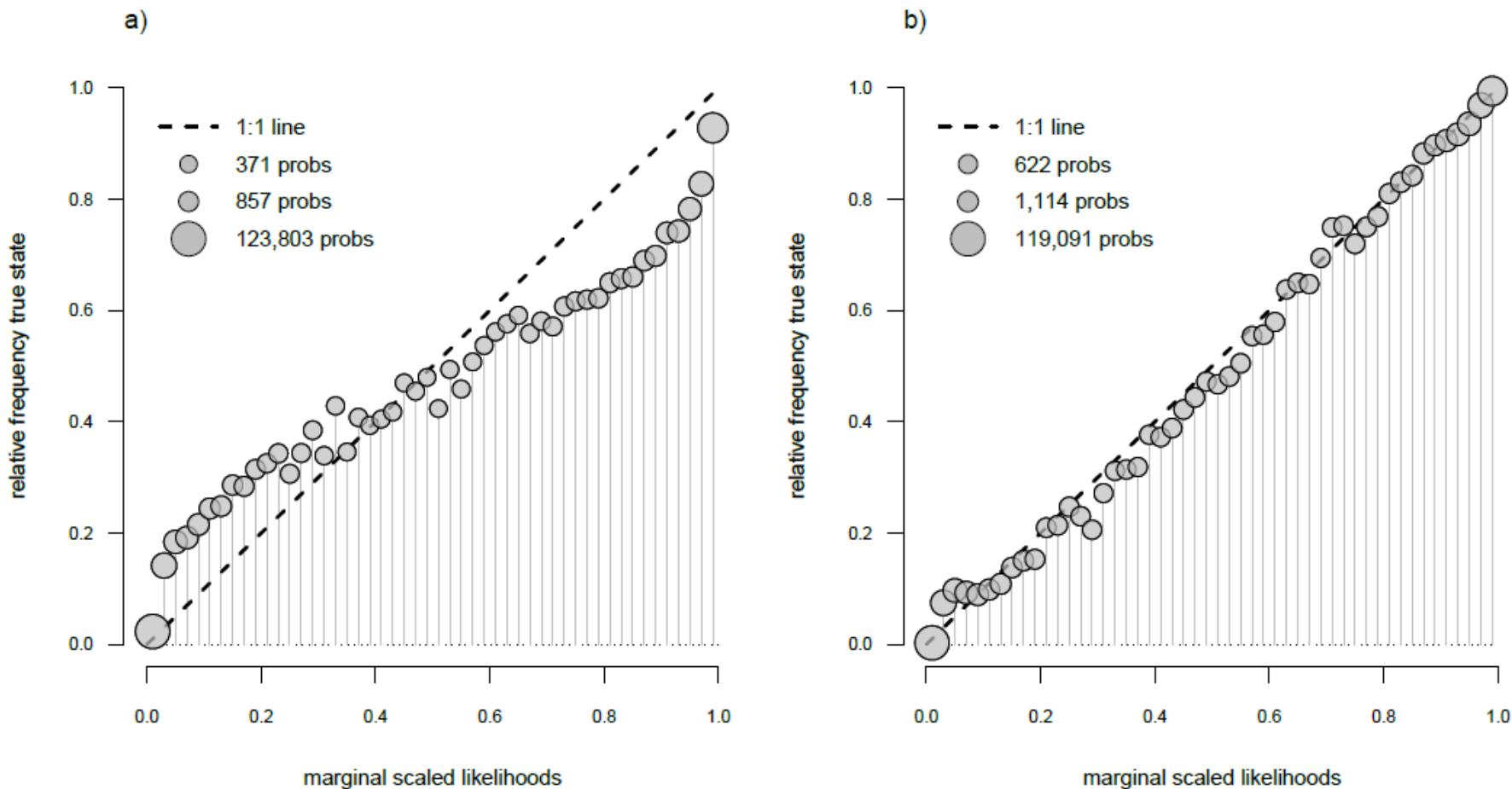
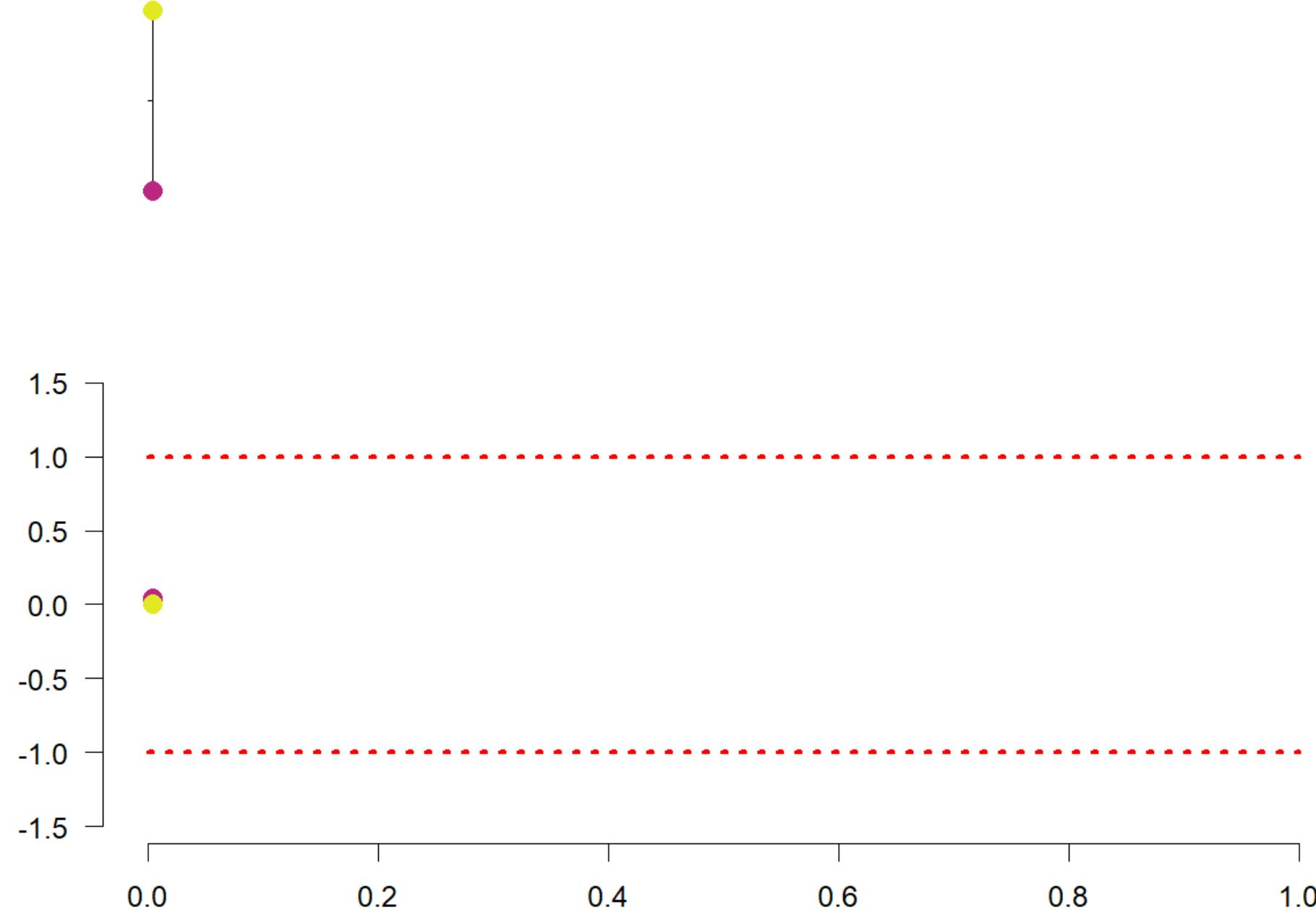


Figure 17: Accuracy of ancestral state reconstruction of discrete characters when the threshold model of Figure 16 was used for simulation. (a) Node marginal scaled likelihoods compared to the relative frequency that each node was in that condition using a standard M_k model for estimation. (b) The same as (a), but in which the generating threshold model was used for estimation. If the scaled likelihoods are an accurate measure of the true probability of that each node was in each character state, then these values should form a 1:1 line. Point diameters have been scaled by the natural logarithm of the sample size (total number of probabilities computed) for each bin. Note that these don't sum to the number of nodes because multiple values are calculated for each node in each simulation, depending on the number of levels (3 or 4) of the trait. See main text for additional details.



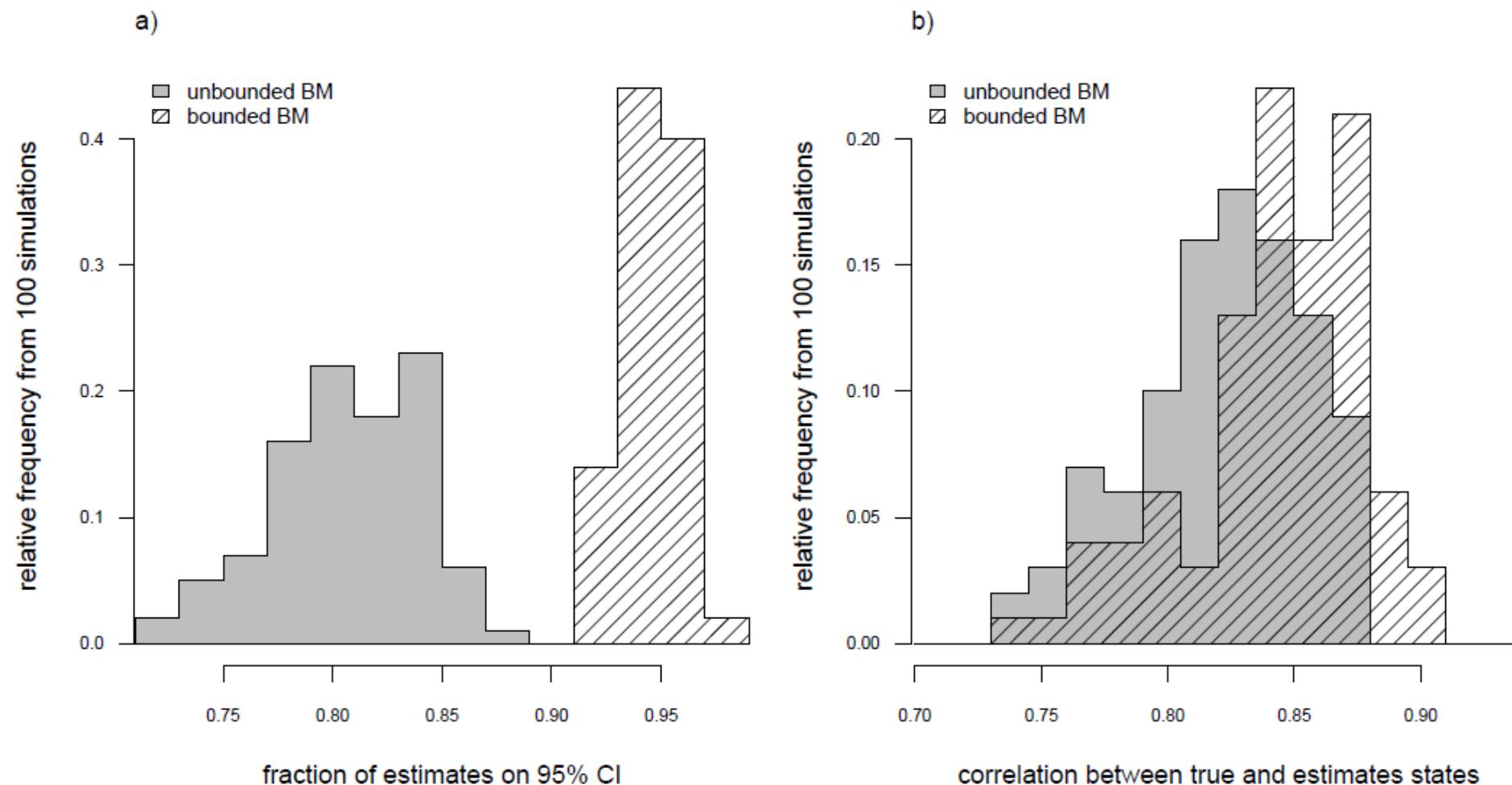


Figure 18: Accuracy of ancestral state reconstruction of continuous characters when data were simulated under Brownian motion evolution with reflective bounds. (a) Frequency distribution of the fraction of nodes falling in within the 95% confidence interval of each node estimate, averaged across all nodes by simulation both when a standard Brownian model (grey) and bounded model (shading lines) was used for estimation. (b) Distribution of correlation between true and estimated ancestral states when the data were generated under bounded Brownian evolution, and either a standard Brownian motion model (grey) or bounded model (shading lines) was used for estimation. See main text for additional details.

We can make our ancestral state estimates more certain by adding direct information from fossils

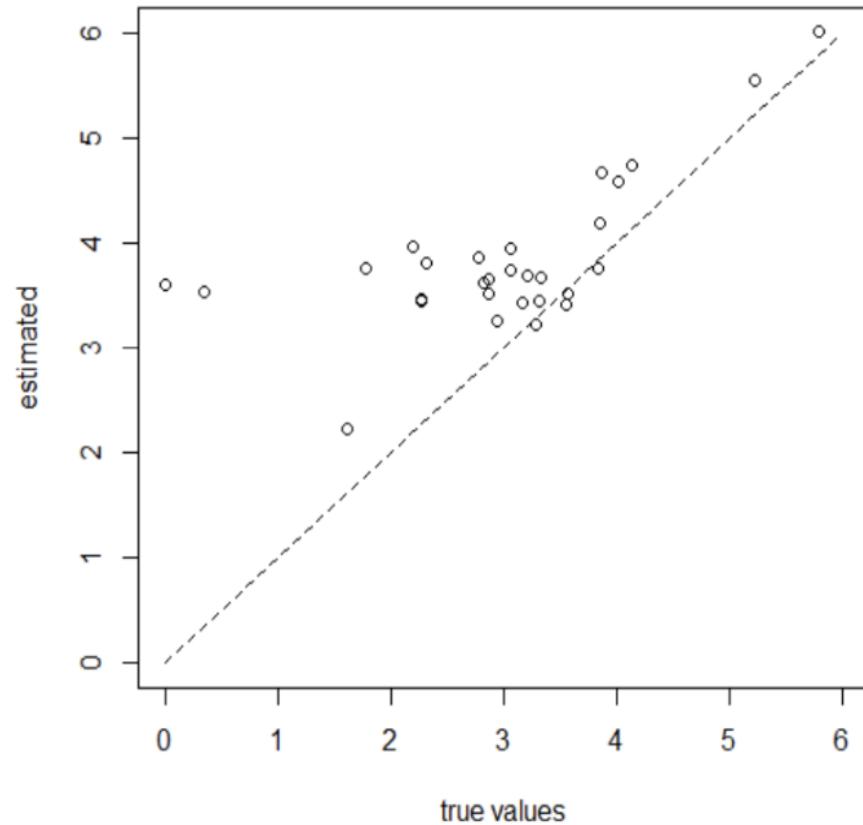


Figure. MLE ancestral states assuming constant rate BM with no prior information about root.

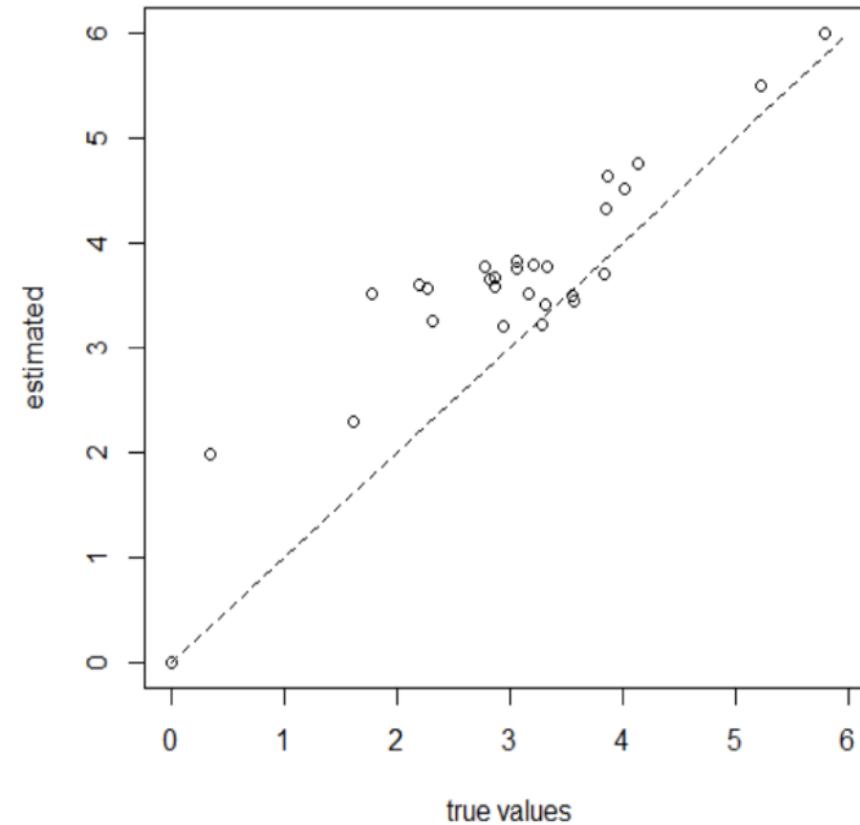


Figure. Bayesian ancestral state estimates with a strong prior density on the root.

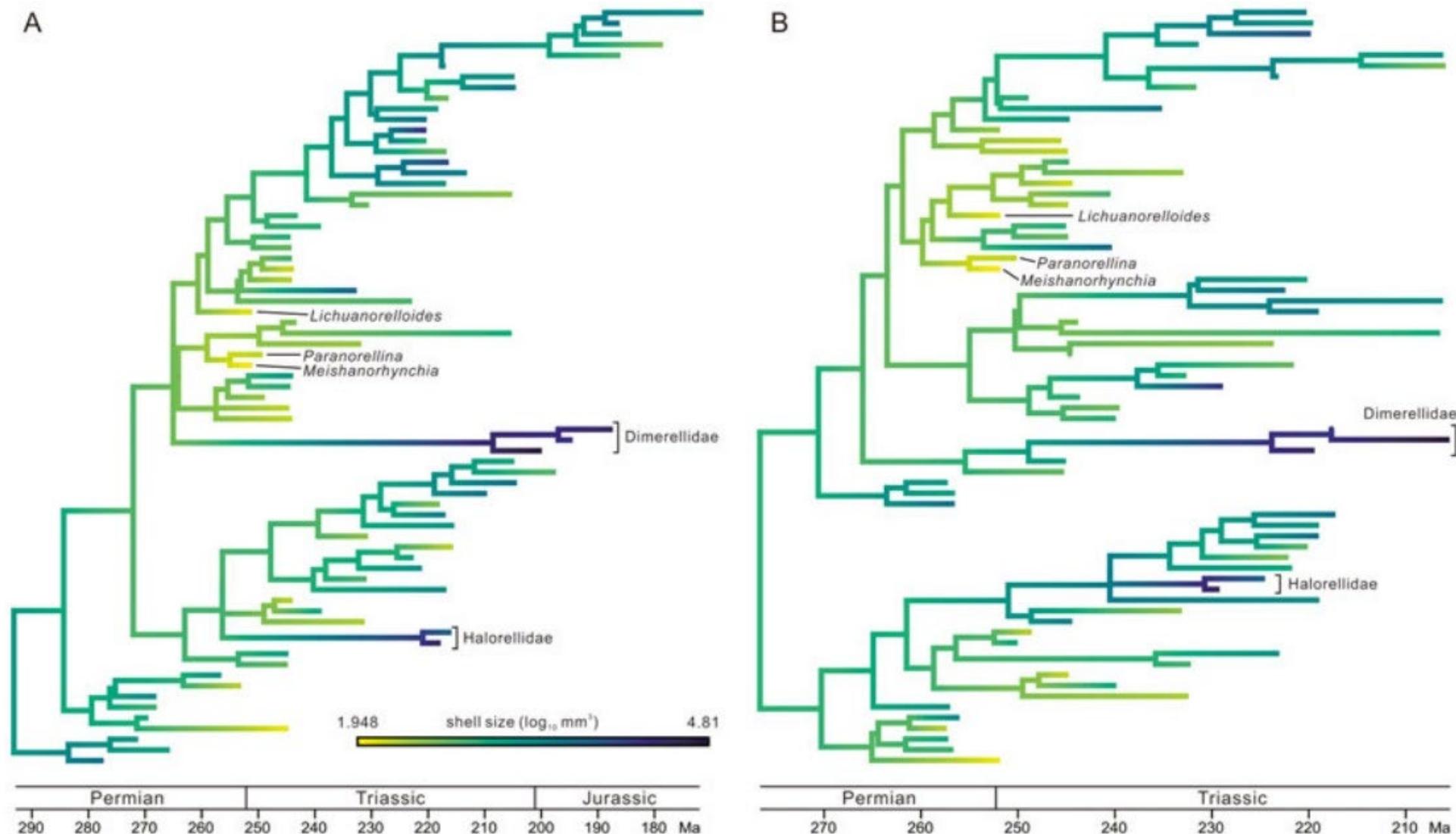


FIGURE 4. Ancestral-state reconstruction of shell size, plotted on the “species-dated” maximum clade credibility (MCC) tree (A) and the “genus-dated” MCC tree (B). Darker color means larger size (online version in color). See Supplementary Material for the same figure with all taxa labeled.