Ancestral state reconstruction II: Discrete characters

Liam J. Revell

15 December 2016

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

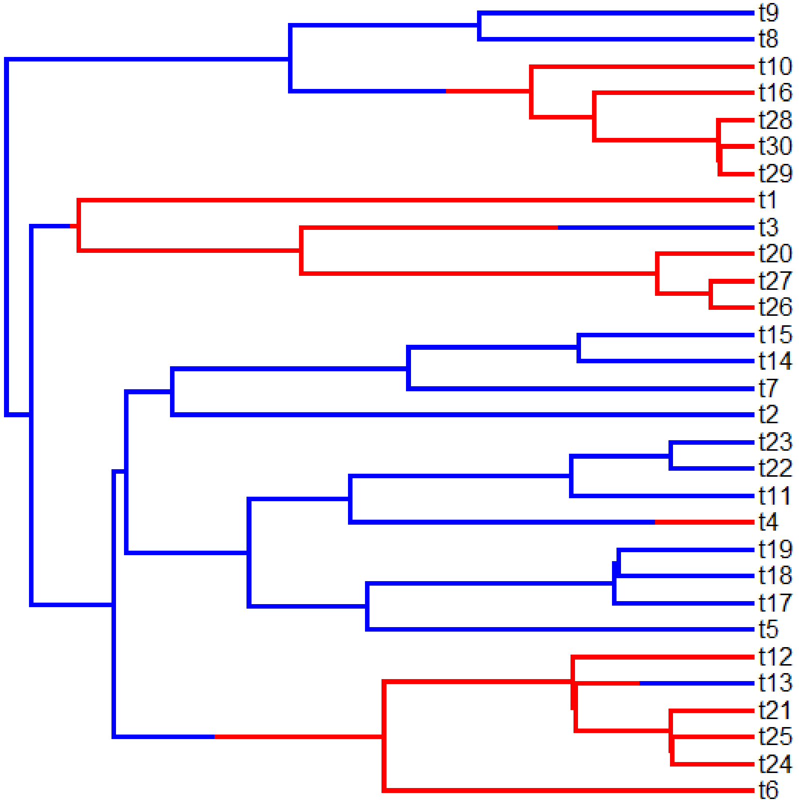


Figure. True history.

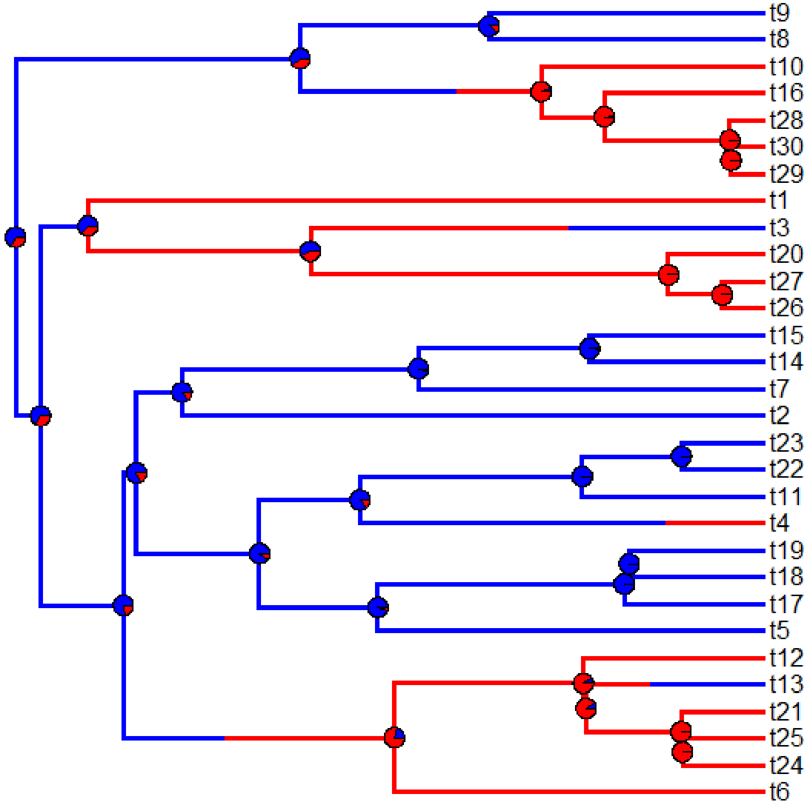


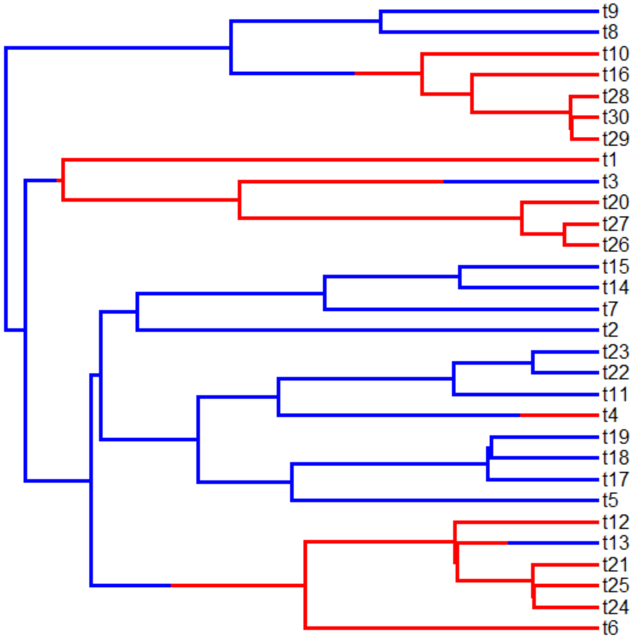
Figure. True binary character history with marginal ancestral reconstructions (empirical Bayes posterior probabilities).

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.



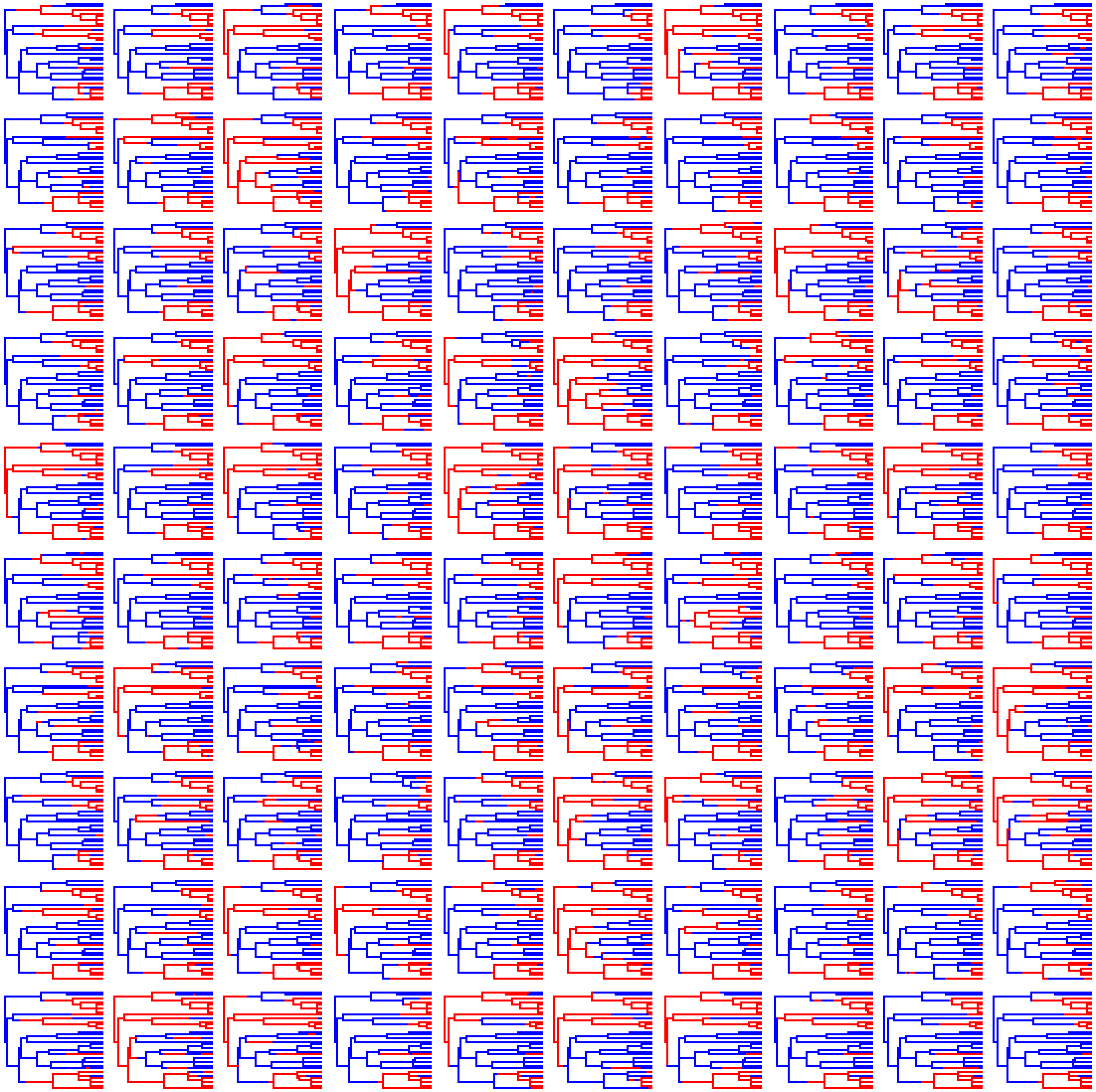


Figure. True history (above) & sample of stochastic character maps from the empirical Bayes posterior distribution (right).

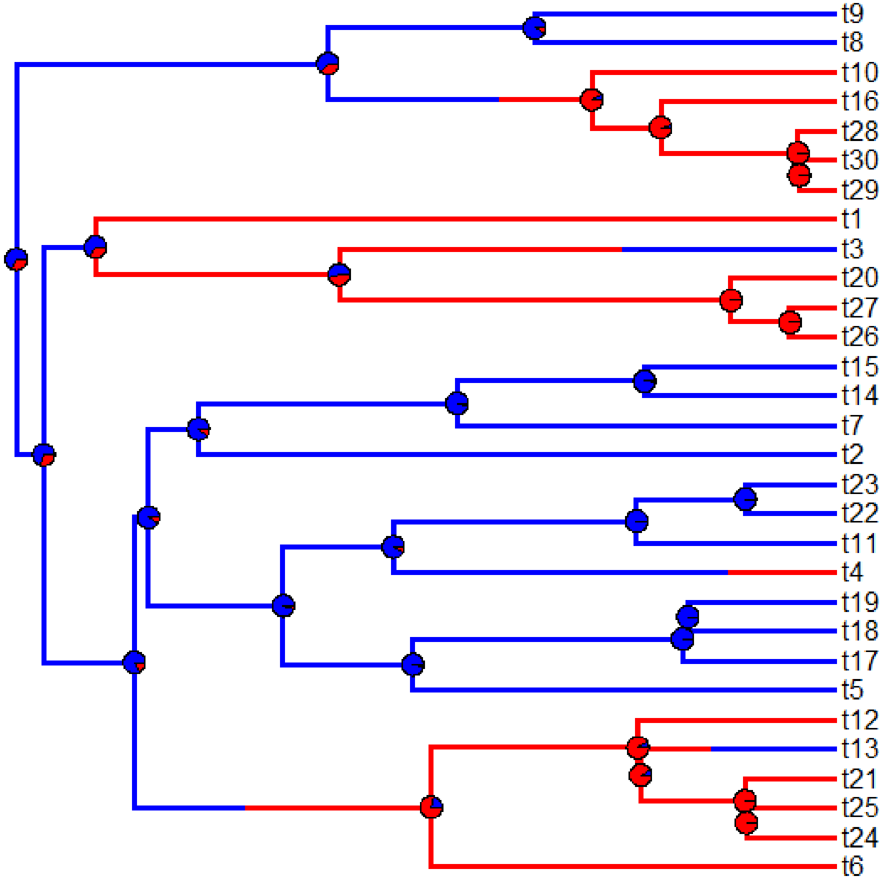


Figure. True history with posterior probabilities from stochastic mapping.

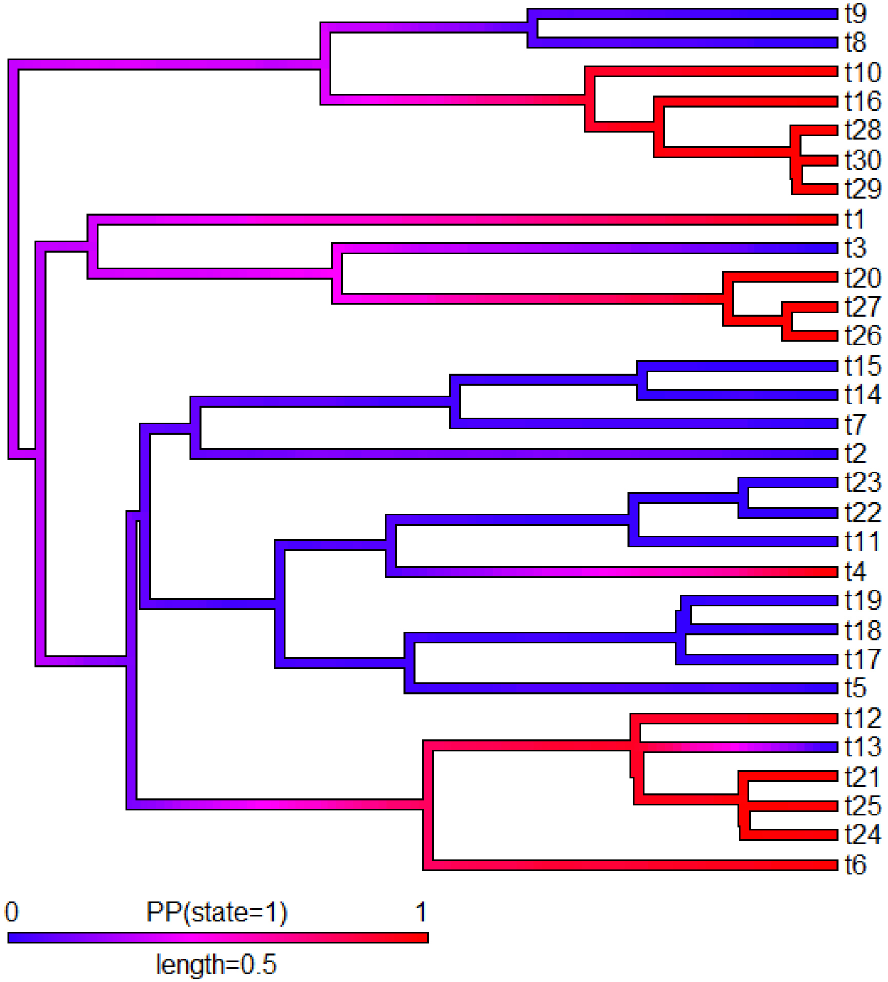
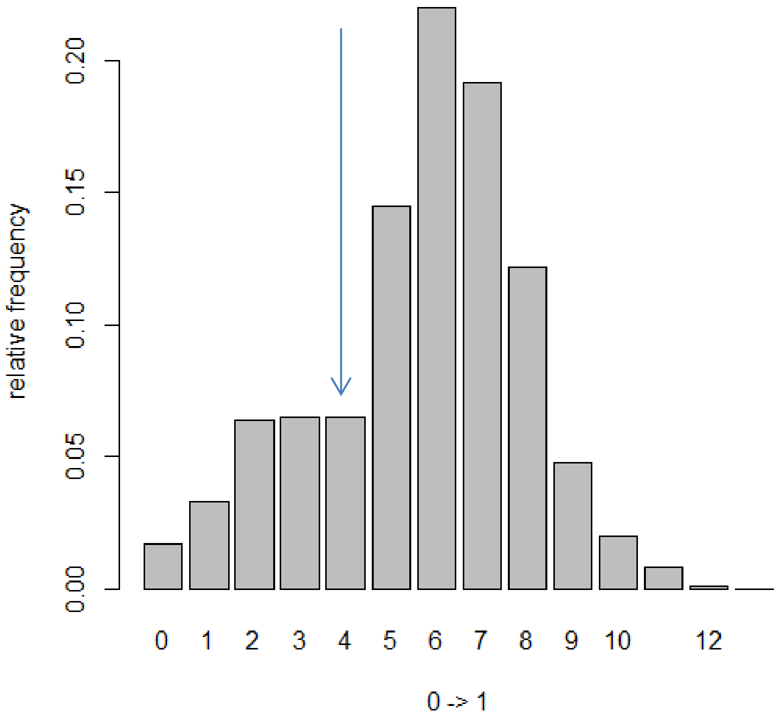
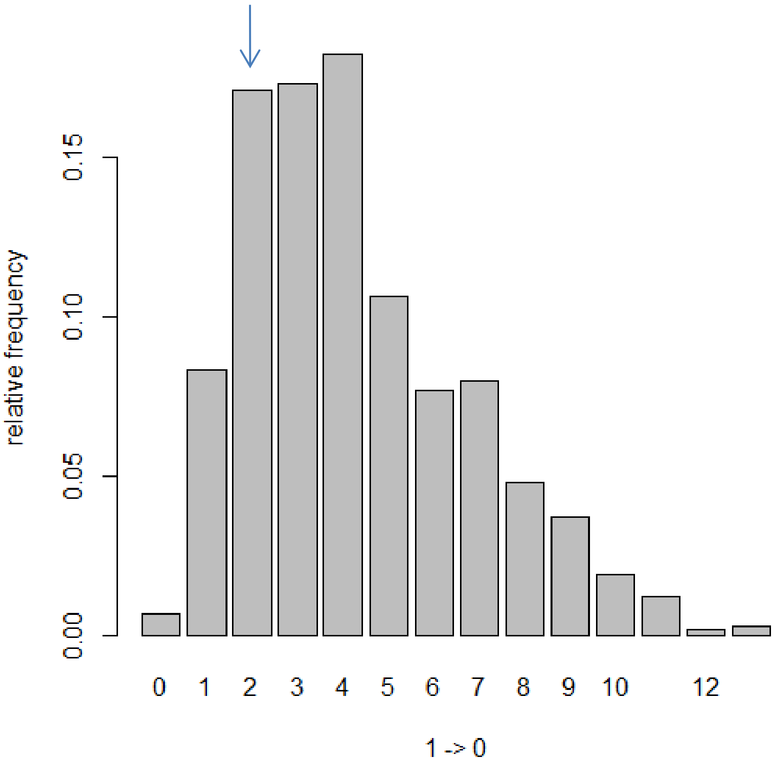


Figure. Posterior density map from stochastic mapping.

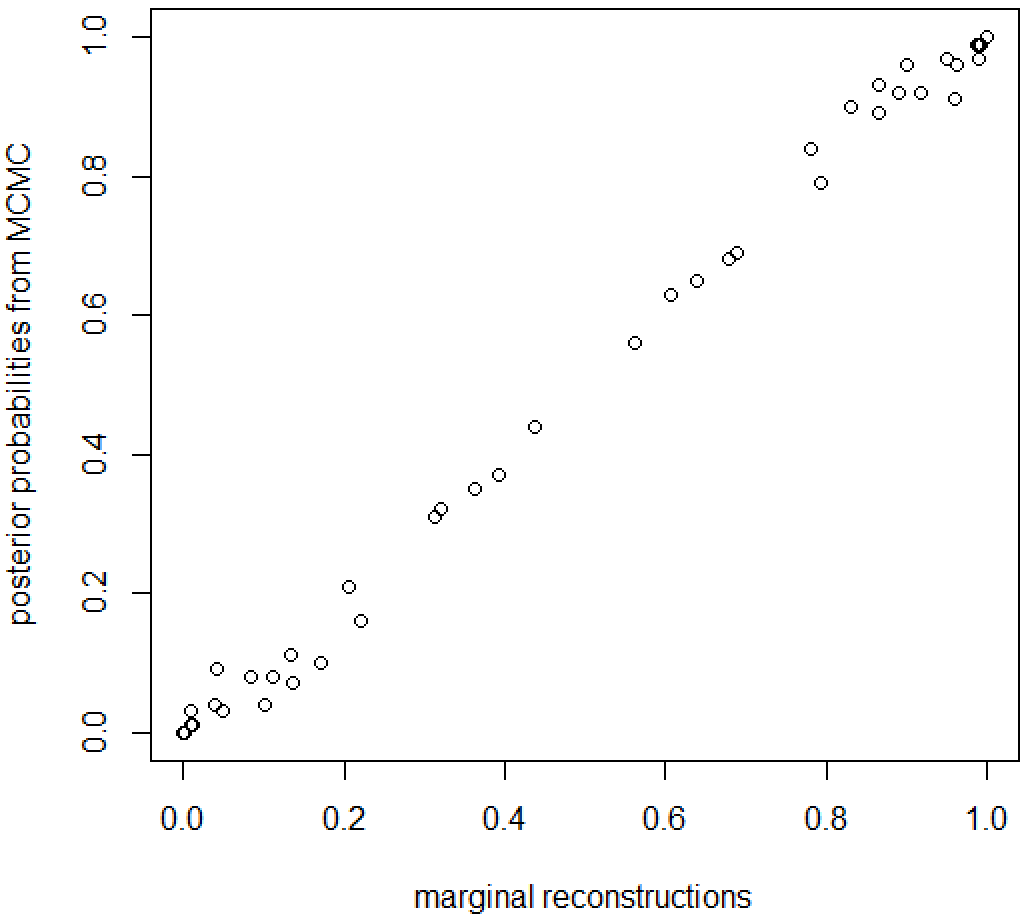
The number of changes on the tree

- 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 $\boldsymbol{Q}$, 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?

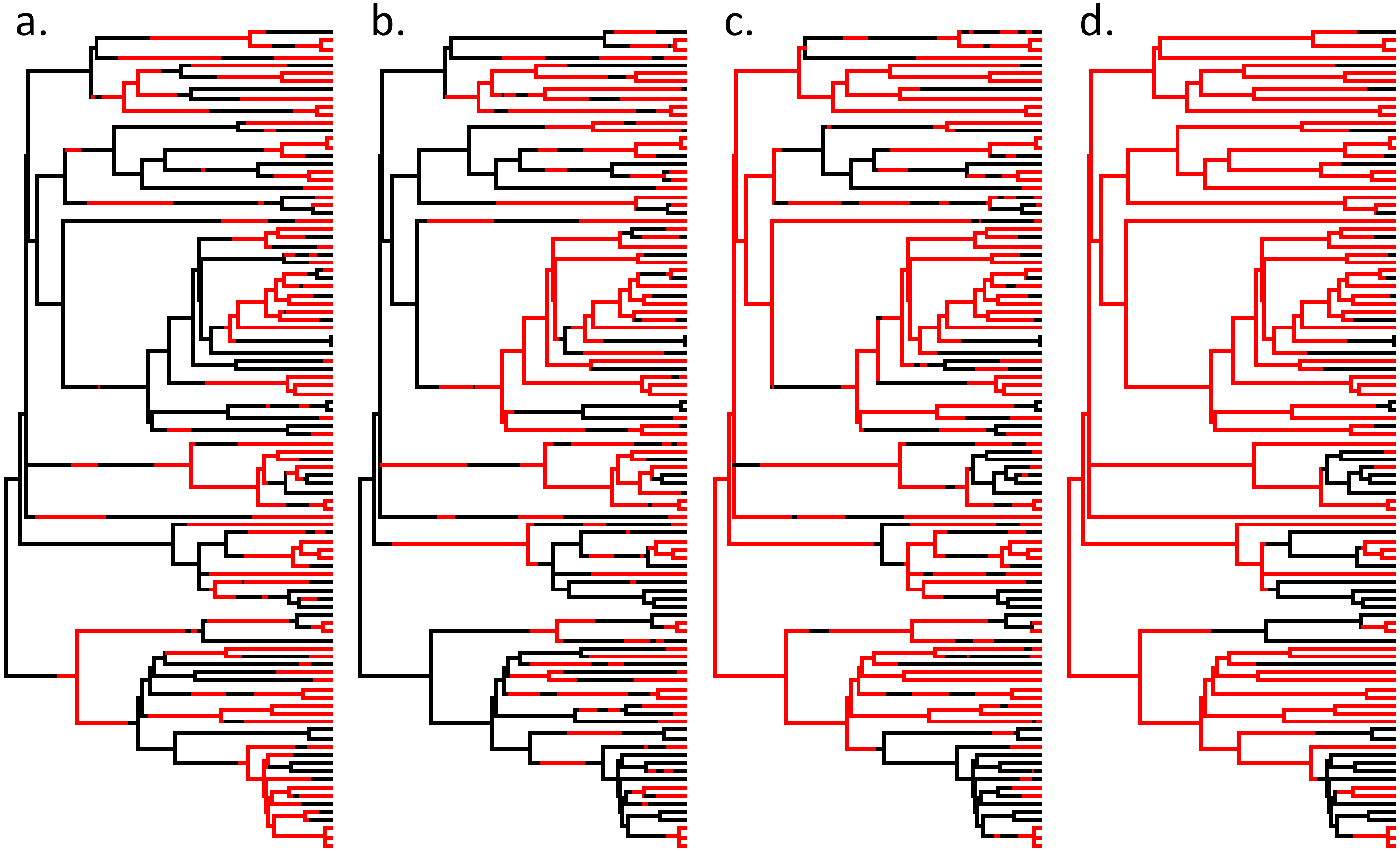


Figure. a) True history. b) Sampled history using empirical ***Q***. c) Sampled history using true ***Q***. d) Sampled history with a strong prior density on ***Q*** to be ***Q*** x 10^{-3}.

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.