

# ANCESTRAL RECONSTRUCTION: COMPARING MAJORITY RULE WITH PARSIMONY

HEYANG LI (THOMAS)

ABSTRACT. Various phylogenetic problems (such as DNA, protein and genome reconstruction) involve estimating the ancestral states at the root of a tree from data at the tips (leaves). A better understanding of various methods of ancestral reconstruction would help us determine their reliability and credibility. This project investigates the upper and lower bound for the correct probability of reconstruction for Yule trees, under different methods of reconstruction. This project also focuses on the question of whether of majority rule or parsimony perform better. In addition, results such as monotonicity of parsimony, and the effect of changing the state of one or two leaves are presented.

## 1. INTRODUCTION AND DEFINITIONS

Ancestral reconstruction has been used to infer various ecological and biological traits associated with ancestral root in a tree ([3] and [4]). However, different methods require different computational complexity and different amounts of information, and also have different degrees of accuracy of reconstruction.

**1.1. Yule and Mutation Processes.** In 1925, George Udny Yule proposed a simple model of speciation. Yule's pure birth model ([8]) is a random process to generate a subclass of phylogenetic trees. The process can start with one or two lineages both of length 0 at time  $t = 0$ . Then each lineage speciates independently of each other with a constant diversification rate  $\lambda$ . The time of any given extant lineage to speciate has an exponential distribution with mean  $\frac{1}{\lambda}$ . In this model, we have three parameters:  $t$  is the depth of the tree,  $\lambda$  is the speciation rate and  $N_t$  is the number of leaves the tree has at time  $t$ .

We now superimpose on this random process for generating a tree a second stochastic (mutation) process. This process is parameterized by the tree and its branch lengths, and its state space is a set of  $k$  discrete states which undergo changes along the edges of the tree according to a finite-state continuous-time Markov process.

The root of the Yule tree starts with an initial state. Each of the  $k$  states is given a different label. Along any edge (from the root towards a leaf), a stochastic process occurs for interstate changes. The current state changes (along an edge) according to

---

*Key words and phrases.* Yule tree, pure-birth model, ancestral reconstruction, majority rule, parsimony.

a symmetric continuous-time Markov process with mutation rate  $\mu$ . This project investigates some reconstruction methods used to estimate the original root state from the states observed at the leaves.

**1.2. Reconstruction Methods.** There are three commonly-used methods to predict the original root state: *majority rule*, *parsimony* and *maximum likelihood*. Majority rule estimates the original root state to be the state that occurs the most frequently amongst the leaves (ties are broken randomly). Parsimony estimates the root state(s) that requires the minimal number of state changes on the edges of the tree to generate the observed leaf states. Maximum likelihood finds the root state that is the most likely to generate the observed leaf states.

Notice that these three methods require increasing information concerning the tree. In particular, majority rule just requires the number of leaves in each state (or just which state occurs the most frequently amongst the leaves); Parsimony requires the tree topology and the leaf states; and maximum likelihood requires the leaf states, tree topology and its edge lengths.

### 1.3. Main Results of this project.

- The monotonicity of parsimony estimation probability with respect to time;
- Upper and lower bounds on the probability of correct root state estimation for various methods;
- Comparing the probability of correct rooted state estimation between different reconstruction methods;
- The  $r$ -state parsimony dynamical equations under a Yule model;
- Investigation into the stability of parsimony reconstruction when one or two leaf states are changed.

**1.4. Definitions.** Let  $\mathcal{T} = \mathcal{T}_t$  be the random variable consisting of the Yule tree (tree topology and branch lengths) at time  $t$  with parameters  $\lambda$  (speciation rate) and  $\mu$  (mutation rate).

Define  $p_{\text{PAR}}(\mathcal{T}_t)$  to be the probability that parsimony would estimate the correct root state with the observed leaf states. Let

$$p_{\text{PAR}}^{\text{Yule}}(t) := \mathbb{E}_{\mathcal{T}}[p_{\text{PAR}}(\mathcal{T}_t)],$$

where  $\mathbb{E}_{\mathcal{T}}$  refers to expectation over all Yule trees  $\mathcal{T}_t$  grown for time  $t$ .

Similarly, define  $p_{\text{MR}}(\mathcal{T}_t)$  and  $p_{\text{ML}}(\mathcal{T}_t)$  to be the probability that majority rule and maximum likelihood would estimate the correct root state with the observed leaf states. Let

$$p_{\text{MR}}^{\text{Yule}}(t) := \mathbb{E}_{\mathcal{T}}[p_{\text{MR}}(\mathcal{T}_t)] \quad \text{and} \quad p_{\text{ML}}^{\text{Yule}}(t) := \mathbb{E}_{\mathcal{T}}[p_{\text{ML}}(\mathcal{T}_t)]$$

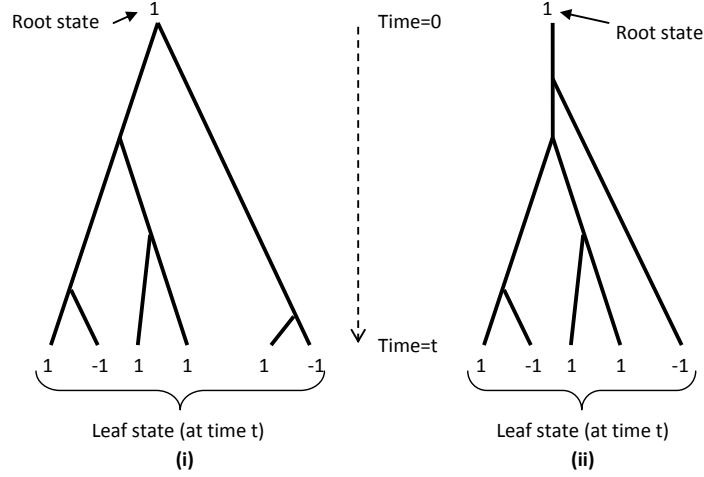


FIGURE 1. Six-leaf example of the 2-state model (with 1 and 2 initial lineages)

to be the expected probability that majority rule and maximum likelihood would estimate the correct root state at time  $t$  for a Yule tree.

Also, define  $p_*^{\text{Yule}} := \lim_{t \rightarrow \infty} p_*^{\text{Yule}}(t)$ , for  $*$  =  $MR$ ,  $ML$ , and  $PAR$ .

## 2. THE MONOTONIC PROPERTY OF $p_{\text{PAR}}^{\text{Yule}}(t)$

For the one initial lineage start. From Section 2 of [2], the authors suggested the following conjecture:

**Conjecture 2.0.1.** *Let  $\rho = \frac{\mu}{\lambda}$ . If  $\rho < \frac{1}{6}$ , then  $p_{\text{PAR}}^{\text{Yule}}(t)$  decreases monotonically with respect to time  $t$ .*

In this section, we establish that this conjecture is true. Following [2], let  $S_t$  be the expected probability of making a correct estimation,  $D_t$  be the expected probability of making a wrong estimation, and  $E_t$  be the expected probability of a tie. The conjecture is equivalent to showing that:

**Theorem 2.0.2.** *Let  $\rho = \frac{\mu}{\lambda}$ . If  $\rho < \frac{1}{6}$ , then*

$$\frac{dp_{\text{PAR}}^{\text{Yule}}(t)}{dt} \leq 0,$$

for all time  $t > 0$ .

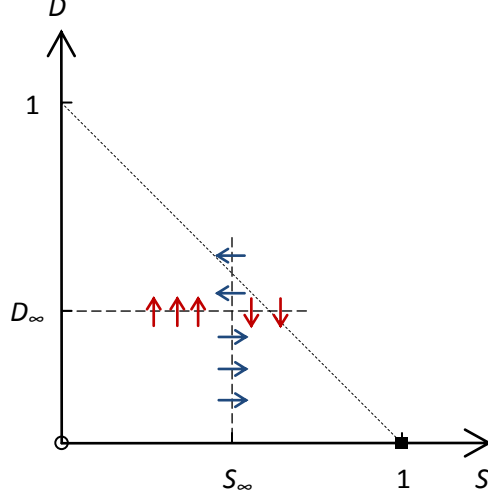


FIGURE 2. Partial vector fields of the system

*Proof.* Since the probability of a wrong estimation is  $1 - p_{\text{PAR}}^{\text{Yule}}(t)$ , we have  $\frac{dp_{\text{PAR}}^{\text{Yule}}(t)}{dt} \leq 0$  if

$$\frac{dp_{\text{PAR}}^{\text{Yule}}(t)}{dt} - \frac{d(1 - p_{\text{PAR}}^{\text{Yule}}(t))}{dt} \leq 0, \text{ for all } t.$$

This holds if

$$(2.0.1) \quad \frac{d(S_t + \frac{1}{2}E_t)}{dt} - \frac{d(D_t + \frac{1}{2}E_t)}{dt} \leq 0.$$

Eqn. (2.0.1) holds provided that

$$\frac{dS_t}{dt} - \frac{dD_t}{dt} \leq 0,$$

which, in turn, holds if  $\frac{dS_t}{dt} \leq 0$  and  $\frac{dD_t}{dt} \geq 0$ .

From here we will consider the 2-D vector field of  $S_t$  and  $D_t$  from the 2-D autonomous differential equation (with a linear time transformation of  $u = \lambda t$ ) obtained from [2] as:

$$\frac{dS}{du} = (1 - \rho)S + (\rho - 2S)D - S^2, \quad \text{and} \quad \frac{dD}{du} = (1 - \rho)D + (\rho - 2D)S - D^2.$$

From [2] we know that  $S_t$  and  $D_t$  converges to  $S_\infty := \frac{1-2\rho+\sqrt{(1-6\rho)(1-2\rho)}}{2}$  and  $D_\infty := \frac{1-2\rho-\sqrt{(1-6\rho)(1-2\rho)}}{2}$  respectively as  $t \rightarrow \infty$ . Now, we firstly show that  $S_t$  will never be less than and  $D_t$  will never be more than  $D_\infty$ . At  $t = 0$ , the system starts

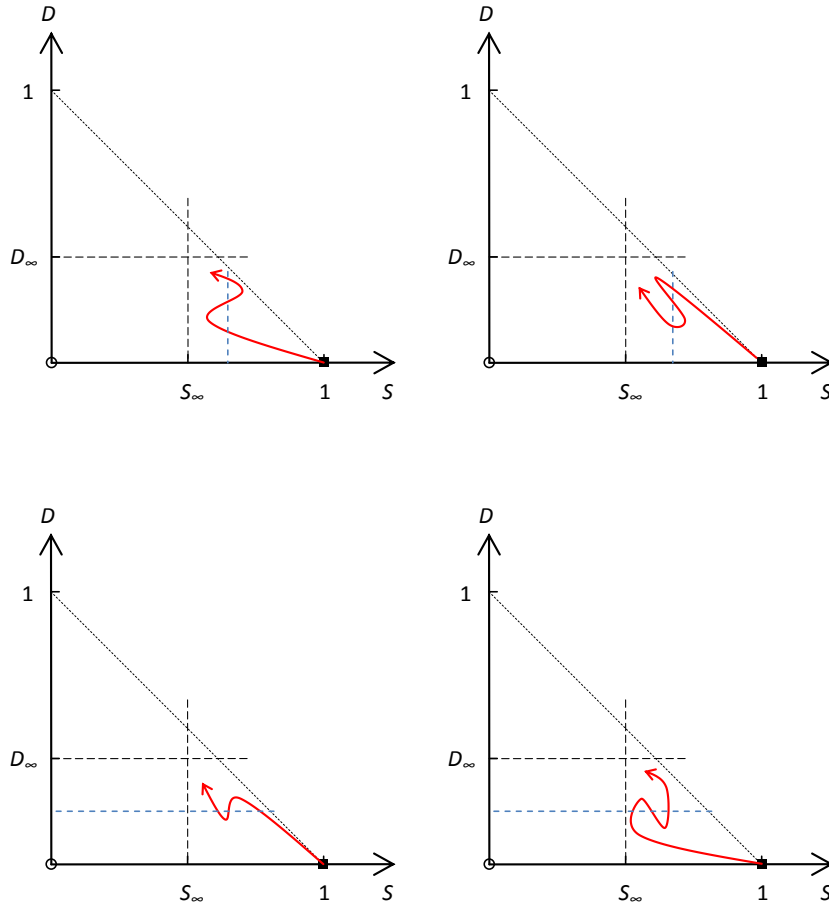


FIGURE 3. The four possible types of path that increase  $S$  or decrease  $D$ .

at  $S_0 = 1$  and  $D_0 = 0$ . Looking at the vector field (see Fig. 2) on the horizontal and vertical line that pass through the limit point  $(S_\infty, D_\infty)$ , assuming  $D_\infty \geq \rho/2$  and  $S_\infty \geq \rho/2$  (this is implied by the limits of  $S$  and  $D$  calculated in [2]).

From Fig. 2, the red arrow is the vertical vector field direction (partial in  $D$ ) and the blue arrow is the horizontal vector field direction (partial in  $S$ ). Fig. 2 shows that by starting at  $(S, D) = (1, 0)$ , the orbit which goes to  $(S_\infty, D_\infty)$  could never cross the horizontal or the vertical dotted line. Hence the only way for  $S$  to decrease or for  $D$

to increase is if one of the following orbits occurs (as in Fig. 3). For each of the orbit, it will cross a horizontal line or a vertical line at least 3 times.

Since  $\frac{dS}{du}$  and  $\frac{dD}{du}$  are linear in  $D$  and  $S$  respective, the orbit could not cross a vertical or horizontal line three times or more. This can be proved by contradiction.

Assume the orbit crosses a vertical or horizontal line three times (or more), then looking at the partial vector field perpendicular to that vertical or horizontal line would need to change in direction (at least) twice. This would imply that  $\frac{dS}{du}$  is not linear in  $D$  and  $\frac{dD}{du}$  is not linear in  $S$ . This forms a contradiction.

Note, Fig. 3 is the only 4 possible types of path that increase  $S$  or decrease  $D$ . Hence  $S$  and  $D$  can only be decreasing and increasing respectively over time. Therefore,  $p_{\text{PAR}}^{\text{Yule}}$  is indeed monotonically decreasing over time.  $\square$

For the two lineages start, let  $\Delta$  be the corresponding quantity in the two lineages state (Fig. 1 (i)). Let  $S_t^\Delta$  be the expected probability of making a correct estimation,  $D_t^\Delta$  be the expected probability of making a wrong estimation, and  $E_t^\Delta$  be the expected probability of a tie. Therefore:  $p_{\text{PAR}}^{\text{Yule}\Delta}(t) = S_t^\Delta + \frac{1}{2}E_t^\Delta$ . We will show:

**Theorem 2.0.3.** *Let  $\rho = \frac{\mu}{\lambda}$ . If  $\rho < \frac{1}{6}$ , then:*

(1)

$$\frac{dp_{\text{PAR}}^{\text{Yule}\Delta}(t)}{dt} \leq 0,$$

for all  $t > 0$ .

(2)

$$p_{\text{PAR}}^{\text{Yule}\Delta}(t) \geq \frac{1}{2} \left( 1 + (1 + 2\rho)\sqrt{(1 - 6\rho)(1 - 2\rho)} \right),$$

for all  $t > 0$ .

*Proof.* For (1):

$$\frac{dp_{\text{PAR}}^{\text{Yule}\Delta}(t)}{dt} \leq 0.$$

The above equation holds, if:

$$\frac{d\left(S_t^\Delta + \frac{1}{2}E_t^\Delta\right)}{dt} \leq 0,$$

or equivalently if:

$$\frac{d\left(S_t^\Delta - D_t^\Delta\right)}{dt} \leq 0.$$

The previous equation holds, if:

$$\frac{d(S_t^2 + 2S_tE_t - (D_t^2 + 2D_tE_t))}{dt} \leq 0,$$

and which holds, if:

$$\frac{d((S_t - D_t)(2 - S_t - D_t))}{dt} \leq 0,$$

or equivalently, if:

$$\frac{d(2S_t - 2D_t - S_t^2 + D_t^2)}{dt} \leq 0.$$

Which in turn holds, if:

$$2(1 - S_t)\frac{dS_t}{dt} - 2(1 - D_t)\frac{dD_t}{dt} \leq 0.$$

The last statement is true, as both  $1 - D_t$  and  $1 - S_t$  is greater or equal to 0, and  $\frac{dS_t}{dt} \leq 0$  and  $\frac{dD_t}{dt} \geq 0$  for  $\rho < \frac{1}{6}$ . Hence part (1) of the theorem is proved.

For (2):

Let

$$p_{\text{PAR}}^{\text{Yule } \Delta} = \lim_{t \rightarrow \infty} p_{\text{PAR}}^{\text{Yule } \Delta}(t),$$

and let  $S^\Delta$ ,  $D^\Delta$  and  $E^\Delta$  be the corresponding limits as  $t \rightarrow \infty$  for  $S^\Delta(t)$ ,  $D^\Delta(t)$  and  $E^\Delta(t)$ .

Then,  $p_{\text{PAR}}^{\text{Yule } \Delta} = S^\Delta + \frac{1}{2}E^\Delta = \frac{1}{2}(1 + (S^\Delta - D^\Delta))$ .

From the proof of part (1),  $S^\Delta - D^\Delta = (S_\infty - D_\infty)(1 + E_\infty)$ , and from [2],

$$(S_\infty - D_\infty)(1 + E_\infty) = \sqrt{(1 - 6\rho)(1 - 2\rho)}, \quad E_\infty = 2\rho.$$

Thus,

$$p_{\text{PAR}}^{\text{Yule } \Delta} = \frac{1}{2} \left( 1 + (1 + 2\rho)\sqrt{(1 - 6\rho)(1 - 2\rho)} \right).$$

Part (2) now follows since part (1) implies that  $p_{\text{PAR}}^{\text{Yule } \Delta}(t) \geq p_{\text{PAR}}^{\text{Yule } \Delta}$ , for all  $t > 0$ .  $\square$

### 3. 2-STATE MODEL ANALYSIS AND BOUNDS

For a 2-state model, it is convenient to let the two states be  $-1, 1$  (1 and  $-1$  are used rather than 0 and 1 because for majority rule we could just add up all the root states and based on the sign of the result to estimate the root state). Let

- $X_0$  be the random variable of the root state of a Yule tree;
- $X = (X_i)_{i=1, \dots, N_{\mathcal{T}}}$  be the random variable of the leaf states of the same Yule tree with root state  $X_0$ .

**Theorem 3.0.4.** *If  $4\mu > \lambda$ , then  $\mathbb{P}(\hat{X}_0 = X_0) \rightarrow \frac{1}{2}$  as  $t \rightarrow \infty$  for any estimator  $\hat{X}_0 = \hat{X}_0(X)$  of  $X$ .*

*Proof.* We apply Theorem 14.1 of [5]. In our context (2-state model, and constant mutation rate) this says that:

$$\mathbb{P}(\hat{X}_0 = X_0 | \mathcal{T}) \leq \frac{1}{2} + N_{\mathcal{T}} \exp(-4\mu t),$$

for any reconstruction method. Note that we are conditioning on the tree  $\mathcal{T}$  (and branch lengths).

Now, for *any* event  $E$ , we have:  $\mathbb{P}(E) = \mathbb{E}_{\mathcal{T}}[\mathbb{P}(E|\mathcal{T})]$ , and so, applying this to  $E = “\hat{X}_0 = X_0”$  gives us:

$$\mathbb{P}(\hat{X}_0 = X_0) \leq \frac{1}{2} + \exp((\lambda - 4\mu)t) \rightarrow \frac{1}{2} \text{ as } t \rightarrow \infty \text{ for any } \lambda < 4\mu.$$

□

**3.1. Majority Rule.** For majority rule, consider the random variable

$$Z_t = X_1 + X_2 + \cdots + X_{N_t} \text{ (where } N_t \sim \text{geo}(\lambda t)\text{)}.$$

Note that the  $X_i$ 's are identically distributed but not independent. Without loss of generality, let the root state be 1. If a realization of  $Z$  is positive, then there are more 1s than  $-1$ s in the leaf states, hence a correct estimation will be made by the majority rule. If  $Z$  is negative, a wrong estimation will be made. If  $Z$  is zero, there is only a 50 percent probability that  $Z$  will be the correct estimation (as a tie is broken uniformly at random).

Let us now examine the expectation and variance of  $Z_t$ .

$$(3.1.1) \quad \mathbb{E}[Z_t] = \mathbb{E}_{\mathcal{T}} \left[ \mathbb{E}_{\mu} \sum_{i=1}^{N_{\mathcal{T}}} [X_i | \mathcal{T}] \right] = \mathbb{E}_{\mathcal{T}} [N_{\mathcal{T}} \mathbb{E}_{\mu} [X_i | \mathcal{T}]] = \mathbb{E}_{\mathcal{T}} [N_{\mathcal{T}} e^{-2\mu t}] = e^{\lambda t} e^{-2\mu t}$$

(recall  $\mathcal{T}$  refers to the topology and edge length of the Yule tree).

For the variance, we obtain:

$$(3.1.2) \quad \text{Var}(Z_t) = \text{Var}_{\mathcal{T}}(\mathbb{E}[Z_t | \mathcal{T}]) + \mathbb{E}_{\mathcal{T}}[\text{Var}(Z_t | \mathcal{T})].$$

The first term of the above equation  $\text{Var}_{\mathcal{T}}(\mathbb{E}[Z_t | \mathcal{T}])$ , can be expressed as:

$$(3.1.3) \quad \text{Var}_{\mathcal{T}}(\mathbb{E}[Z_t | \mathcal{T}]) = \text{Var}_{\mathcal{T}}(N_{\mathcal{T}} e^{-2\mu t}) = e^{(2\lambda - 4\mu)t} (1 - e^{-\lambda t}).$$



Because  $\mathbb{E}_{\mathcal{T}}[\text{Var}(Z_t|\mathcal{T})] \geq 0$ , we conclude that:

$$(3.1.4) \quad \text{Var}(Z_t) = e^{(2\lambda-4\mu)t} (1 - e^{-\lambda t}) + \mathbb{E}_{\mathcal{T}}[\text{Var}(Z_t|\mathcal{T})] \geq e^{(2\lambda-4\mu)t} (1 - e^{-\lambda t}).$$

Therefore, for any  $\epsilon > 0$ , there exist  $t_0$  such that:

$$(3.1.5) \quad \frac{\text{Var}(Z_t)}{\mathbb{E}[Z_t]^2} > 1 - \epsilon \text{ for any } t > t_0 \text{ for all } \lambda > 4\mu.$$

The variance of  $Z_t$  occurs from the value of  $N_{\mathcal{T}}$  and the mutation process.

In the light of this result, let's refocus our attention on the random variable  $Y_t = \frac{Z_t}{N_{\mathcal{T}}}$ . Let's consider the  $\mathbb{E}[Y_t]$  and  $\text{Var}(Y_t)$ . Note that  $Y_t$  also can't be asymptotically normal, as  $-1 \leq Y_t \leq 1$  and yet  $\sigma^2/\mu^2 \not\rightarrow 0$ .

**Theorem 3.1.1.**

$$(3.1.6) \quad p_{\text{MR}}^{\text{Yule}} \geq 2 - \frac{1}{(1 - 4\mu/\lambda)^2}.$$

*Proof.* For  $\mathbb{E}[Y_t]$ , we obtain:

$$\mathbb{E}[Y_t] = \mathbb{E}_{\mathcal{T}} \left[ \mathbb{E}_{\mu} \left[ \frac{Z_t}{N_{\mathcal{T}}} | \mathcal{T} \right] \right] = \mathbb{E}_{\mathcal{T}} [e^{-2\mu t}] = e^{-2\mu t}.$$

For  $\text{Var}(Y_t)$ , we obtain:

$$\text{Var}(Y_t) = \text{Var}_{\mathcal{T}}(\mathbb{E}_{\mu}[Y_t|\mathcal{T}]) + \mathbb{E}_{\mu}[\text{Var}_{\mathcal{T}}(Y_t|\mathcal{T})].$$

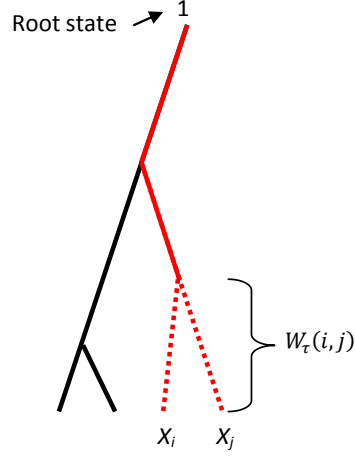
Firstly we find that:

$$\text{Var}_{\mathcal{T}}(\mathbb{E}_{\mu}[Y_t|\mathcal{T}]) = \text{Var}_{\mathcal{T}}(e^{-2\mu t}) = 0.$$

Secondly we find an upper bound for  $\mathbb{E}_{\mathcal{T}}[\text{Var}_{\mu}(Y_t|\mathcal{T})]$ .

Let  $\alpha = 4\mu$  and  $W_{\mathcal{T}}(i, j)$  be the length from the first split to time  $t$  for leaf  $X_i$  and  $X_j$  (in Fig. 4). We could derive the following:

$$\mathbb{E}_{\mathcal{T}}[\text{Var}_{\mu}(Y_t|\mathcal{T})] = \mathbb{E}_{\mathcal{T}} \left[ \frac{1}{N_{\mathcal{T}}^2} \sum_{i,j=1}^{N_{\mathcal{T}}} \text{Cov}_{\mu}(X_i, X_j) \right] = \mathbb{E}_{\mathcal{T}} \left[ \frac{1}{N_{\mathcal{T}}^2} \sum_{i,j=1}^{N_{\mathcal{T}}} (e^{\alpha W_t} - 1) e^{-4\mu t} \right].$$

FIGURE 4. Defining  $W_{\mathcal{T}}(i, j)$ .

This can be manipulated to:

$$\mathbb{E}_{\mathcal{T}}[\text{Var}_{\mu}(Y_t|\mathcal{T})] = \mathbb{E}_{\mathcal{T}} \left[ \frac{1}{N_{\mathcal{T}}^2} \sum_{i,j=1}^{N_{\mathcal{T}}} e^{\alpha W_{\mathcal{T}}(i,j)} \right] e^{-4\mu t} - e^{-4\mu t}.$$

Let's consider  $\mathbb{E}_{\mathcal{T}} \left[ \frac{1}{N_{\mathcal{T}}^2} \sum_{i,j=1}^{N_{\mathcal{T}}} e^{\alpha W_{\mathcal{T}}(i,j)} \right]$ , and let

$$d_{\mathcal{T}}(t) := \frac{1}{N_{\mathcal{T}}^2} \sum_{i,j=1}^{N_{\mathcal{T}}} e^{\alpha W_{\mathcal{T}}(i,j)} \text{ and } D(t) := \mathbb{E}_{\mathcal{T}}[d(t)].$$

**Lemma 3.1.2.** *The squared coefficient of variation of the random variable  $Y_t$ , i.e.  $\frac{\sigma_t^2}{\mu_t^2}$ , is given by  $D(t) - 1$ .*

Consider the change from time  $t$  to  $t + \delta$ , three events could happen: No speciation occurs with probability  $1 - N_{\mathcal{T}}\lambda\delta + O(\delta^2)$ ; One speciation occurs with probability  $N_{\mathcal{T}}\lambda\delta + O(\delta^2)$ ; Two or more speciation occurs with probability  $O(\delta^2)$ .

This lead the following equation:

$$(3.1.7) \quad d(t+\delta) = d(t) (1 - N_{\mathcal{T}}\lambda\delta + \delta^2) + \frac{1}{(N_{\mathcal{T}} + 1)^2} \sum_{i,j=1}^{N_{\mathcal{T}}+1} \left( e^{\alpha W'_{\mathcal{T}}(i,j)} \right) (N_{\mathcal{T}}\delta\lambda\delta + O(\delta^2)) + O(\delta^2).$$

Considering the distribution of  $W'_{\mathcal{T}}(i, j)$ , we found that since when one of the leaf speciates, each of the leaf has the same chance of speciation. Hence:

$$(3.1.8) \quad \mathbb{E}_{\mathcal{T}} \left[ \frac{1}{(N_{\mathcal{T}} + 1)^2} \sum_{i,j=1}^{N_{\mathcal{T}}+1} \left( e^{\alpha W'_{\mathcal{T}}(i,j)} \right) (N_{\mathcal{T}}\lambda) \right] = \mathbb{E}_{\mathcal{T}} \left[ \frac{\lambda(N_{\mathcal{T}} + 2)(N_{\mathcal{T}})^2}{(N_{\mathcal{T}} + 1)^2(N_{\mathcal{T}})^2} \sum_{i,j=1}^{N_{\mathcal{T}}} e^{\alpha W_{\mathcal{T}}(i,j)} + \frac{\lambda N_{\mathcal{T}}}{(N_{\mathcal{T}} + 1)^2} e^{\alpha t} \right].$$

Now let  $\delta \rightarrow 0$ , rearrange, and take the expectation of both sides of Eqn. 3.1.7 to obtain:

$$(3.1.9) \quad \frac{dD(t)}{dt} = -\lambda \mathbb{E}_{\mathcal{T}} [N_{\mathcal{T}}d(t)] + \lambda \mathbb{E}_{\mathcal{T}} \left[ \frac{1}{(N_{\mathcal{T}} + 1)^2} \sum_{i,j=1}^{N_{\mathcal{T}}+1} \left( e^{\alpha W'_{\mathcal{T}}(i,j)} \right) N_{\mathcal{T}} \right].$$

This equation can be simplified with the help of Eqn. 3.1.8.

$$\frac{dD(t)}{dt} = \lambda \mathbb{E}_{\mathcal{T}} \left[ \frac{N_{\mathcal{T}}}{(N_{\mathcal{T}} + 1)^2} \right] e^{\alpha t} - \lambda \mathbb{E}_{\mathcal{T}} \left[ \frac{N_{\mathcal{T}}}{(N_{\mathcal{T}} + 1)^2} d_{\mathcal{T}}(t) \right].$$

This equation can be further simplified to:

$$\frac{dD(t)}{dt} = \lambda \mathbb{E}_{\mathcal{T}} \left[ \frac{N_{\mathcal{T}}}{(N_{\mathcal{T}} + 1)^2} (e^{\alpha t} - d_{\mathcal{T}}(t)) \right],$$

this can be turned into an inequality:

$$\frac{dD(t)}{dt} \leq \lambda \mathbb{E}_{\mathcal{T}} \left[ \frac{1}{N_{\mathcal{T}}} (e^{\alpha t} - d_{\mathcal{T}}(t)) \right].$$

Now,  $d_{\mathcal{T}}(t) \geq 1$ , for all  $t \geq 0$ , so:

$$\frac{dD(t)}{dt} \leq \lambda \mathbb{E}_{\mathcal{T}} \left[ \frac{1}{N_{\mathcal{T}}} (e^{\alpha t} - 1) \right].$$

Moreover,  $\mathbb{E}_{\mathcal{T}} \left[ \frac{1}{N_{\mathcal{T}}} \right] = \frac{\lambda t e^{-\lambda t}}{1 - e^{-\lambda t}}$ , so

$$D(t) \leq \int_0^t \frac{\lambda^2 s e^{-\lambda s}}{1 - e^{-\lambda s}} (e^{\alpha s} - 1) ds + D(0).$$

Where  $D(0) = 1$ , and  $\frac{\sigma^2}{\mu^2} = D(t) - 1$ , by lemma 3.1.2.

Therefore:

$$(3.1.10) \quad r := \lim_{t \rightarrow \infty} \frac{\sigma^2}{\mu^2} \leq \int_0^t \frac{\lambda^2 s e^{-\lambda s}}{1 - e^{-\lambda s}} (e^{\alpha s} - 1) ds.$$

Put  $u = \lambda s$ ,  $du = \lambda ds$ , so that Eqn. 3.1.10 becomes:

$$(3.1.11) \quad r \leq \int_0^t \frac{u e^{-u}}{1 - e^{-u}} \left( e^{(\alpha/\lambda)u} - 1 \right) du.$$

Now,  $\frac{u}{1 - e^{-u}} \leq 1 + u$  for all  $u \geq 0$ . Hence, from, Eqn. 3.1.11:

$$(3.1.12) \quad r \leq \int_0^t (1 + u) \left( e^{(\alpha/\lambda-1)u} - e^{-u} \right) du.$$

Using integration by parts with  $W = (1 + u)$ ,  $dW/du = 1$ ,  $dZ = e^{-u(1-\alpha/\lambda)} - e^{-u}$ ,  $Z = \frac{-1}{1-\alpha/\lambda} e^{-u(1-\alpha/\lambda)} + e^{-u}$  we obtain:

$$\int_0^t (1 + u) \left( e^{(\alpha/\lambda-1)u} - e^{-u} \right) du = \frac{1}{(1 - \alpha/\lambda)^2} - 1.$$

Thus, from Eqn. 3.1.10 and  $\alpha = 4\mu$  we get:

$$\lim_{t \rightarrow \infty} \frac{\sigma^2}{\mu^2} \leq \frac{1}{(1 - \alpha/\lambda)^2} - 1 \rightarrow 0 \text{ as } \frac{\mu}{\lambda} \nearrow 0.$$

Hence, by Chebyshev's inequality, we obtain the desired theorem.  $\square$

In particular, we obtain:

**Corollary 3.1.3.**  $p_{MR}^{\text{Yule}} \rightarrow 1$  as  $\mu/\lambda \rightarrow 0$ .

We can show  $Z_t$  is not always asymptotically normal as  $t \rightarrow \infty$ . Suppose to the contrary that  $Z_t$  were asymptotically normal as  $t \rightarrow \infty$ . By Corollary 3.1.3 we can conclude that  $\mathbb{P}(Z_\infty > 0) \geq 0.9$  for  $\frac{\mu}{\lambda}$  sufficiently small. However if  $Z_t$  is asymptotically normal when  $t \rightarrow \infty$ , by Eqn. 3.1.5 and this implies that:

$$(3.1.13) \quad \mathbb{P}(Z_t < 0) = \mathbb{P} \left( (Z_t - \mathbb{E}[Z_t]) / \sqrt{\text{Var}(Z_t)} < \frac{\mathbb{E}[Z_t]}{\sqrt{\text{Var}(Z_t)}} \right) \geq \mathbb{P}((Z_t - \mathbb{E}[Z_t]) / \sqrt{\text{Var}(Z_t)} < 1 - \epsilon) \approx 0.16.$$

This forms a contradiction with  $\mathbb{P}(Z_\infty > 0) \geq 0.9$  hence  $Z_\infty$  is not always asymptotically normal.

I would like to propose the following conjecture that:

**Conjecture 3.1.4.**

$$(3.1.14) \quad p_{\text{MR}}^{\text{Yule}} > 1/2 \text{ if } \lambda > 4\mu.$$

Here, it's worth mentioning a result in ([1] page 413) which implies that:

$$p_{\text{MR}}^{\text{Yule}}(t) \geq \frac{1}{2} \text{ for all } \mu, \lambda \text{ and all } t > 0.$$

**3.2. Parsimony and Majority Rule Upper Bound.** Without loss of generality, let the root state be 1. A result from [2] states:  $p_{\text{PAR}}^{\text{Yule}}(t) > \frac{1}{2}$  if  $\mu < \frac{1}{6}\lambda$  otherwise  $p_{\text{PAR}}^{\text{Yule}} \rightarrow \frac{1}{2}$  as  $t \rightarrow \infty$ .

We can see that for the one lineage start Yule tree:

$$(3.2.1) \quad p_{\text{PAR}}^{\text{Yule}} \leq 1 - Q(p_{\text{PAR}}^{\text{Yule}}) \text{ as } t \rightarrow \infty.$$

The quantity  $Q$  can be calculated as:

$$Q = \int_0^\infty \frac{1}{2} (1 - \exp(-2l\mu)) \lambda \exp(-\lambda l) dl,$$

which can be calculated to be:

$$Q = \frac{\mu}{\lambda + 2\mu} = \frac{\rho}{1 + 2\rho} \text{ for } \rho = \frac{\lambda}{\mu}.$$

**Theorem 3.2.1.** *Therefore, solving the linear equation above (for the one lineage start) results in:*

$$(3.2.2) \quad p_{\text{PAR}}^{\text{Yule}} \leq \frac{1}{1 + Q}.$$

*A similar upper bound calculation could be done for majority rule as the same inequality constraints exists for  $p_{\text{MR}}^{\text{Yule}}$ , hence we can conclude:*

$$(3.2.3) \quad p_{\text{MR}}^{\text{Yule}} \leq \frac{1}{1 + Q}.$$

Combined with the monotonicity of  $p_{\text{PAR}}^{\text{Yule}}$ , we can conclude the following:

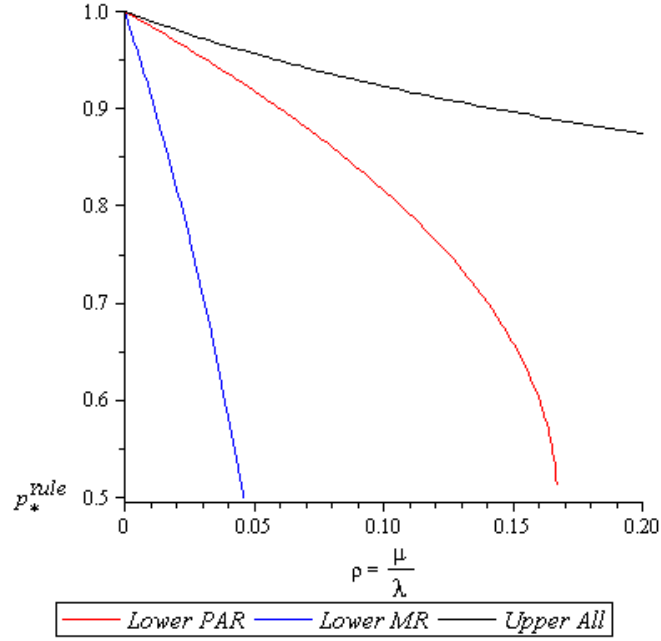


FIGURE 5. Upper and lower correct estimation probability bounds as  $t \rightarrow \infty$ .

**Corollary 3.2.2.**

$$(3.2.4) \quad p_{\text{PAR}}^{\text{Yule}}(t) \leq \frac{1}{1+Q}, \text{ for all } t \geq 0.$$

All the correct reconstruction bounds for parsimony and majority rule are plotted in Fig. 5.

**3.3. Star-tree.** Consider now a star-tree as shown in Fig. 6, and suppose that  $N_t \sim \text{geo}(\lambda t)$ . Let the symbol  $*$  be conditioning  $\mathcal{T}$  to be a star tree with  $N_t$  leaves. Define

$$p_{\text{MR}}^*(t) := \mathbb{E}_{\mathcal{T}}[p_{\text{MR}}(\mathcal{T}_t^*)].$$

Then we obtain the following result:

**Theorem 3.3.1.**

$$(3.3.1) \quad p_{\text{MR}}^*(t) \rightarrow 1 \text{ as } t \rightarrow \infty \text{ for all } \lambda > 4\mu.$$

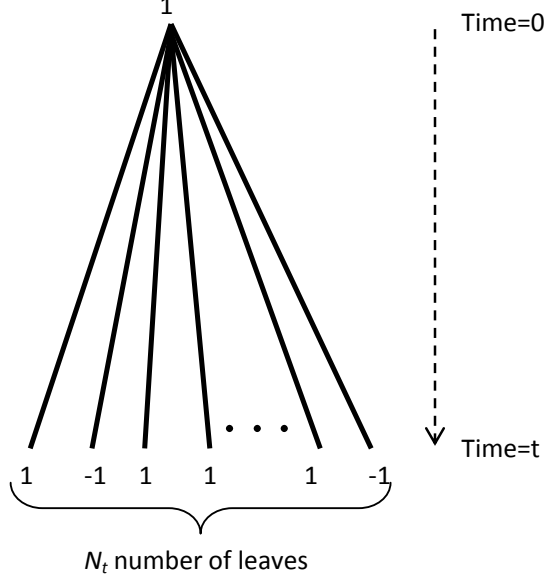


FIGURE 6. Example of a Star-tree.

*Proof.* Let  $\mathcal{T}^*$  be a star-tree with  $N_t \sim \text{geo}(\lambda t)$ . Let  $Y_t = \frac{\sum_{i=1}^{N_t} X_i}{N_{\mathcal{T}^*}}$ . Let's consider the first two moments  $\mathbb{E}[Y_t]$  and  $\text{Var}(Y_t)$ . For  $\mathbb{E}[Y_t]$ , we obtain:

**Lemma 3.3.2.**

$$\mathbb{E}[Y_t] = \mathbb{E}_{\mathcal{T}^*} \left[ \mathbb{E}_{\mu} \left[ \frac{Z_t}{N_{\mathcal{T}^*}} \mid \mathcal{T}^* \right] \right] = \mathbb{E}_{\mathcal{T}^*} [e^{-2\mu t}] = e^{-2\mu t}.$$

For  $\text{Var}(Y_t)$ , we obtain:

$$\text{Var}(Y_t) = \text{Var}_{\mathcal{T}^*}(\mathbb{E}_{\mu}[Y_t \mid \mathcal{T}^*]) + \mathbb{E}_{\mu}[\text{Var}_{\mathcal{T}^*}(Y_t \mid \mathcal{T}^*)].$$

Firstly, it is clear that:

$$\text{Var}_{\mathcal{T}^*}(\mathbb{E}_{\mu}[Y_t \mid \mathcal{T}^*]) = \text{Var}_{\mathcal{T}^*}(e^{-2\mu t}) = 0.$$

Secondly, we can derive an upper bound for  $\mathbb{E}_{\mathcal{T}^*}[\text{Var}_{\mu}(Y_t \mid \mathcal{T}^*)]$ :

$$\mathbb{E}_{\mathcal{T}^*}[\text{Var}_{\mu}(Y_t \mid \mathcal{T}^*)] = \mathbb{E}_{\mathcal{T}^*} \left[ \text{Var}_{\mu} \left( \sum_{i=1}^{N_{\mathcal{T}^*}} \frac{X_i}{N_{\mathcal{T}^*}} \mid \mathcal{T}^* \right) \right].$$

Now, in a star tree, each  $X_i$  is independently (and identically) distributed, hence:

$$\mathbb{E}_{\mathcal{T}^*}[Var_{\mu}(Y_t|\mathcal{T}^*)] = \mathbb{E}_{\mathcal{T}^*}\left[\frac{Var_{\mu}(X_i)}{N_{\mathcal{T}^*}}\right].$$

We know that  $Var_{\mu}(X_i) = 1 - e^{-4\mu t}$ , therefore:

$$\mathbb{E}_{\mathcal{T}^*}[Var_{\mu}(Y_t|\mathcal{T}^*)] = \mathbb{E}_{\mathcal{T}^*}\left[\frac{1 - e^{-4\mu t}}{N_{\mathcal{T}^*}}\right].$$

Since  $\mathbb{E}_{\mathcal{T}^*}\left[\frac{1}{N_{\mathcal{T}^*}}\right] = \frac{\lambda t e^{-\lambda t}}{1 - e^{-\lambda t}}$ , we obtain:

$$\mathbb{E}_{\mathcal{T}^*}[Var_{\mu}(Y_t|\mathcal{T}^*)] = \frac{(1 - e^{-4\mu t})\lambda t e^{-\lambda t}}{1 - e^{-\lambda t}}.$$

Hence, for all  $\lambda > 4\mu$ , we obtain:

$$\frac{\mathbb{E}[Y_t]^2}{Var(Y_t)} = e^{-4\mu t} \div \frac{(1 - e^{-4\mu t})\lambda t e^{-\lambda t}}{1 - e^{-\lambda t}} = \frac{e^{(\lambda - 4\mu)t}(1 - e^{-\lambda t})}{\lambda t(1 - e^{-4\mu t})} \rightarrow \infty, \text{ as } t \rightarrow \infty.$$

Applying Chebyshev's inequality suffices to prove the theorem.  $\square$

#### 4. WHICH METHOD IS BETTER? MAXIMUM LIKELIHOOD VS. MAJORITY RULE VS. PARSIMONY

Firstly, let us consider binary trees with just three leaves. Such a tree will have eight different combinations of leaf states. There are two states that cause majority rule and parsimony to have a different probability of estimating the right root state. Assume the root state is 1. For leaf combination -1,-1,1, majority rule has probability 0 of correct estimate, while parsimony has probability 0.5 of correct estimate; On the other hand for leaf combination 1,1,-1, majority rule has probability 1 of correct estimate, while parsimony has a probability 0.5 of correct estimate (see Fig. 7). Therefore, there are scenarios in which either method can perform more accurately than the other.

What if we condition on the topological tree, with given edge lengths? Consider the same three-leaf example, let the speciation occur in the left branch at time  $x$  (for some fixed  $x < t$ ), where  $t$  is the total time (see Fig. 7). For leaf state -1, -1, 1 (say state P), majority rule has probability 0 of correct estimate, while parsimony only



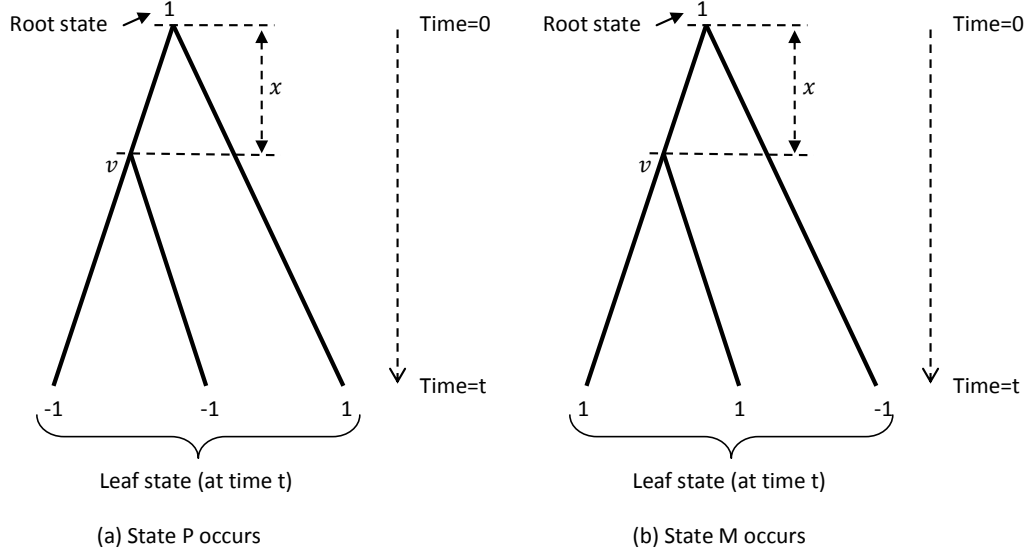


FIGURE 7. Three-leaf example of the 2-state model

got probability 0.5 of correct estimate. On the other hand for leaf state 1, 1, -1 (say state M), majority rule has probability 1 of correct estimate, while the parsimony has probability 0.5 of correct estimate. While all the other outcomes, majority rule and parsimony have the same probability of correct estimate. Which method would perform better overall depends on the probability of each outcome state (P and M).

The following lemma will be useful in this section:

**Lemma 4.0.3.** *Let  $X_i$  be the state of the  $i$ th leaf and let  $X_1, X_2, \dots, X_N$  be the set of random variables for the states of each leaf.  $X_i$ 's are identically distributed (but not independent). We have:*

$$\mathbb{P}(X_i = 1) = \frac{1}{2}(1 + \exp[-2\mu t])$$

and

$$\mathbb{P}(X_i = -1) = \frac{1}{2}(1 - \exp[-2\mu t]).$$

Therefore:

$$\mathbb{E}[X_i] = \exp[-2\mu t].$$

By using Lemma 4.0.3, we can calculate  $\mathbb{P}(\text{state P}|x)$  and  $\mathbb{P}(\text{state M}|x)$  by conditioning the two cases where vertex  $v$  in Fig. 7 is state 1 or -1:

**Proposition 4.0.4.**

$$\begin{aligned} \mathbb{P}(\text{state } P|x) &= \left(\frac{1}{16} (1 + e^{-2\mu x})\right) \left(1 - e^{-2\mu(t-x)}\right)^2 (1 + e^{-2\mu t}) \\ &\quad + \left(\frac{1}{16} (1 - e^{-2\mu x})\right) \left(1 + e^{-2\mu(t-x)}\right)^2 (1 + e^{-2\mu t}), \\ \mathbb{P}(\text{state } M|x) &= \left(\frac{1}{16} (1 + e^{-2\mu x})\right) \left(1 + e^{-2\mu(t-x)}\right)^2 (1 - e^{-2\mu t}) \\ &\quad + \left(\frac{1}{16} (1 - e^{-2\mu x})\right) \left(1 - e^{-2\mu(t-x)}\right)^2 (1 - e^{-2\mu t}). \end{aligned}$$

By the above proposition we obtain the following:

**Theorem 4.0.5.**  $\mathbb{P}(\text{state } P|x) - \mathbb{P}(\text{state } M|x) = -(1/4)e^{-2\mu t} + (1/4)e^{-2\mu(3t-2x)} < 0$  for all  $t > x$ .

*In particular, for any tree (including Yule tree) with three leaves with clocked edge lengths, state M is more likely than state P for any value of x on a 2-state symmetric model.*

Going back to Theorem 4.0.5, it is interesting to note the discontinuity of the difference as  $\mathbb{P}(\text{state } P|x=0) = \mathbb{P}(\text{state } M|x=0)$ . On the other hand,  $\mathbb{P}(\text{state } P|x) - \mathbb{P}(\text{state } M|x) \rightarrow -(1/4)e^{-2\mu t} + (1/4)e^{-2\mu(3t)} < 0$  as  $x \rightarrow 0^+$ .

In the light of this result, we pose the following two conjectures:

**Conjecture 4.0.6.**

$$(4.0.2) \quad p_{\text{MR}}^{\text{Yule}}(t) > p_{\text{PAR}}^{\text{Yule}}(t), \text{ for all } t > 0.$$

**Conjecture 4.0.7.**

$$(4.0.3) \quad p_{\text{MR}}^{\text{Yule}}(t; n) > p_{\text{PAR}}^{\text{Yule}}(t; n), \text{ for all } t > 0.$$

The above 3-leaf analysis established the affirmative result for Conjecture 4.0.7 when  $n = 3$ . Note also that the truth of Conjecture 4.0.7 would imply the truth of Conjecture 4.0.6.

Fig. 8 shows a plot of simulations by Olivier Gascuel (pers. comm.) for 1000 taxa (leaves, species), on 1000 randomly generated Yule trees (for each  $\mu/\lambda$  ratio) and 1000 mutation process in each generated Yule tree. Different methods are tested to find the percentage of correct estimation for each method in terms of  $\mu/\lambda$  ratio. Maximum likelihood method outperforms every other type of reconstruction (this can be formally proven, due to the uniform prior). Also from the graph, it seems majority rule outperforms parsimony, even though majority rule uses less information than parsimony! This is interesting in the light of a quotation from [3]:

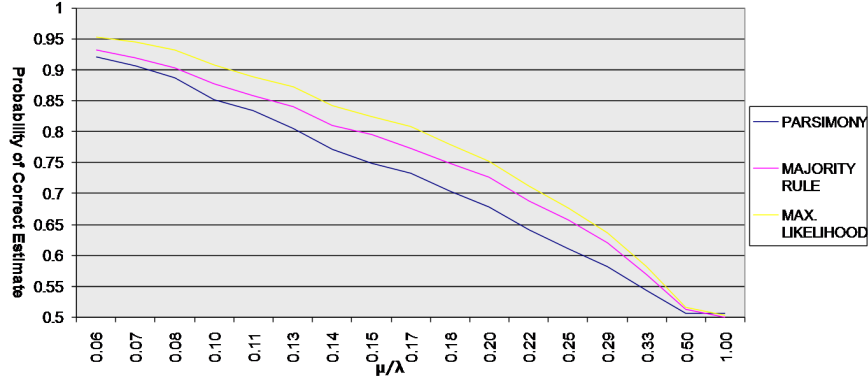


FIGURE 8. Plot for simulation on 1000 taxa

*“Incorporation phylogenetic uncertainty very rarely changes the inferred ancestral state and does not improve the accuracy of the reconstructed ancestral sequence.”*

In fact our proofs and simulations support this idea, since majority rule does not use the structure of the tree and seems to have a higher correct reconstruction probability than parsimony. One interesting fact is the existence of a ultrametric binary tree for which parsimony performs better than majority rule, as shown in Fig. 9.

By Lemma 4.0.3, the probability of a correct estimation by majority rule and parsimony given the tree  $\mathcal{T}_t^\tau$  as Fig. 9 is (letting  $a = \frac{1}{2}(1 + \exp[-2\mu t]) > 0.5$  and letting  $\epsilon \rightarrow 0$ ):

$$p_{\text{MR}}^{\mathcal{T}^\tau}(t) = a,$$

and, a case analysis shows that:

$$p_{\text{PAR}}^{\mathcal{T}^\tau}(t) = \frac{1}{2} \binom{4}{2} a^2 (1-a)^2 + \binom{4}{1} a^3 (1-a) + \binom{4}{0} a^4 = 3a^2 - 2a^3.$$

The last two equations imply:

$$p_{\text{PAR}}^{\mathcal{T}^\tau}(t) \geq p_{\text{MR}}^{\mathcal{T}^\tau}(t), \text{ for any } a \text{ such that } \frac{1}{2} \leq a \leq 1.$$

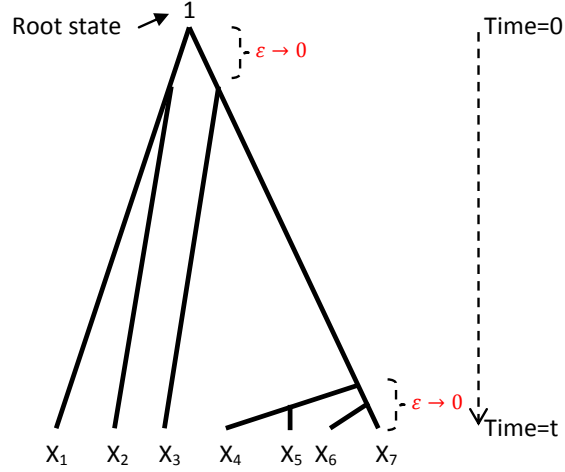


FIGURE 9. Seven-leaf example where Parsimony outperforms Majority Rule

## 5. K-STATE MODEL

Since DNA sequences are made up of four types of nucleotides and various other bio-molecular data has more than 2 states, the study of a  $k$ -state model (where  $k \geq 2$ ) would help to understand more realistic evolutionary processes.

In this section, we still work under the Yule model, with parameters  $t$  (time),  $\lambda$  (speciation rate) and  $\mu$  (mutation rate). The variables  $t$  and  $\lambda$  are the same as in the last section, however  $\mu$  is the rate of changing from state  $i$  to any particular other state  $j$ .

**5.1. Preliminary.** Let the probability of the state change from  $a$  to  $b$ , conditioned on starting with state  $a$ , along a edge length  $y$  be  $\mathbb{P}_{ab}(y)$ .  $\mathbb{P}_{ab}(y)$  can be calculated by considering how it changes as the edge length  $y$  increases to  $y + \delta$  (with  $\delta$  small):

$$\mathbb{P}_{ab}(y + \delta) = (1 - (r - 1)\mu\delta)\mathbb{P}_{ab}(y) + \frac{(r - 1)\mu\delta}{r - 1}(1 - \mathbb{P}_{ab}(y)) + O(\delta^2).$$

Letting  $\delta \rightarrow 0$  we obtain:

$$\frac{d\mathbb{P}_{ab}(t)}{dt} = \mu - r\mu\mathbb{P}_{ab}(t).$$

Solving this differential equation (by the integration factor method) with the initial condition  $\mathbb{P}_{ab}(0) = 0$  if  $a \neq b$  and  $\mathbb{P}_{ab}(0) = 1$  if  $a = b$  we obtain:

**Lemma 5.1.1.**

$$(5.1.1) \quad \mathbb{P}_{ab}(t) = \begin{cases} \frac{1}{r}(1 - e^{-r\mu t}), & \text{if } a \neq b; \\ \frac{1}{r}(1 + (r-1)e^{-r\mu t}), & \text{if } a = b. \end{cases}$$

**5.2. System of DEs for k-state Parsimony.** For the  $k$ -state model, consider the parsimony method. We define  $P_m(t)$  to be the probability the root estimation contains  $m$  states and one of those state is the true state;  $Q_m(t)$  to be the probability the root estimation contains  $m$  states and none of those state is the true state. Those two sets of probabilities are dependent on  $t$  (time),  $\lambda$  (speciation rate) and  $\mu$  (mutation rate).

Consider the first  $\delta$  period of time (for the Yule tree with  $k$  states), there are four events that could happen. Either:

- (i) no speciation nor mutation occurs,
- (ii) one speciation and no mutation occurs,
- (iii) no speciation occurs and one mutation occurs, Or
- (iv) the number of mutation plus speciation events is greater than one.

Also each event has a corresponding probability. There is a probability of  $1 - \lambda\delta - \mu\delta + O(\delta)^2$  that neither speciation nor mutation occurs; a probability of  $\lambda\delta + O(\delta)^2$  that one speciation and no mutation occurs; a probability of  $\mu\delta + O(\delta)^2$  that no speciation and one mutation occurs; and a probability of  $O(\delta)^2$  that the number of mutation plus speciation evnts is greater than one.

In this way we can generate a system of differential equations for  $P_m(t)$  and  $Q_m(t)$  (note that the sum of  $P_m$  and  $Q_m$  equals 1).

$$\begin{aligned} \frac{dP_m(t)}{dt} &= (-\mu - \lambda)P_m(t) + \mu \left( P_m(t) \cdot \frac{m-1}{r-1} + Q_m(t) \cdot \frac{m}{r-1} \right) \\ &+ \lambda \left( \sum_{v,w \geq m} P_v(t)P_w(t) \cdot \frac{\binom{r-1}{m-1} \binom{r-m}{v-m} \binom{r-v}{w-m}}{\binom{r-1}{v-1} \binom{r-1}{w-1}} \right) \\ &+ \lambda \left( \sum_{v+w=m} 2Q_v(t)P_w(t) \cdot \frac{\binom{r-1}{v-1} \binom{r-v}{w}}{\binom{r-1}{v-1} \binom{r-1}{w}} \right). \\ \frac{dQ_m(t)}{dt} &= (-\mu - \lambda)Q_m(t) + \mu(P_m(t) \cdot \frac{r-m}{r-1} + Q_m(t) \cdot \frac{r-m-1}{r-1}) \end{aligned}$$

$$\begin{aligned}
& +\lambda \left( \sum_{v,w \geq m} Q_v(t)Q_w(t) \cdot \frac{\binom{r-1}{m} \binom{r-m-1}{v-m} \binom{r-v-1}{w-m}}{\binom{r-1}{v} \binom{r-1}{w}} \right) \\
& +\lambda \left( \sum_{v+w=m} Q_v(t)Q_w(t) \cdot \frac{\binom{r-1}{v} \binom{r-v-1}{w}}{\binom{r-1}{v} \binom{r-1}{w}} \right) \\
& +\lambda \left( \sum_{v \geq m, w \geq m+1} 2Q_v(t)P_w(t) \cdot \frac{\binom{r-1}{m} \binom{r-m-1}{v-m} \binom{r-m-1}{w-m-1}}{\binom{r-1}{v} \binom{r-1}{w-1}} \right).
\end{aligned}$$

We now check that the above equations, when  $r = 2$ , agrees with the D.E. equation from [2].

$$\frac{dP_1(t)}{dt} = (-\mu - \lambda)P_m(t) + \mu Q_m(t) + \lambda \left( \sum_{v,w \geq 1} P_v(t)P_w(t) \cdot \frac{\binom{1}{0} \binom{1}{v-1} \binom{2-v}{w-1}}{\binom{1}{v-1} \binom{1}{w-1}} \right).$$

Hence:

$$\frac{dP_1(t)}{dt} = (-\mu - \lambda)P_m(t) + \mu Q_m(t) + \lambda \left( \sum_{(v,w)=(1,1),(1,2),(2,1)} P_v(t)P_w(t) \right).$$

So the first equation reduces to the 2-state D.E. equation from [2].

Possible future work involves checking the other two equations agree with the two state model, then try to extend this DE for 3-state and solve for a solution.

**5.3. Which method is better? Majority rule vs. Parsimony.** Firstly, let us consider binary trees with just three leaves. Assume the root state is 0. There are two states that cause majority rule and parsimony to have a different probability of correctly estimating the root state. Consider a three-leaf tree, let the speciation occur in the left branch at time  $x$  (for some fixed  $x < t$ ), where  $t$  is the total time (see Fig 10). For leaf states  $a, a, 0$  (say state P), majority rule has probability 0 of a correct estimate, while parsimony only got probability 0.5 of a correct estimate. On the other hand for leaf states  $0, 0, a$  (say state M), majority rule has probability 1 of a correct estimate, while the parsimony has probability 0.5 of a correct estimate. For all the other leaf states outcomes, majority rule and parsimony have the same probability of a correct estimate. Which would perform better overall depends on the relative probabilities of the particular outcome states P and M.

**Lemma 5.3.1.**

$$\mathbb{P}(\text{state P}|x) - \mathbb{P}(\text{state M}|x)$$

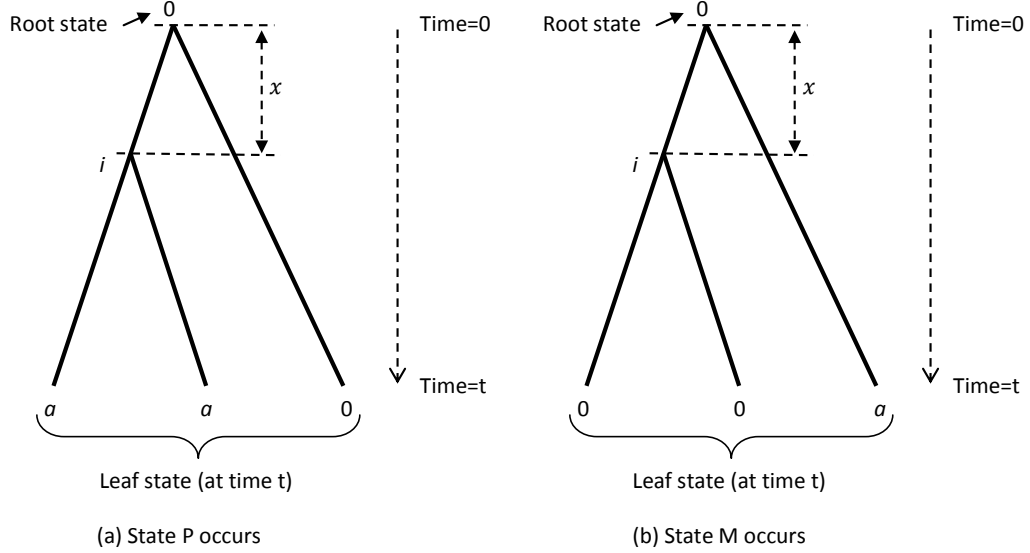


FIGURE 10. Three-leaf example of the  $r$ -state model

$$= \left[ e^{-r\mu(3t-2x)}r - e^{-r\mu(-x+2t)}r + 2e^{-r\mu(-x+2t)} - e^{-r\mu(3t-2x)} - e^{-r\mu t} \right] / r^2.$$

*Proof.* At the speciation point there are  $r$  possibilities for state P and M with the possibilities denoted by  $i$ . The probability of state P conditional on the speciation at time  $x$  with  $0 \leq x \leq t$  is:

$$\mathbb{P}(\text{state P} | x) = \sum_{i=0}^{r-1} \mathbb{P}(\text{state P} \cap i | x).$$

The above equation can be simplified to:

$$(r-2)\mathbb{P}(\text{state P} \cap i \neq 0, a | x) + \mathbb{P}(\text{state P} \cap i = 0 | x) + \mathbb{P}(\text{state P} \cap i = a | x).$$

Each of those probabilities could be calculated by using 5.1.1, to obtain:

$$\mathbb{P}(\text{state P} \cap i \neq 0, a | x) = (1 - e^{-r\mu x}) \left(1 - e^{-r\mu(t-x)}\right)^2 (1 + (r-1)e^{-r\mu t}) / r^4,$$

$$\mathbb{P}(\text{state P} \cap i = 0 | x) = (1 + (r-1)e^{-r\mu x}) \left(1 - e^{-r\mu(t-x)}\right)^2 (1 + (r-1)e^{-r\mu t}) / r^4,$$

$$\mathbb{P}(\text{state P} \cap i = a | x) = (1 - e^{-r\mu x}) \left(1 + (r-1)e^{-r\mu(t-x)}\right)^2 (1 + (r-1)e^{-r\mu t}) / r^4,$$

from which we obtain:

**Proposition 5.3.2.**

$$\begin{aligned} \mathbb{P}(\text{state P}|x) &= (r-2)(1-e^{-r\mu x})\left(1-e^{-r\mu(t-x)}\right)^2(1+(r-1)e^{-r\mu t})/r^4 \\ &\quad + (1+(r-1)e^{-r\mu x})\left(1-e^{-r\mu(t-x)}\right)^2(1+(r-1)e^{-r\mu t})/r^4 \\ &\quad + (1-e^{-r\mu x})\left(1+(r-1)e^{-r\mu(t-x)}\right)^2(1+(r-1)e^{-r\mu t})/r^4. \end{aligned}$$

Similarly, let the probability of state M conditional on the speciation at time  $x$  with  $0 \leq x \leq t$  is:

$$\mathbb{P}(\text{state M}|x) = \sum_{i=0}^{r-1} \mathbb{P}(\text{state M} \cap i|x).$$

The above equation can be simplified to:

$$(r-1)\mathbb{P}(\text{state M} \cap i \neq 0|x) + \mathbb{P}(\text{state M} \cap i = 0|x).$$

Each of these probabilities can be calculated by using 5.1.1, to obtain:

$$\mathbb{P}(\text{state M} \cap i \neq 0|x) = (1-e^{-r\mu x})\left(1-e^{-r\mu(t-x)}\right)^2(1-e^{-r\mu t})/r^4, \text{ and}$$

$$\mathbb{P}(\text{state M} \cap i = 0|x) = (1+(r-1)e^{-r\mu x})\left(1+(r-1)e^{-r\mu(t-x)}\right)^2(1-e^{-r\mu t})/r^4.$$

Hence, we obtain:

**Proposition 5.3.3.**

$$\begin{aligned} \mathbb{P}(\text{state M}|x) &= (r-1)(1-e^{-r\mu x})\left(1-e^{-r\mu(t-x)}\right)^2(1-e^{-r\mu t})/r^4 \\ &\quad + (1+(r-1)e^{-r\mu x})\left(1+(r-1)e^{-r\mu(t-x)}\right)^2(1-e^{-r\mu t})/r^4. \end{aligned}$$

Subtracting (Prop. 5.3.2) from (Prop. 5.3.3) and simplifying, leads to the expression in the lemma.  $\square$

In lemma 5.3.1, notice the denominator is positive, the first 2 terms of the nominator is negative or 0 (by properties of exponential and  $0 \leq x \leq t$ ) and the last 3 terms of the nominator is negative or 0 by the AM-GM inequality. Hence, we obtain the following results.

**Theorem 5.3.4.** *For any tree with three leaves with clocked edge lengths, state M is more likely than state P for any value of  $x$  on a  $r$ -state symmetric model.*

*In particular, for a Yule tree with three leaves, state M is more likely than state P for any value of  $x$  on a  $r$ -state symmetric model.*



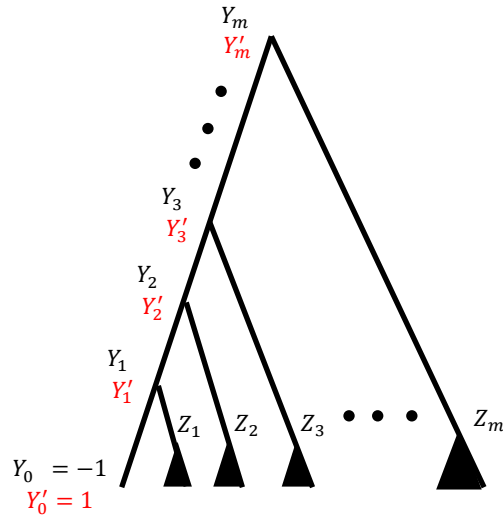


FIGURE 11. 2-state parsimony, one leaf state change

### 6. CHANGING ONE OR MORE ROOT STATE

6.1. **Case I: 2-state parsimony.** What are the possible effects of changing one leaf state from -1 to 1 on the two state model? Define the root estimation before the change to be *original estimation*, and the estimation after the change to be the *modified estimation* (an ' is used).

For majority rule, if the original estimation is  $\{1\}$  or  $\{-1, 1\}$ , then the modified estimation is always  $\{1\}$ ; If the original estimation is  $\{-1\}$ , then the modified estimation is  $\{-1, 1\}$  or  $\{-1\}$ .

For parsimony, consider the tree in Fig. 11 and let:  
 $[Y] = \sum$  all elements in  $Y$ , where  $Y$  is the parsimony estimation set.

**Lemma 6.1.1.** *For a tree with two initial branches, let the original estimation for the branches be  $A$  and  $B$ , while the modified estimation for the branches is  $A'$  and  $B'$ . Let the original estimation for the root be  $C$  and the modified estimation for the branches be  $C'$ . If  $[A']=[A]$ ,  $[A]+1$  or  $[A]+2$ , then  $[C']=[C]$ ,  $[C]+1$  or  $[C]+2$ . [This is both necessary and sufficient]*

This lemma is proved by considering all possible cases of  $A$ ,  $C$  and  $A'$ .

Let the parsimony estimation set be as labeled in Fig. 11, then as the consequence of lemma 6.1.1, we are given  $[Y'_0] = [Y_0]$ ,  $[Y_0] + 1$  or  $[Y_0] + 2$ ; implies  $[Y'_1] = [Y_1]$ ,  $[Y_1] + 1$  or  $[Y_1] + 2$ ; ...  $[Y'_m] = [Y_m]$ ,  $[Y_m] + 1$  or  $[Y_m] + 2$ . Hence we obtain the theorem below:

**Theorem 6.1.2.** *Under a 2-state model with a binary tree, changing one of the leaf states from -1 to 1 can only increase the sum of the root estimation set by 2, 1 or make no change for the root estimation set.*

**6.2. Case II:  $r$ -state parsimony.** What are the possible effects of changing one leaf state from  $a$  to 0 in the  $r$ -state model? As before, we define the root estimation before the change to be *original estimation*, and the estimation after the change to be the *modified estimation* (an ' is used).

For majority rule the outcome is trivial. If the original estimation is 0 or contains 0, then the modified estimation will be the set  $\{0\}$ . If the original estimation does not contains 0, then the modified estimation will either be the same as the original estimation, or it will be the original estimation union  $\{0\}$ .

No similar results for  $r$ -state parsimony holds for the 2-state setting, in fact changing one leaf state from  $a \neq 0$  to 0 can cause the modified estimation to fail to contain 0, while the original estimation does! Similarly, changing one leaf state from  $a \neq 0$  to 0 can cause the modified estimation to contain 0 and some other state, while the original estimation only contains 0. Here is one such example for the latter case (the former case can be obtained by starting the tree in Fig. 12 at the first speciation point, as labeled by \*).

Combining the former and the latter case for  $r$ -state parsimony, it is possible to change two leaf states from  $a$  to 0 and cause the modified estimation to be  $b$ , while the original estimation is 0. An example is shown in Fig. 13.

A remaining problem is whether or not it is possible to change one leaf state from  $a$  to 0 and obtain a modified estimation of  $\{b\}$ ,  $b \neq 0$ , which the original estimation is  $\{0\}$ ? This turns out to be impossible by the following result.

**Proposition 6.2.1.** *Changing one leaf state from  $a \neq 0$  to 0 cannot result in the modified estimation to be  $b \neq 0$ , if the original estimation is  $\{0\}$ .*

*Proof.* We apply proof by contradiction. Assume there is a tree such that the modified estimation to be  $b \neq 0$ , while the original estimation is  $\{0\}$ . Then consider the two edges that lead to the root, as in Fig. 14: Firstly, one side of the root has to contain the change, and the modified and original parsimony estimation for that side must be 0,  $S_1$  and  $S_2$  respectively (as the single root state must be the result of intersections) and such that  $S_2 \cap S_1 = \phi$  and  $\{0\} \cap S_2 = \phi$ . Find the lowest position on the tree such  $S_2$  and 0,  $S_1$  occurs.

Secondly, consider  $A$  and  $A'$  through the following four cases:

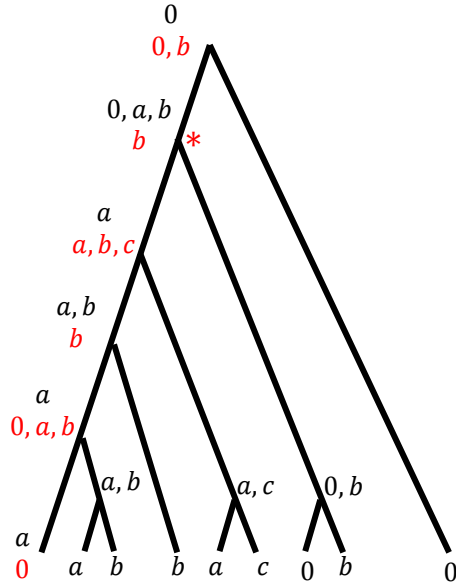


FIGURE 12. Counter example for the  $r$ -state model change

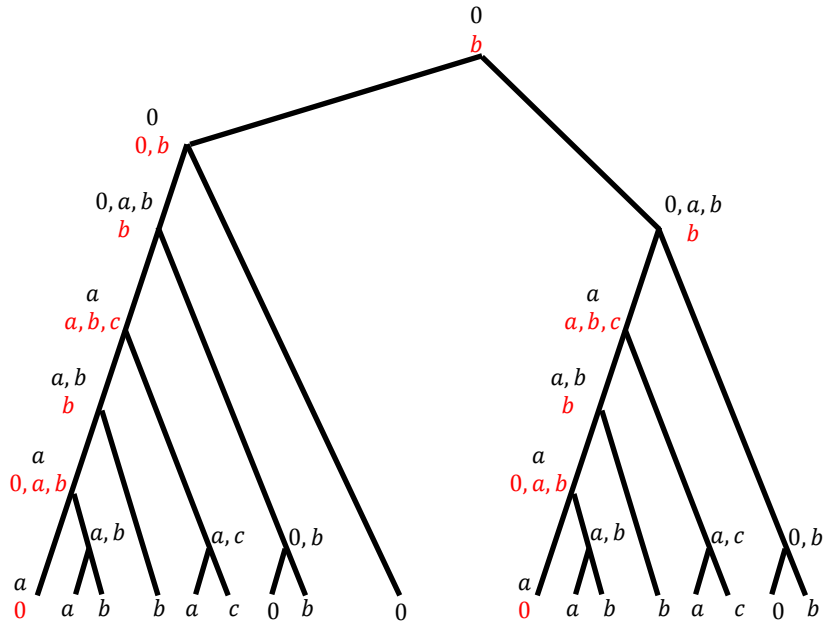
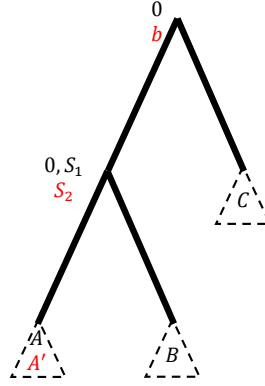


FIGURE 13. Counterexample for the  $r$ -state model with 2 changes

FIGURE 14. Representation of  $\mathcal{T}$  for the  $r$ -state model with 1 changes

|          |              |              |              |       |
|----------|--------------|--------------|--------------|-------|
| $0, S_1$ | intersection | intersection | union        | union |
| $S_2$    | intersection | union        | intersection | union |

Here ‘intersection’ means the set is obtained by parsimony through intersection, and ‘union’ means the set is obtained by parsimony through union.

If  $0, S_1$  and  $S_2$  both come from intersection, then  $A$  and  $A'$  would satisfy the requirements for  $S_2$  and  $0, S_1$ , hence we obtain a contradiction (as no root state satisfies this requirement); If  $0, S_1$  comes from intersection and  $S_2$  comes from union, then  $0 \in S_2$ , which is again a contradiction; If  $0, S_1$  comes from union and  $S_2$  comes from intersection, then  $S_2 \in S_1$ , a contradiction with  $S_2 \cap S_1 = \phi$ ; If  $0, S_1$  and  $S_2$  both come from union, then  $A$  and  $A'$  would satisfy the requirements for  $S_2$  and  $0, S_1$ , hence, once again, we obtain a contradiction.  $\square$

#### ACKNOWLEDGMENT

I would like to thank my supervisor Prof. M. Steel for the kind time spent for this project. Also, thanks to Prof Oliver Gascuel for providing the simulations results presented in Fig. 8, to friends in University of Canterbury and to my parents who supported me and showed interest in this project this year.

## REFERENCES

- [1] Evans, W., Kenyon, C., Peres, Y. and Schulman L.J. (2000). Broadcasting on trees and the using model. *The Annals of Applied Probability*, 10(3): 410-433.
- [2] Gascuel, O. and Steel, M. (2010). Inferring ancestral sequences in taxon-rich phylogenies. *Mathematical Biosciences*, 227: 125-135.
- [3] Hanson-Smith, V., Kolaczkowski, B. and J. W. Thornton. (2010). Robustness of Ancestral Sequence Reconstruction to Phylogenetic Uncertainty. *Molecular Biology Evolution*, 27(9): 1988-1999.
- [4] Liberles, L. (2007). Ancestral sequence reconstruction. Oxford University Press. ISBN13: 978-0-19-929918-8.
- [5] Mossel, E. and Steel, M. (2005). How much can evolved characters tell us about the tree that generated them? pp. 384-412 In "Mathematics of Evolution and Phylogeny," Oxford University Press.
- [6] Sober, E. and Steel, M. (2002). Testing the hypothesis of common ancestry. *Journal of Theoretical Biology*, 218: 395-408.
- [7] Steel, M. and Mooers, A. (2009). Expected length of pendant and interior edges of a Yule tree. *Applied Mathematics Letters*, 23(11): 1315-1319
- [8] Yule, G.U. (1925). A Mathematical Theory of Evolution, based on the Conclusions of Dr. J. C. Willis, F.R.S.. *Philosophical Transactions of the Royal Society of London Ser. B*, 213: 21-87.

*E-mail address:* thomasli1990@hotmail.com or hli77@student.canterbury.ac.nz