Ancestral state reconstruction for discrete & continuous traits

Liam J. Revell

3 July 2015

Ancestral state reconstruction

- In ancestral character reconstruction our goal is to estimate the ancestral condition of phenotypic traits – usually at internal nodes.
- Ideally, we should also obtain a measurement of the *uncertainty* associated with our ancestral state estimate.

Why do we want to reconstruct ancestral states?

- We have a data set of body size in *Anolis* lizards of the Caribbean, and we are interested in the ancestral value of one or multiple nodes in the tree.

Anolis cuvieri

Anolis occultus

Why do we want to reconstruct ancestral states?

- We have digit number in *Lerista* skinks & we are interested in the ancestral state at the root or the number of times digits have been lost in the group.

Disclaimer

- I’m not going to cover all methods of ancestral character reconstruction – in part because many have been proposed & so doing would be tedious.
- I’m going to concentrate on “statistical” methods for ancestral character reconstruction .
- I define this as including methods with an explicit underlying model (i.e., “model-based”); and for which we can compute a measure of our confidence in our inference to compare against alternative scenarios.

Ancestral state reconstruction

- The first step in ancestral character estimation involves identifying the type of data we are interested in analyzing.
- For instance, we might have data measured on a continuous scale (“continuous characters”) or discrete characters (qualitative features or characteristics that we count).
- The *distinction* between continuous & discrete characters is not always straightforward.
- For instance: *Drosophila* bristle number; scale counts on the midline; etc.

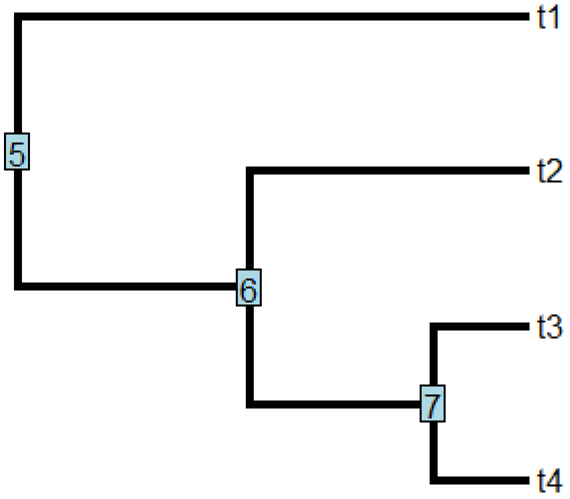
Ancestral state reconstruction

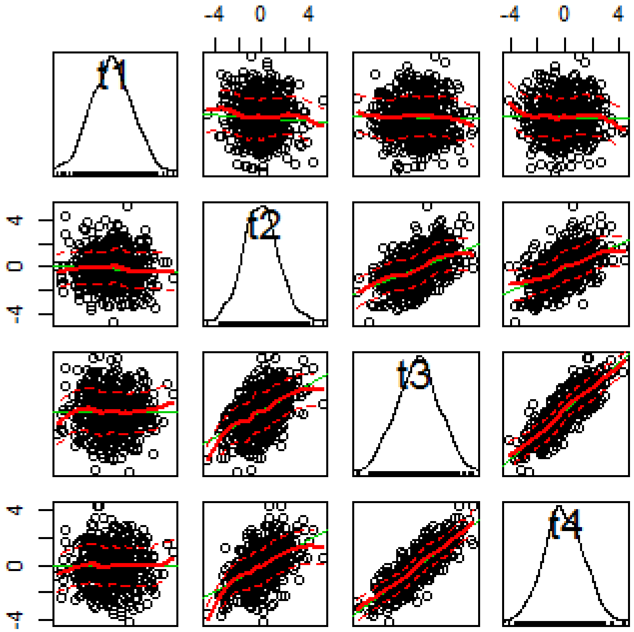
- We need to think not only about the character but also about the *model* that’s appropriate to our data.
- For instance, whether or character is meristic or metric – is it more appropriate to think that it evolves by Brownian evolution (wandering up & down gradually through time); or via (more or less) instantaneous leaps between state.

Brownian motion

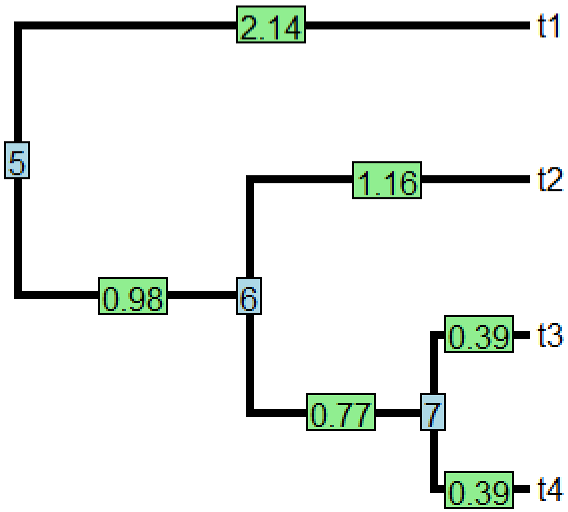
- Remember Brownian motion?
- Brownian motion is a continuous-time stochastic process.
- The expected distribution of the phenotypic trait data at the tips the tree is *multivariate normal*.

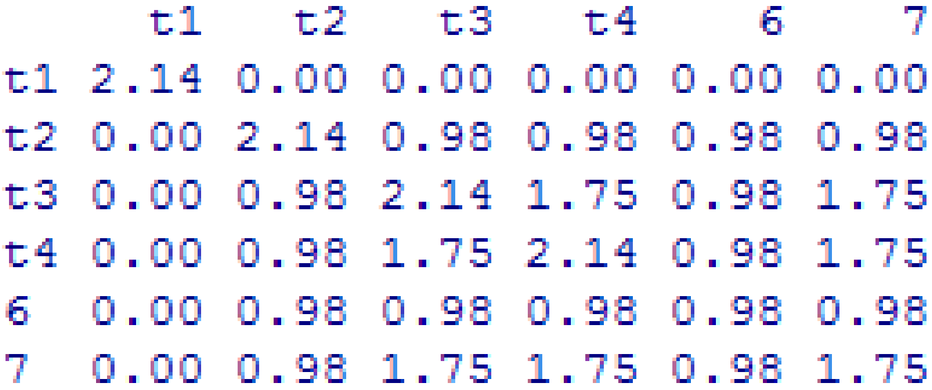
Brownian motion (on a phylogeny)





Brownian motion (on a phylogeny)





```
      t1   t2   t3   t4    6    7
t1 2.14 0.00 0.00 0.00 0.00 0.00
t2 0.00 2.14 0.98 0.98 0.98 0.98
t3 0.00 0.98 2.14 1.75 0.98 1.75
t4 0.00 0.98 1.75 2.14 0.98 1.75
6  0.00 0.98 0.98 0.98 0.98 0.98
7  0.00 0.98 1.75 1.75 0.98 1.75
```

The expected distribution of the tips & *nodes* of the tree under Brownian motion is multivariate normal with variance-covariance matrix in which each *i,j*th term is *proportional* to the height above the root for the common ancestor of *i and j*.

Ancestral state construction under Brownian motion

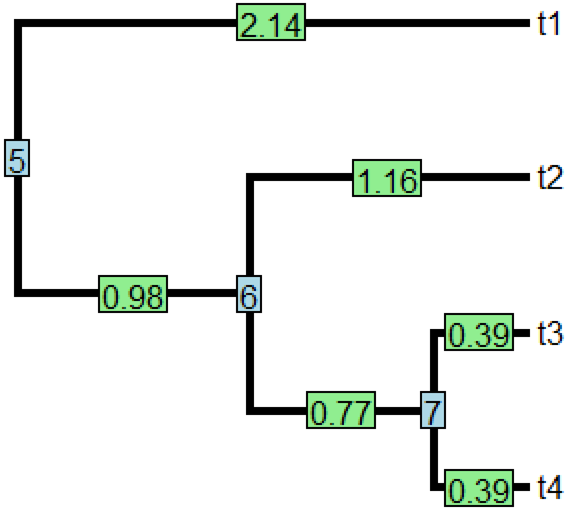
- The tips & nodes of the tree have a multivariate normal density.
- One choice of ancestral states that would make sense is to pick the set of ancestral states that maximize the probability of our data & tree.

$$L(\mathbf{a}, a_0, \sigma^2 \mid \mathbf{T}, \mathbf{x}) = \frac{\exp[-\tfrac{1}{2}([\mathbf{x}, \mathbf{a}] - a_0 \mathbf{1})'(\sigma^2 \mathbf{T})^{-1}([\mathbf{x}, \mathbf{a}] - a_0 \mathbf{1})]}{\sqrt{(2\pi)^{n+m-1} \left|\sigma^2 \mathbf{T}\right|}}$$

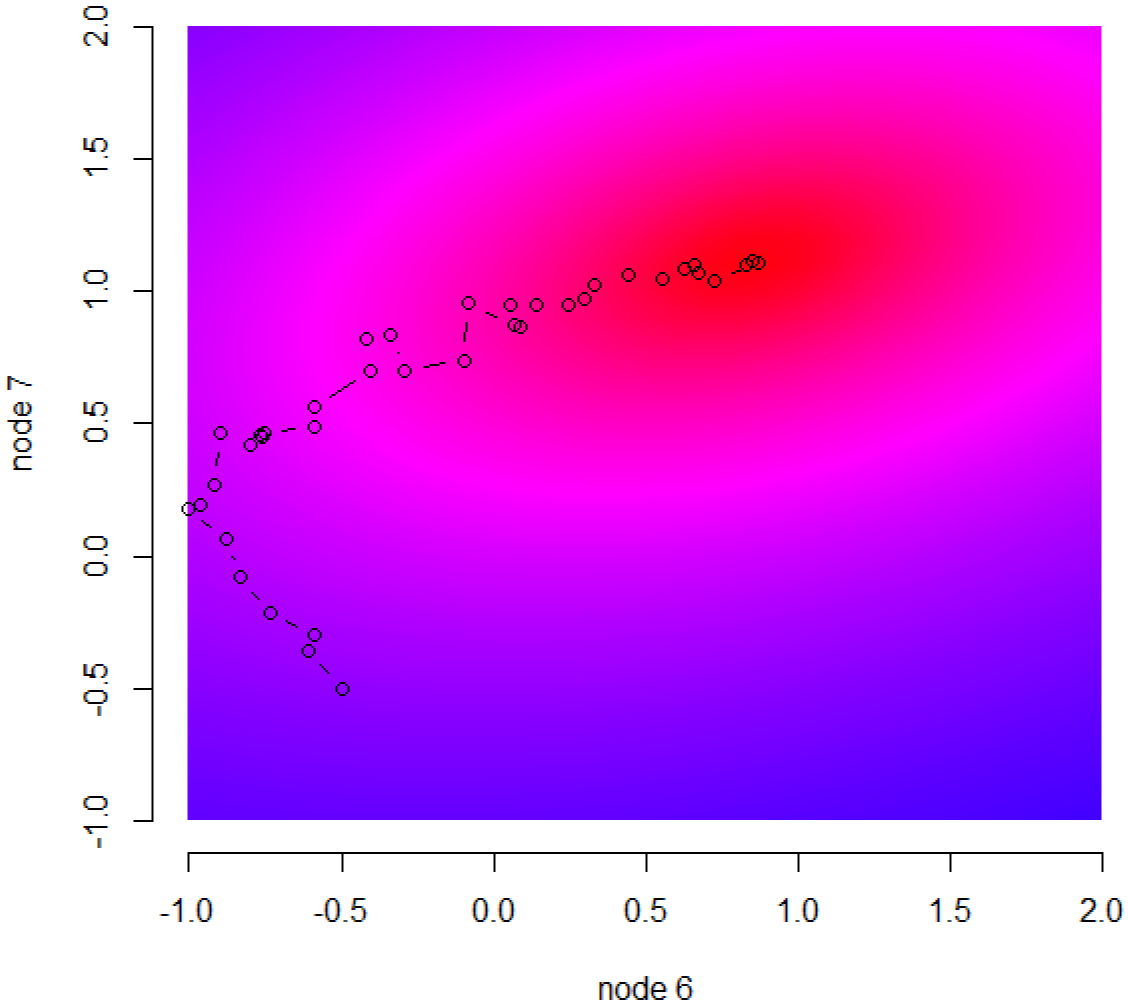
- These ancestral states are the *maximum likelihood estimates* (MLEs).

Ancestral state construction under Brownian motion

- How do we find the ancestral states that maximize the likelihood?
- We could simply try different values for the ancestral states to try and find values that maximized the likelihood…..



```
       t1        t2        t3        t4
0.0540405 0.7435602 1.1046901 1.2590358
```



This does work – however it will become *extremely* inefficient as the number of dimensions grows.

Ancestral state construction under Brownian motion

- There are a variety of routines for numerical multivariate optimization.
- However in the case of ancestral state estimation assuming Brownian motion, we can do even better by taking advantage of the contrasts algorithm.
- It turns out to be the case that the root node estimated during the contrasts algorithm is also the MLE of the root.
- To get the MLEs at every other node in the tree, we can just re-root at that node.

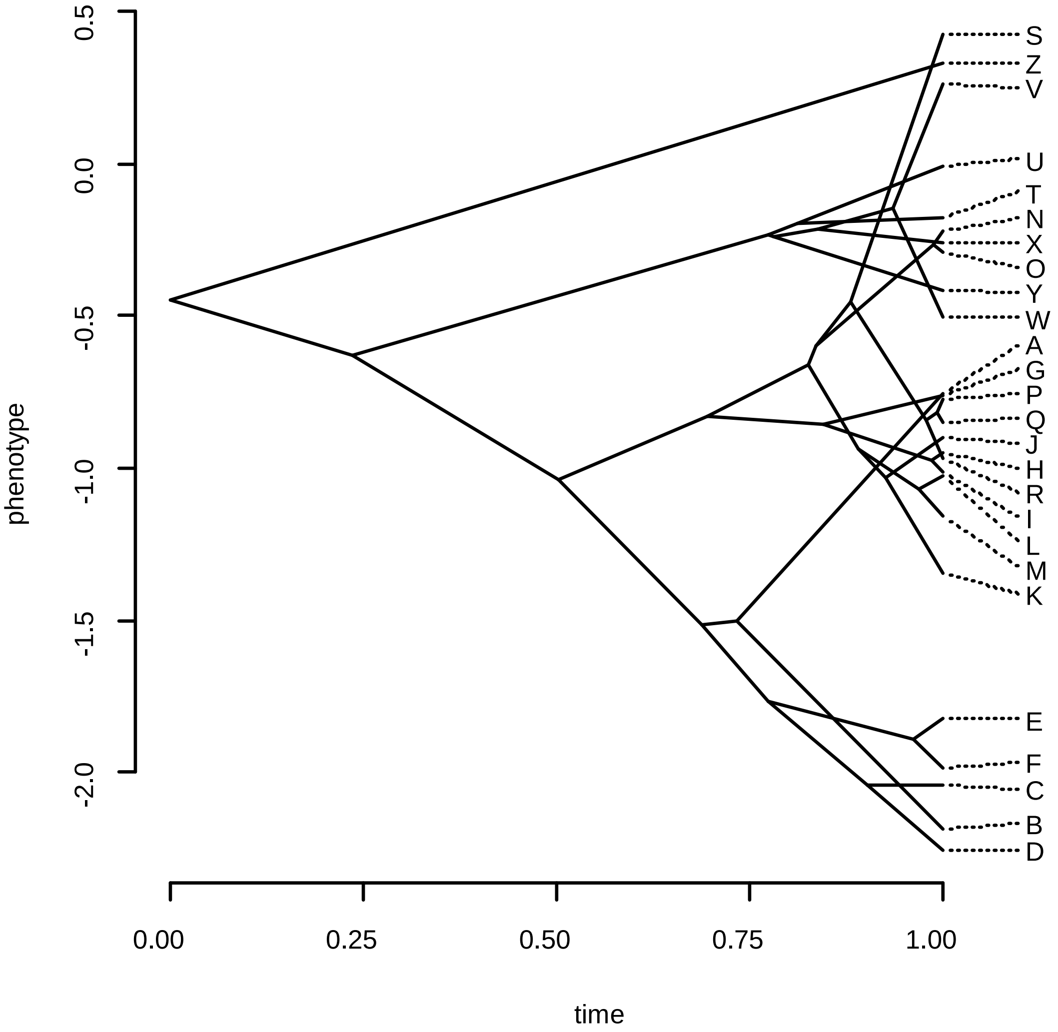


Figure. A projection of the tree into phenotype space. The vertical position of internal nodes correspond with MLE ancestral states.

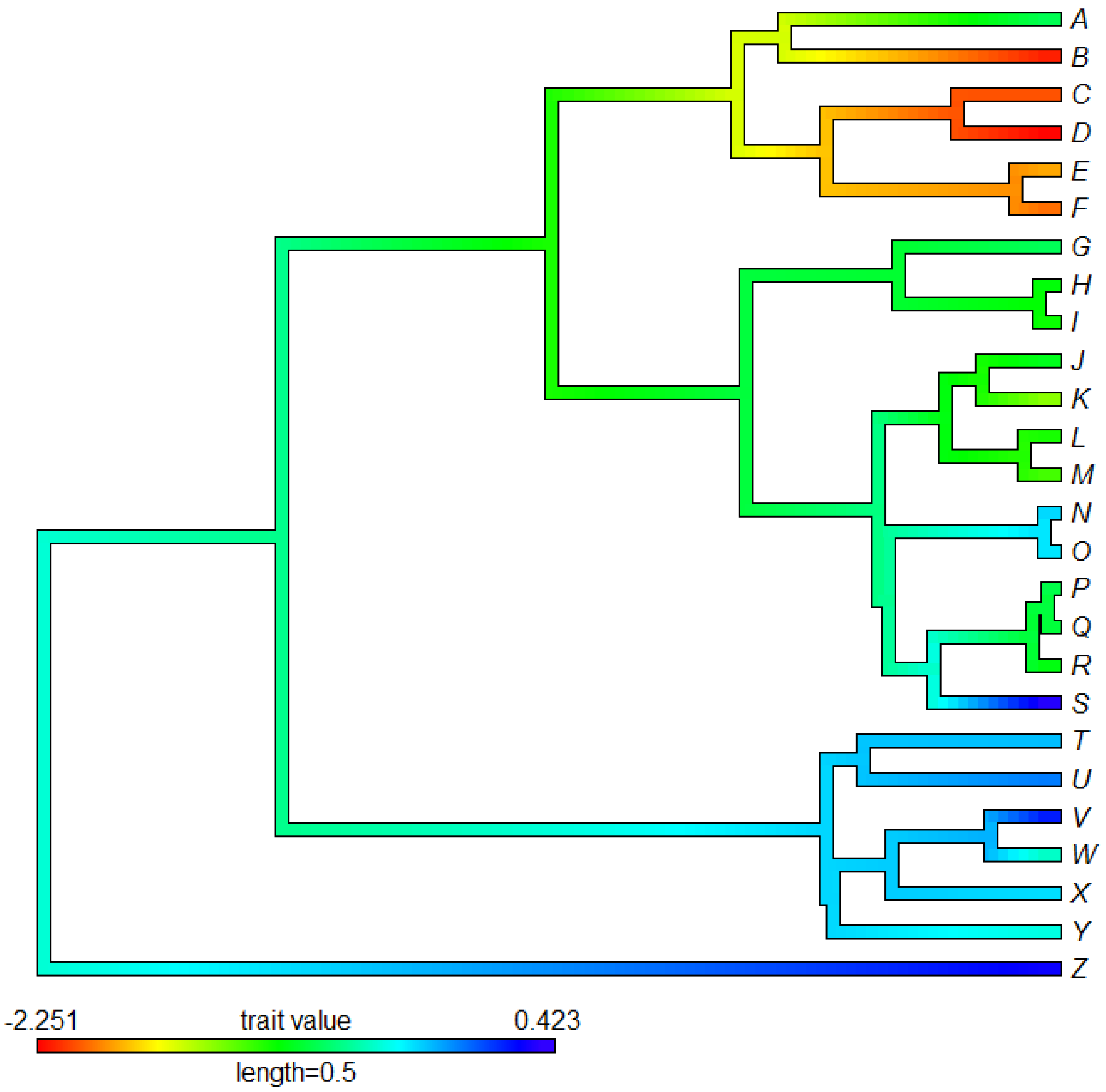
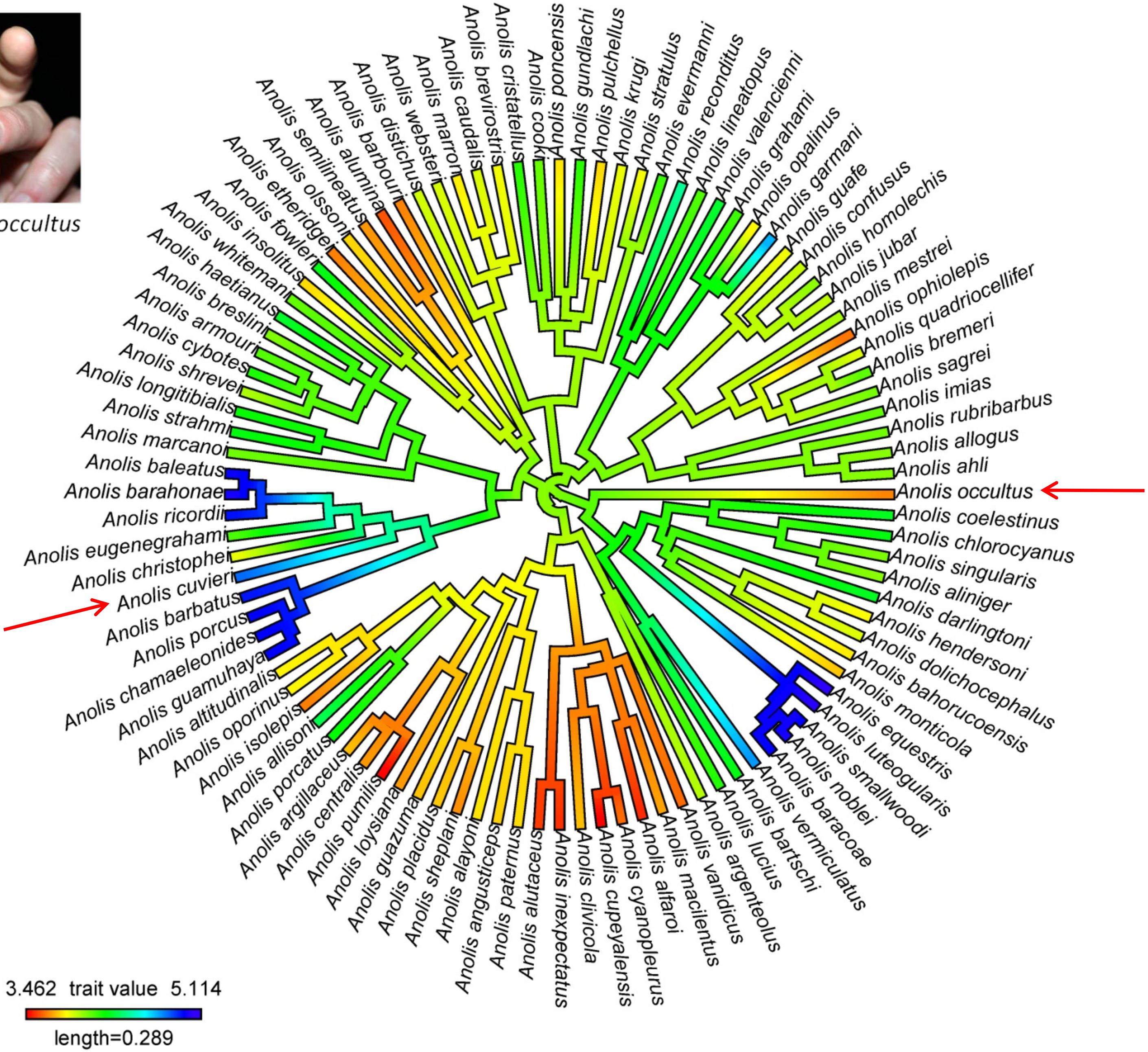


Figure. A projection of the observed & reconstructed trait values onto the phylogeny.

Anolis cuvieri

Anolis occultus

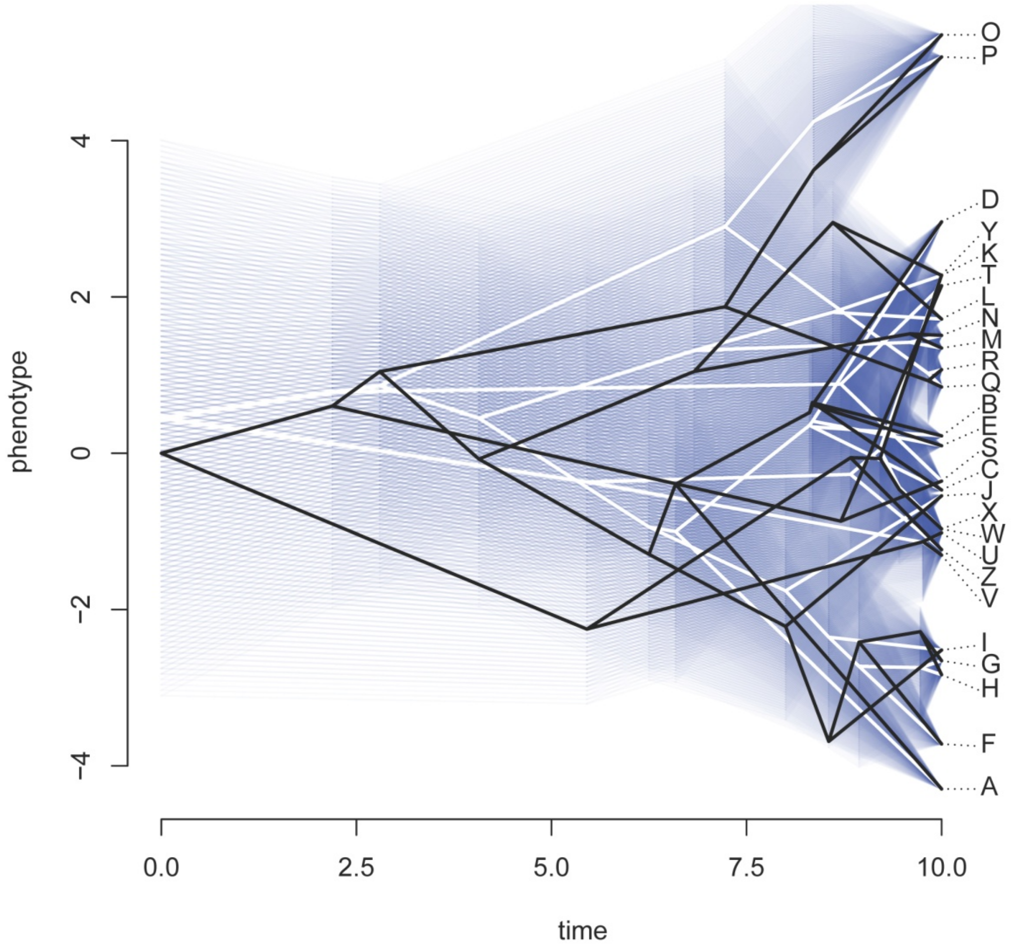


But what about uncertainty….

- We could estimate the variance (uncertainty) in our fitted ancestral states using the Hessian matrix.
- However, we are lucky here too – because there are also analytical solutions (and the Hessian turns out to be quite bad for relatively small trees).
- We can use the variances to compute 95% CI around ancestral values, and test any hypothesis we might have about ancestral states in our character of interest.

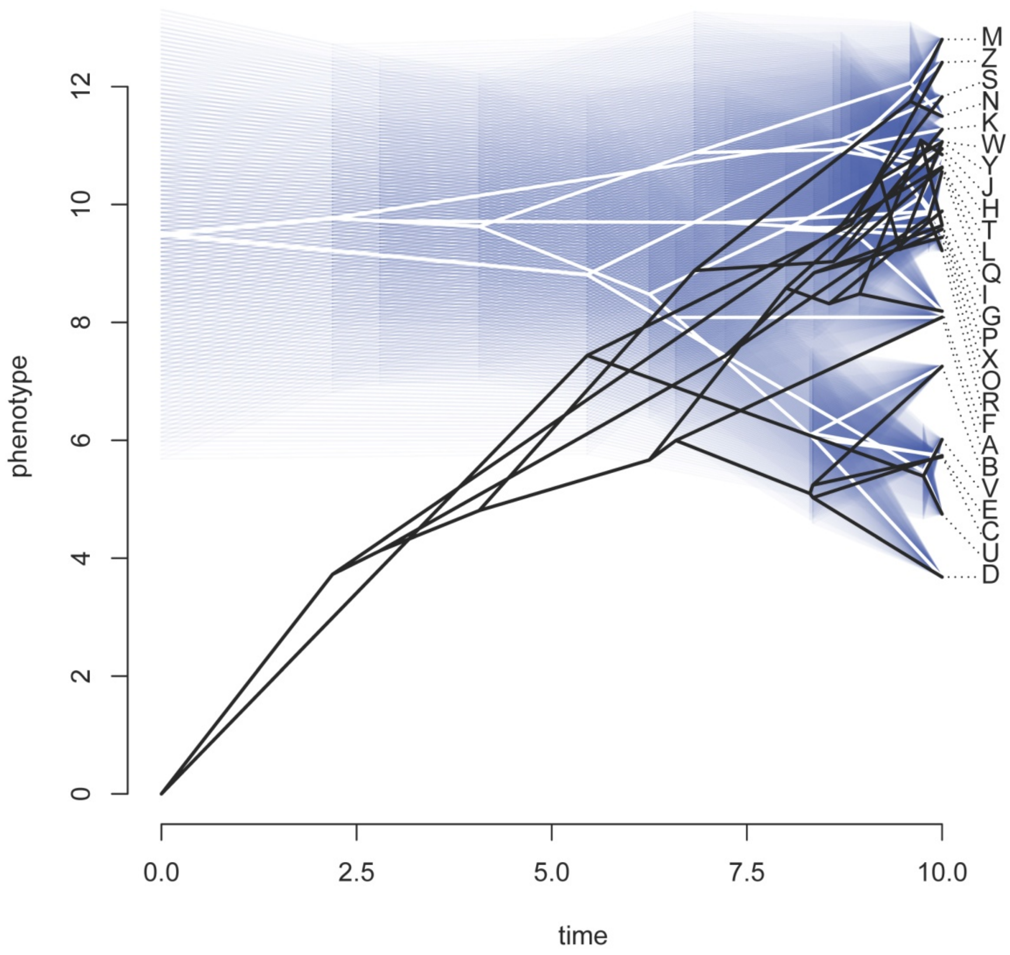
Problem 1:

- 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.



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.



What about fossils?

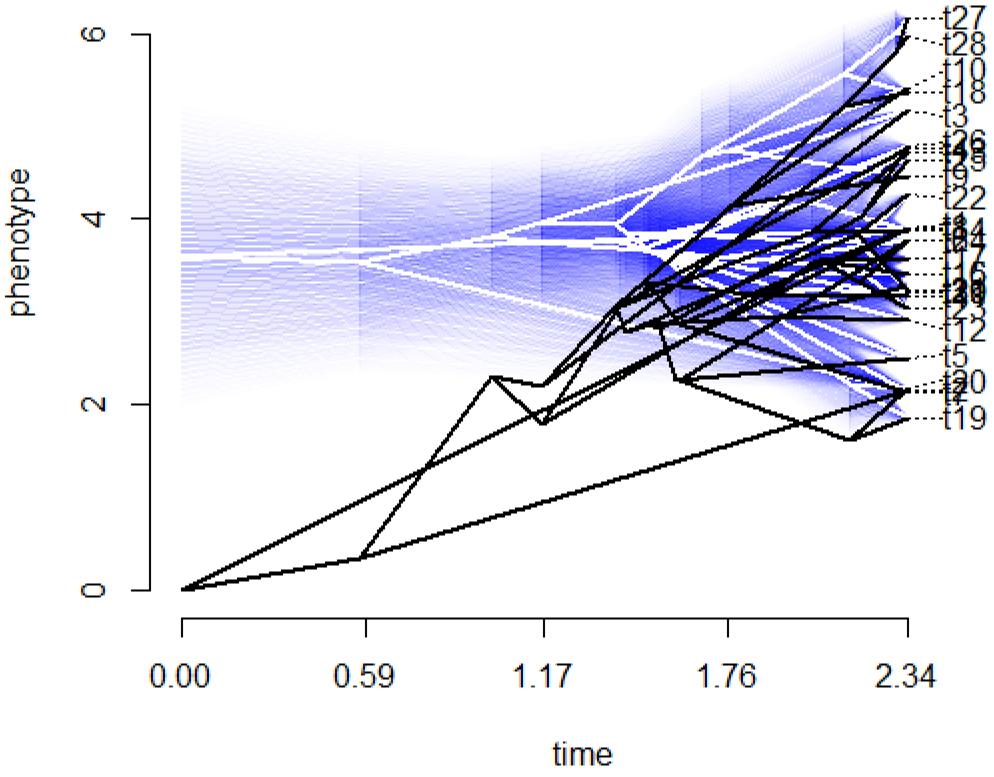


Figure. 95% CI traitgram (blue); true trait history (black). Recovering a trend is hard!

What about fossils?

- We can use *Bayesian* ancestral character estimation to incorporate prior information about the root nodes.
- For instance, we can impose an informative prior distribution on one or multiple nodes based on information from the fossil record about ancestral phenotypes

What about fossils?

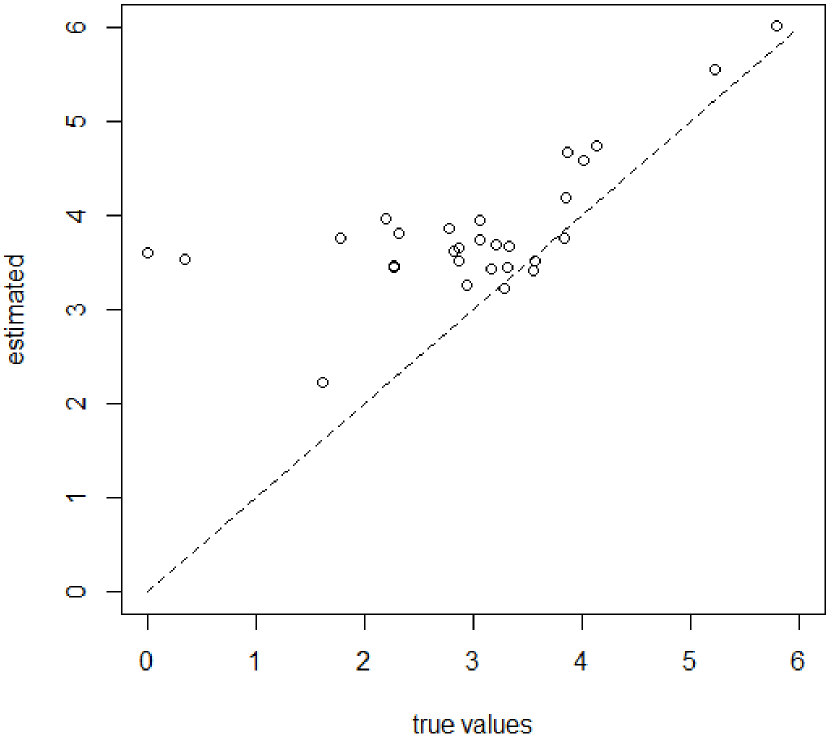


Figure. MLE ancestral states assuming constant rate BM with no prior information about root.

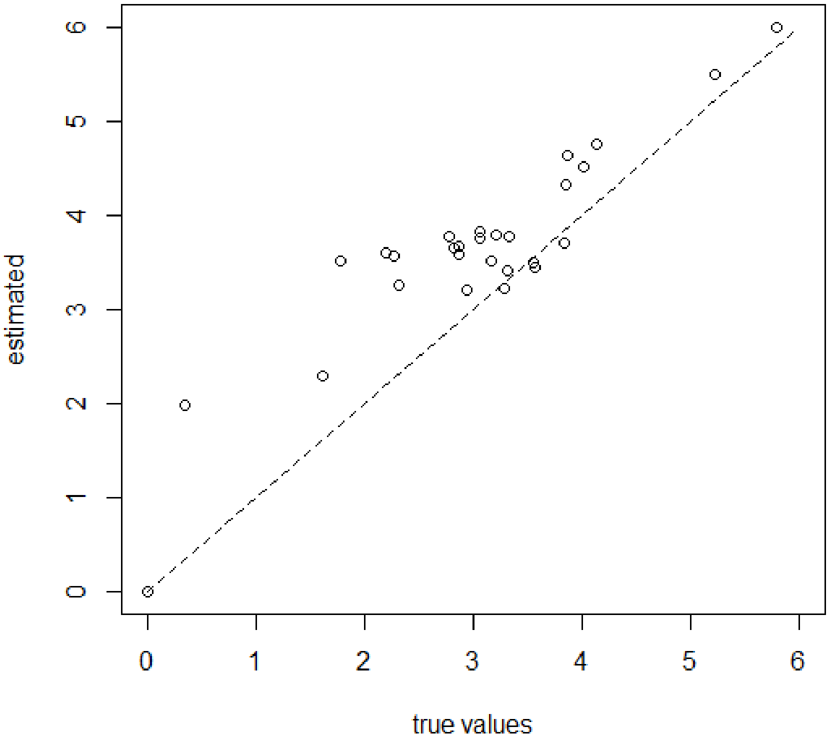


Figure. Bayesian ancestral state estimates with a strong prior density on the root.

Conclusions from ancestral state reconstruction of continuous traits

- We can estimate states using likelihood.
- This approach is unbiased (if our model is correct) and our 95% CIs accurately reflect uncertainty about our estimates.
- *However*, uncertainty can be very large – making inference about ancestral nodes (particularly deep in the tree) difficult.
- Furthermore, if our model of evolution is “badly” wrong – our estimates about ancestral character states can be very biased.
- Independently of our model – we will tend to get better ancestral estimates if we have *prior* information about the states at some nodes in the tree.

Discrete characters

- The most commonly used model for discrete character evolution on trees is a model called the *Mk* model.
- *M* stands for Markov – because the modeled process is a continuous-time Markov chain; and *k* because the model is generalized to include an arbitrary number (*k*) states.
- The central attribute of the *Mk* model is a transition matrix, ***Q*****.**
- ***Q*** gives the instantaneous transition rates between states.
- The rows (or columns, depending on the convention) must sum to zero.
- And we can compute the probability of being in each state after time *t* as:

$$\mathbf{Q} = \begin{bmatrix} q_{00} & q_{01} \\ q_{10} & q_{11} \end{bmatrix}$$

$$\mathbf{p}_t = \exp(\mathbf{Q}t)\mathbf{p}_0$$

Joint vs. marginal reconstruction

- An important distinction in ancestral character reconstruction for discrete characters is *joint* vs. *marginal* reconstruction.

- *Joint reconstruction* is finding the set of character states at all nodes that (jointly) maximize the likelihood.

- *Marginal reconstruction* is finding the state at the current node that maximizes the likelihood integrating over all other states at all nodes, in proportion to their probability.

Marginal reconstruction

- We perform marginal ancestral state reconstruction by at each node computing the set of empirical Bayesian posterior probabilities that each node is in each state.

$$P(x = x_j \mid \mathbf{x}, T, \theta) = \frac{\pi_j L(x_j)}{\sum_i \pi_i L(x_i)}$$

- This is equivalent (and sometimes referred to) as the *scaled likelihoods* – because (if the prior is ignored) the empirical Bayes posterior is the same as scaling the likelihood of *x*=*i* but the sum of the likelihoods that *x* is any *i*.

Marginal reconstruction

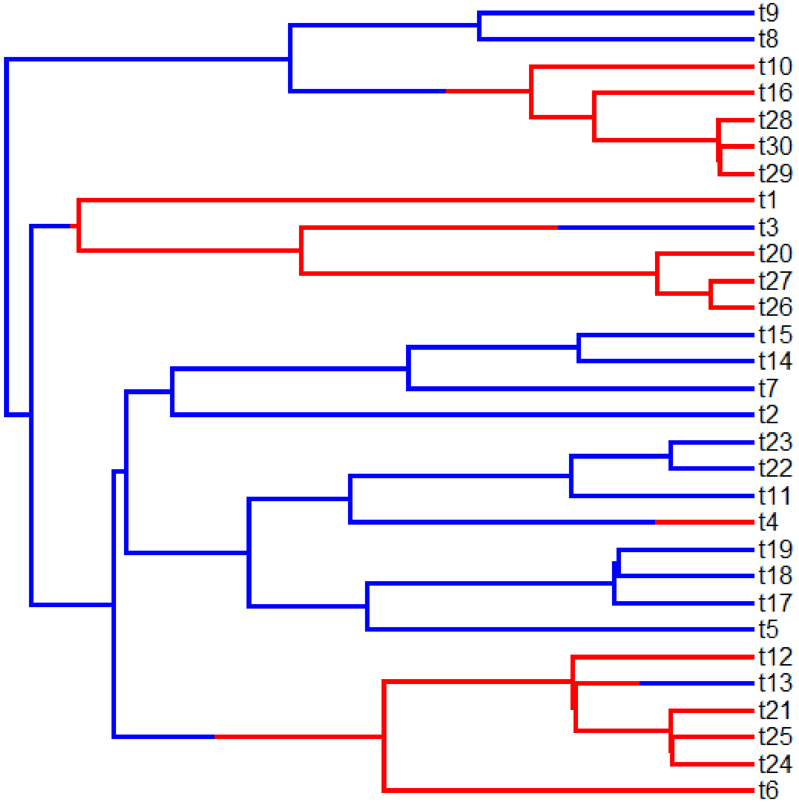


Figure. True history.

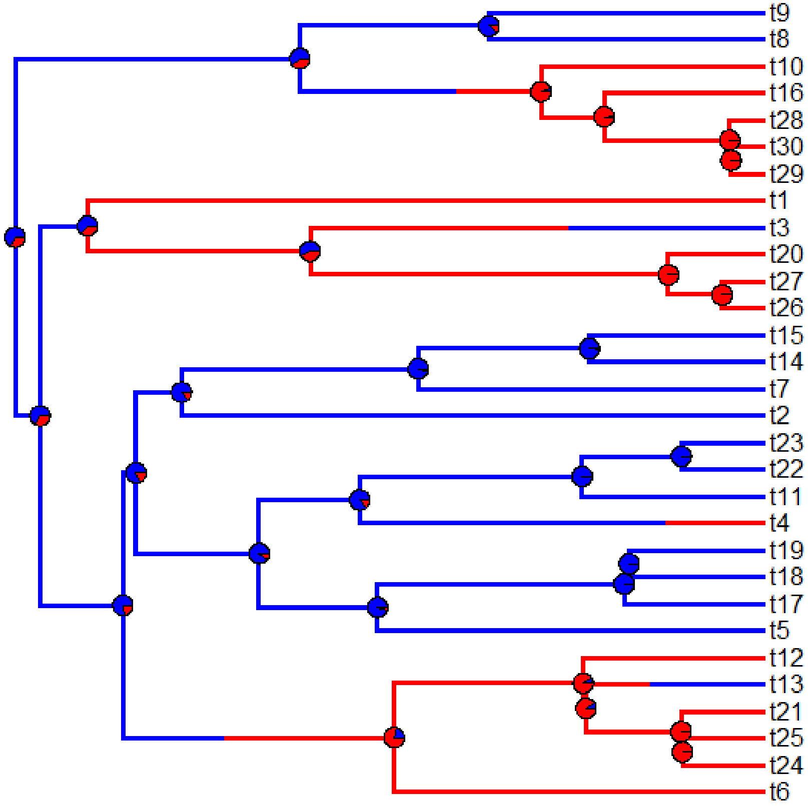


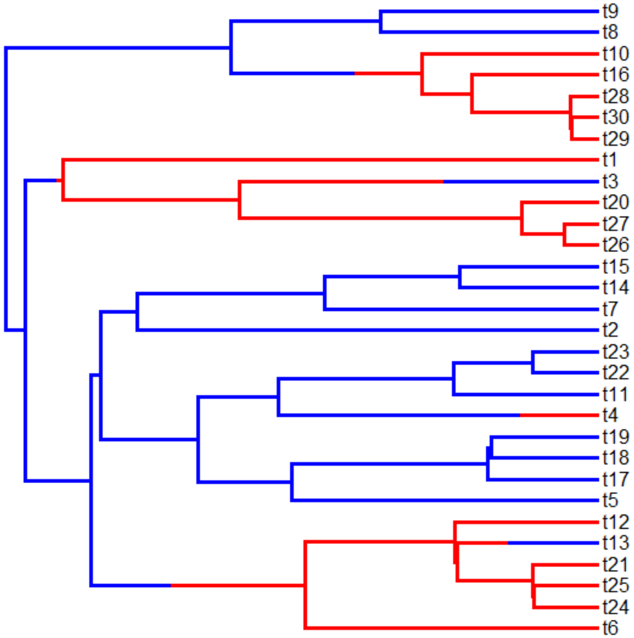
Figure. True binary character history with marginal ancestral reconstructions (empirical Bayes posterior probabilities).

Joint reconstruction

- *Joint reconstruction* is finding the set of states at all internal nodes that maximize the likelihood.
- This is ***not*** (necessarily) equivalent to picking the state at each node with the highest probability.
- We can find the *single* character history with the highest likelihood – but this is just one sample from the distribution. It happens to be the most likely, but it doesn't contain any information about uncertainty.
- One option is to *sample* node states and character histories from their joint (empirical or heirarchical) Bayesian posterior distribution. This is called *stochastic character mapping*.

Stochastic character mapping

- Stochastic character mapping is a procedure whereby we sample character histories in direct proportion to their posterior probability under a model.
- This is accomplished by first sampling a transition matrix ***Q*** (from its posterior probability distribution), then sampling a set of ancestral states at the nodes of the tree from their joint conditional probability distribution given ***Q***. Finally, we simulate character histories along all the edges of the tree conditioned on ***Q*** and our sampled node states.



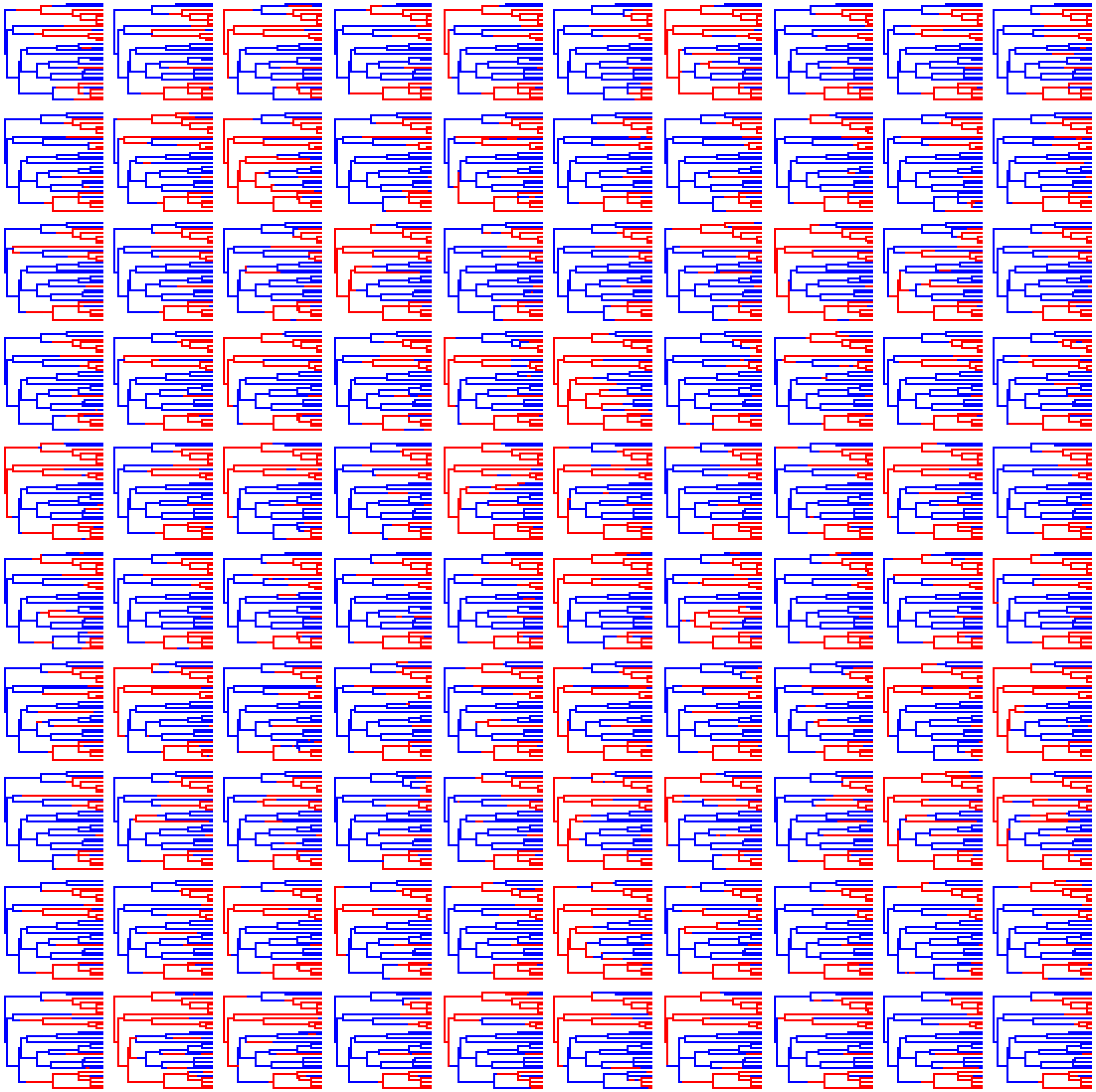


Figure. True history (above) & sample of stochastic character maps from the empirical Bayes posterior distribution (right).

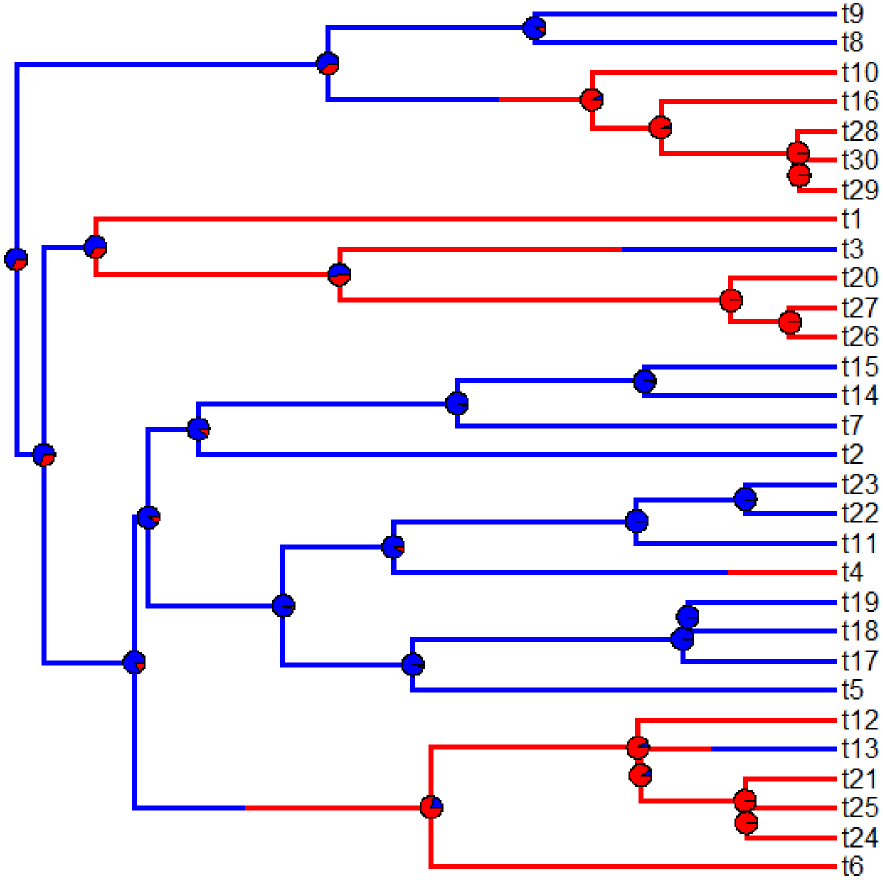


Figure. True history with posterior probabilities from stochastic mapping.

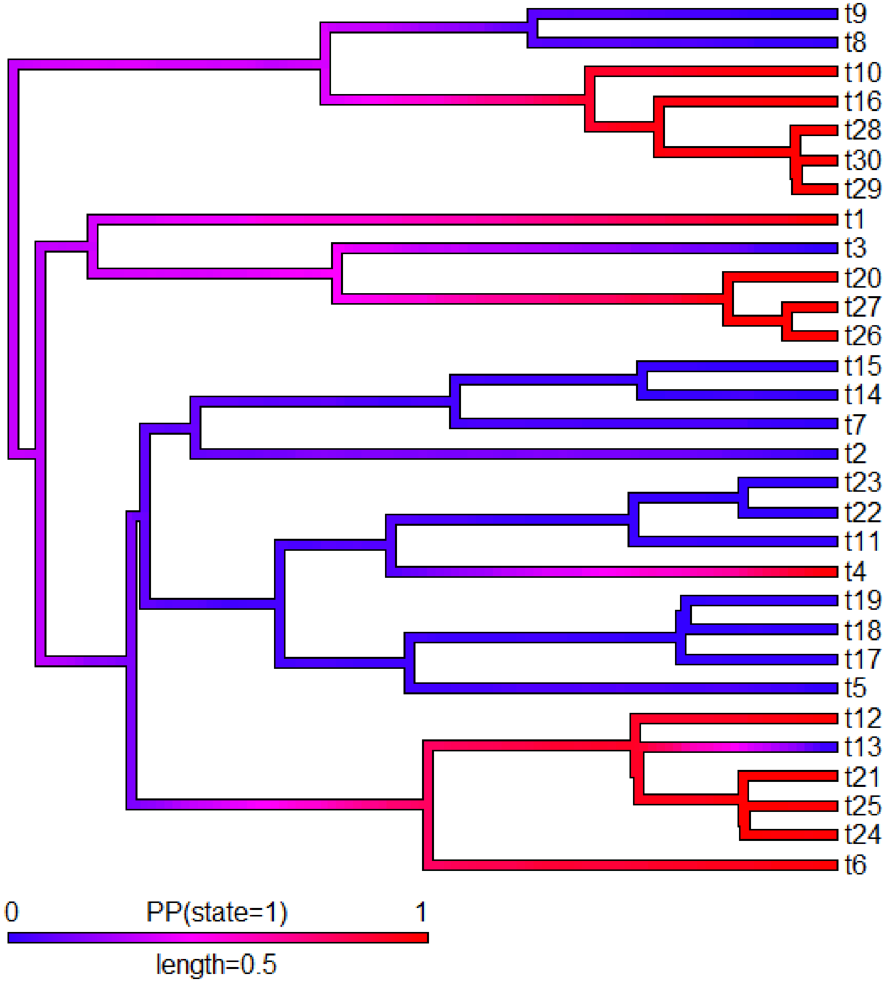
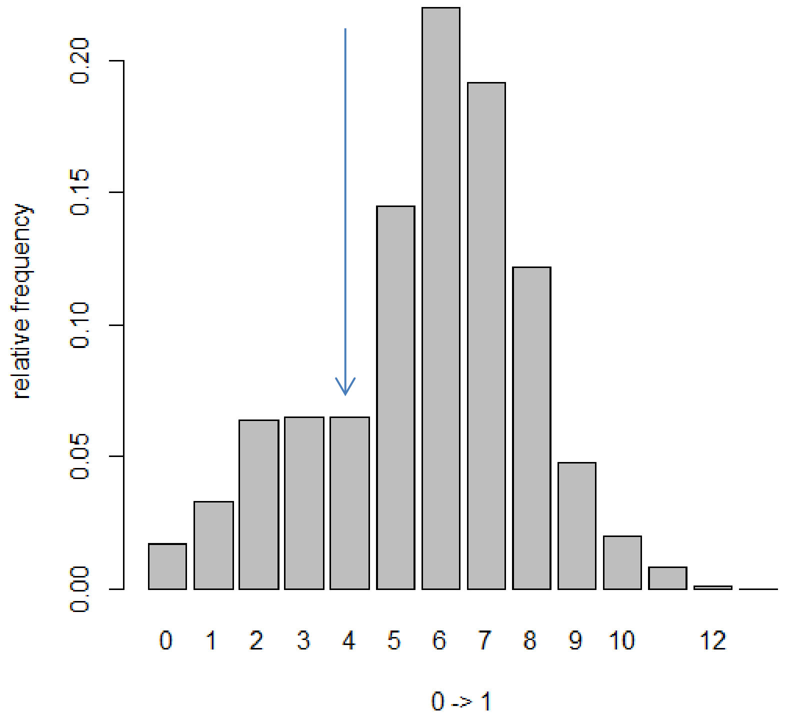
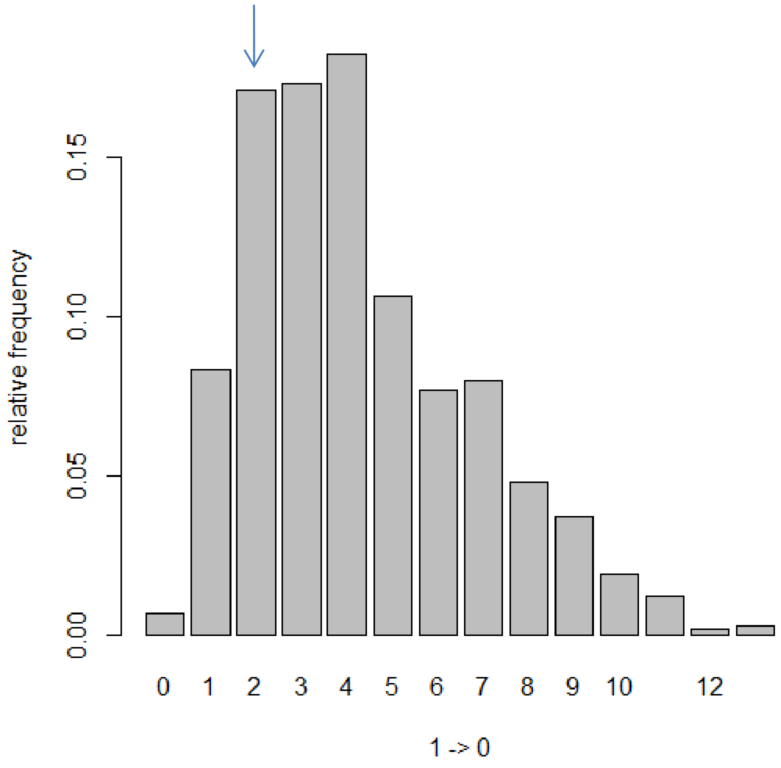


Figure. Posterior density map from stochastic mapping.

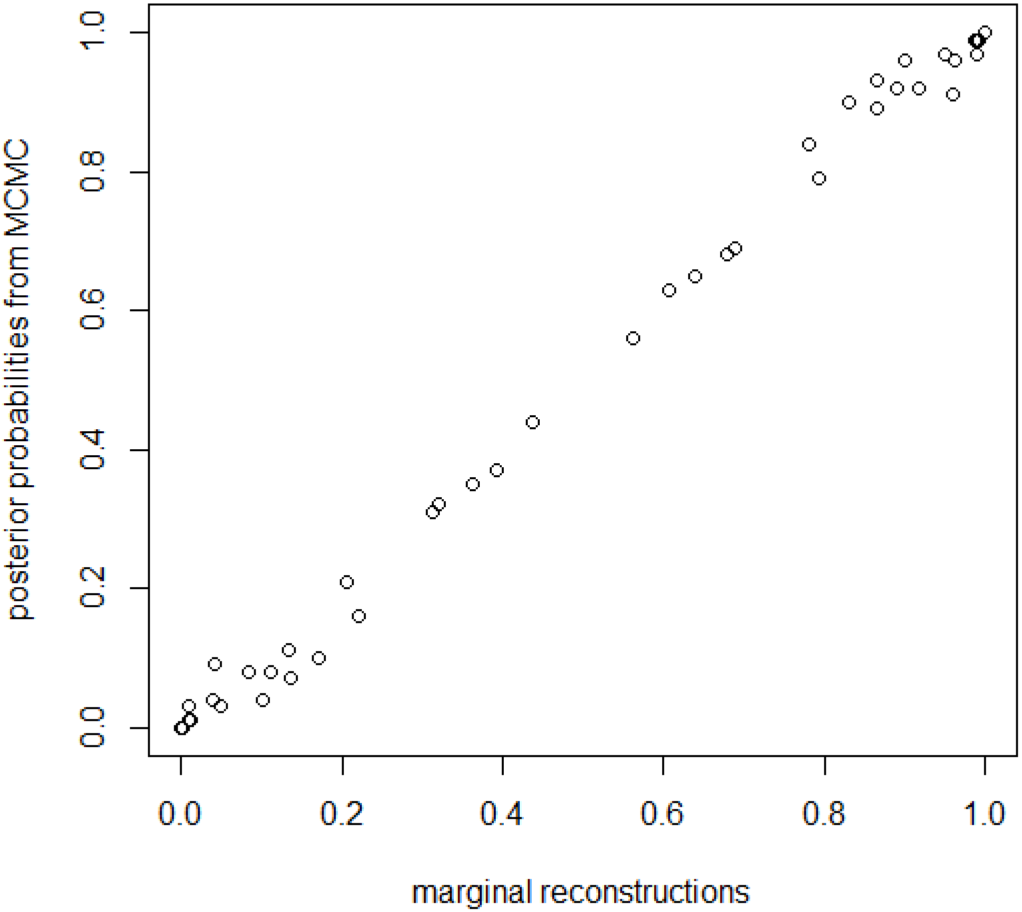
The number of changes on the tree

- We can obtain a probability distribution on the *number* of changes of each type on the tree.





Marginal vs. joint reconstruction



Priors

- In both marginal & joint reconstruction, we need to specify (or implicitly assume) a prior probability distribution for the global root, π_0.
- There is some debate over what constitutes the *best* prior distribution.
- Possibilities include: a flat prior, the stationary distribution given the fitted or sampled transition matrix ***Q***, and the empirical distribution at the tips of the tree.
- This decision can theoretically play a large role in influencing the inferred ancestral character values in the tree.

What about parsimony?

- It turns out that we get the (or a) parsimony reconstruction of our character on the tree from stochastic mapping *if* we put a very strong prior on ***Q*** to be small.
- This suggests that parsimony implicitly assumes that ***Q*** is very small – even if contrary evidence exist in our data suggesting ***Q*** is large.
- This means that the parsimony reconstruction will only accurately reflect the evolutionary process for our character when ***Q*** is very small.

What about parsimony?

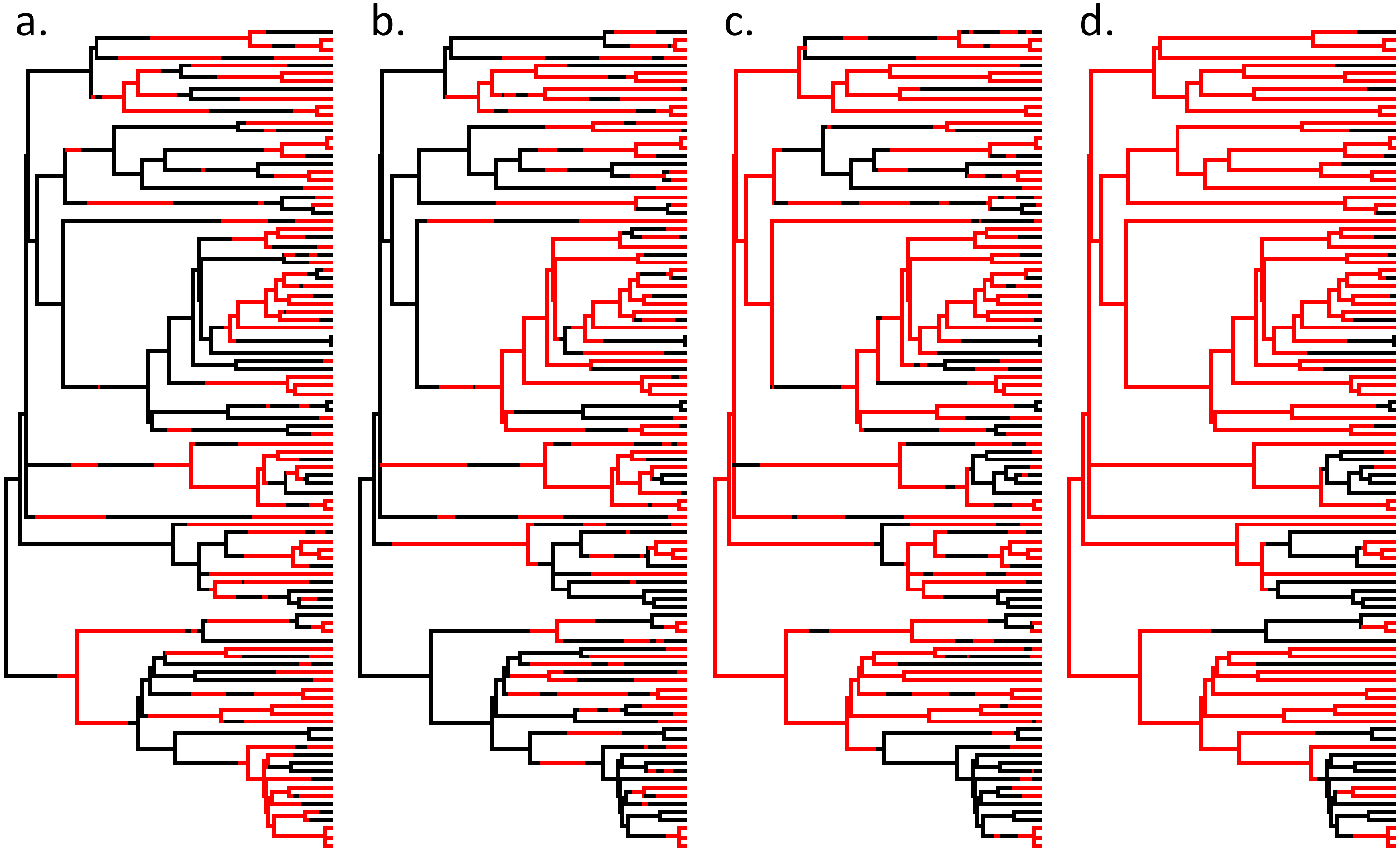


Figure. a) True history. b) Sampled history using empirical ***Q***. c) Sampled history using true ***Q***. d) Sampled history with a strong prior density on ***Q*** to be ***Q*** x 10^{-3}.

Conclusions from ancestral character reconstruction of discrete characters

- Marginal ancestral state reconstruction finds the MLE at a node (empirical Bayes posterior probabilities) integrating over all other nodes.
- Joint ancestral state reconstruction finds the set of states at nodes that maximize the likelihood. This need not be the set of states with the highest empirical Bayesian posterior probabilities.
- We can use stochastic mapping to sample from the joint posterior probability distribution of node states & changes along edges.
- Parsimony reconstruction is akin to assuming that the transition rates between states are very low – sometimes much lower than empirical estimates of those rates.