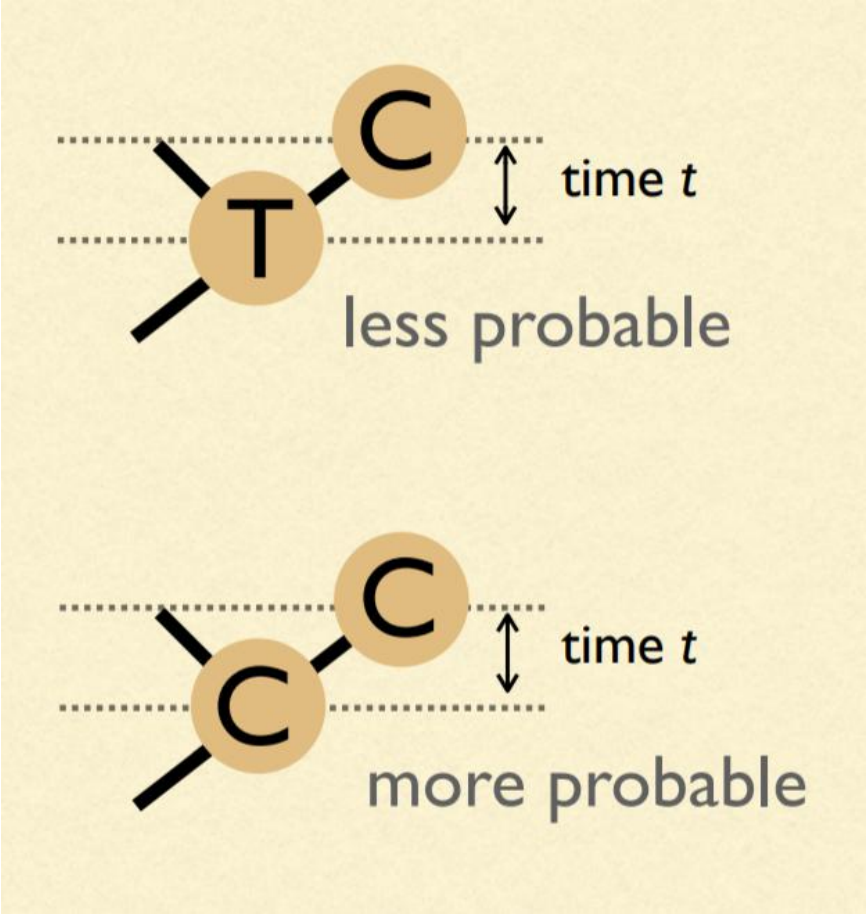
Continuous-Time Markov Models

Markov assumption **= probability of change depends only on current state, not how long it has been in that state**

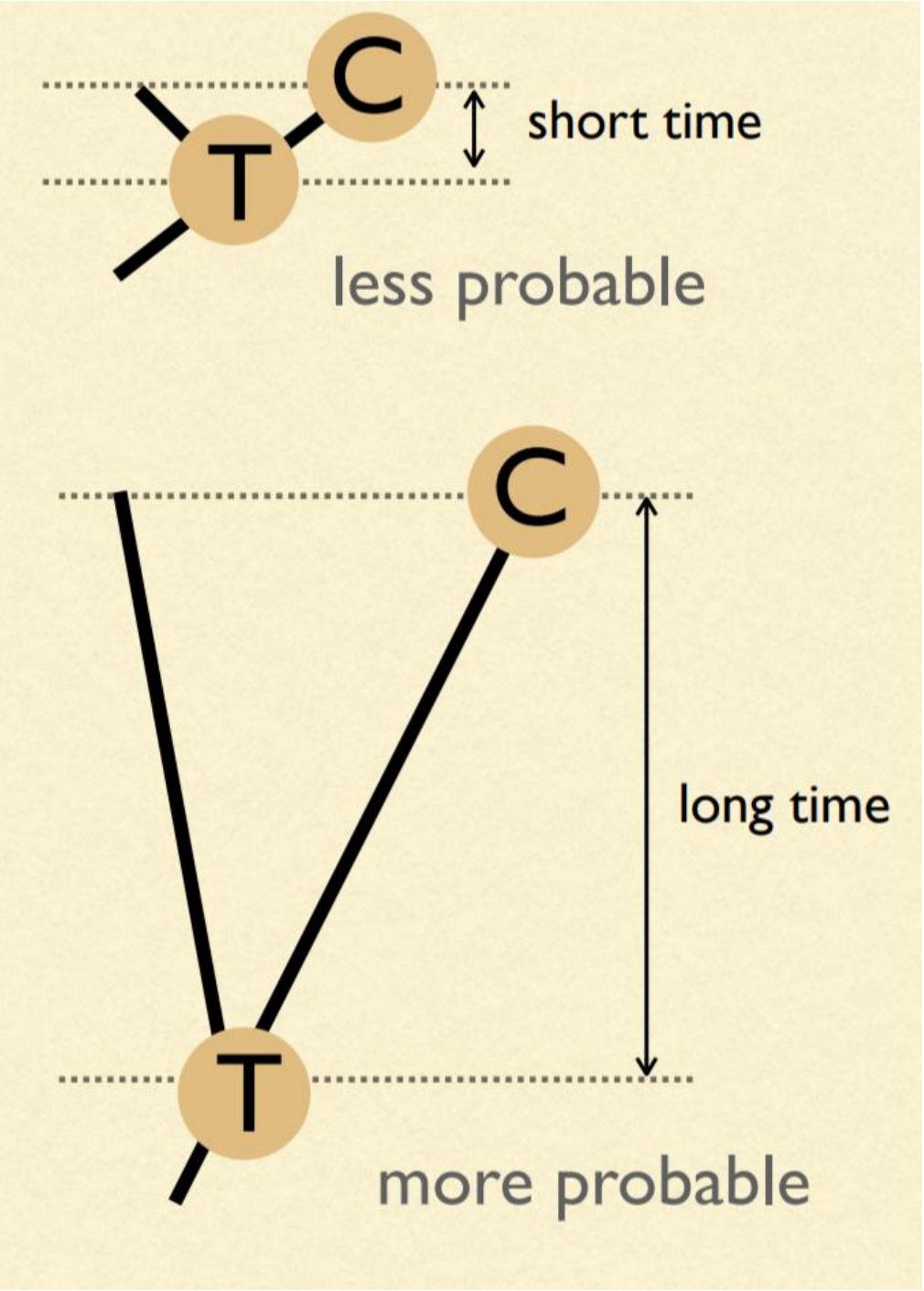
Our model of change depends on time: We must estimate branch lengths

Units of branch length will be expected number of substitutions per site

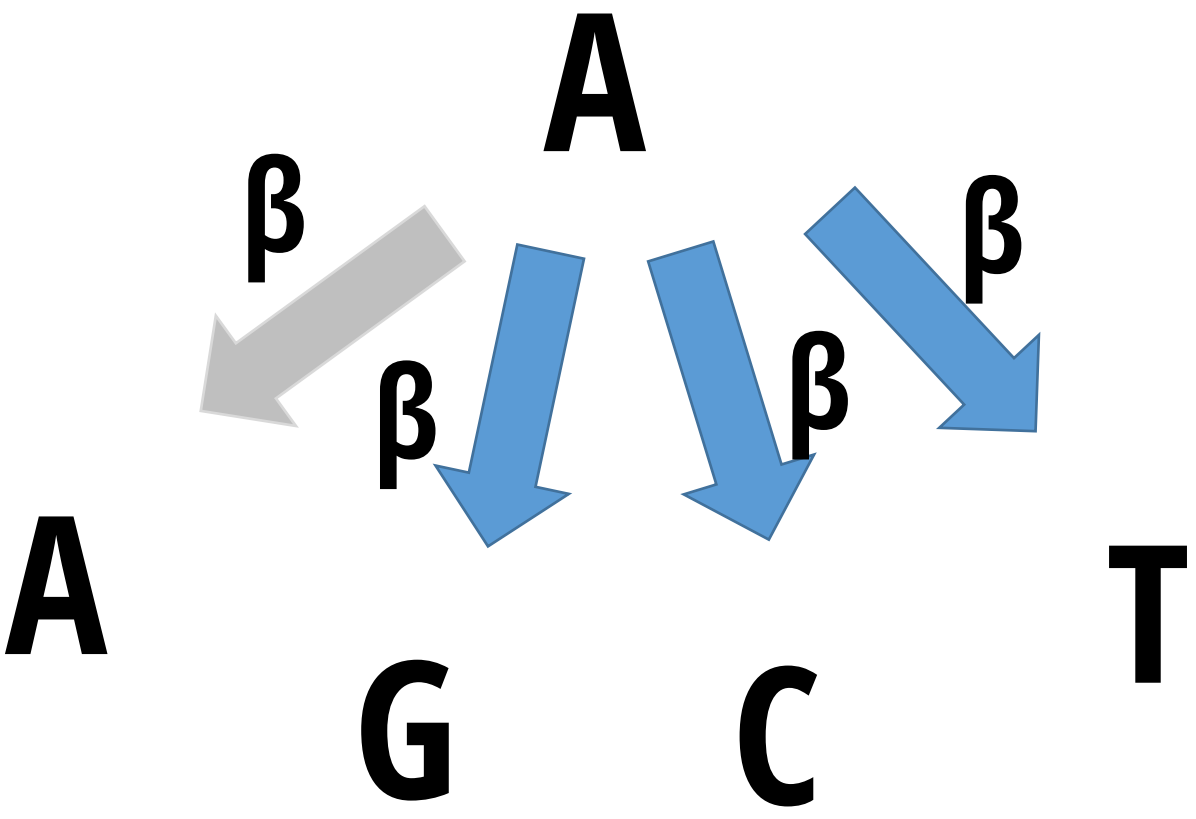
(= rate of substitions x time)



Probabilities are dependent on time



$\mu = 4\beta$



A (very) simple phylogeny…

P_{AA} =

Probability nothing happened +

Probability something happened, but that the last thing that happened ended in an A

A (very) simple phylogeny...

$$P_{AA} = (e^{-\mu t}) + (1 - e^{-\mu t})(1/4)$$

Probability something doesn't happen

Probability at least one thing happens

Probability that the last thing that happened ends in an A

A (very) simple phylogeny…

P_{AG} =

$(1 - e^{-\mu t})(1/4)$

Probability at least one thing happens

Probability that the last thing that happened ends in an G

A (very) simple phylogeny…

P_{AC} =

$(1 - e^{-\mu t})(1/4)$

Probability at least one thing happens

Probability that the last thing that happened ends in an C

A (very) simple phylogeny…

$P_{AT} = (1 - e^{-\mu t})(1/4)$

Probability at least one thing happens

Probability that the last thing that happened ends in an T

One last bit...substitutions vs. "events"

$$\nu = (3/4)\mu t = 3\beta t$$

$$4\nu/3 = \mu t$$

Only 3 out of 4 events results in a substitution. Thus, we can define the expected number of substitutions for a given time interval t as ν.

$$P_{AA} = (e^{-\mu t}) + (1 - e^{-\mu t})(1/4)$$

$$P_{AG} = (1 - e^{-\mu t})(1/4)$$

$$P_{AC} = (1 - e^{-\mu t})(1/4)$$

$$P_{AT} = (1 - e^{-\mu t})(1/4)$$

$$P_{AA} = (1/4) + (3/4)(e^{-4v/3})$$

$$P_{AG} = (1/4) - (1/4)(e^{-4v/3})$$

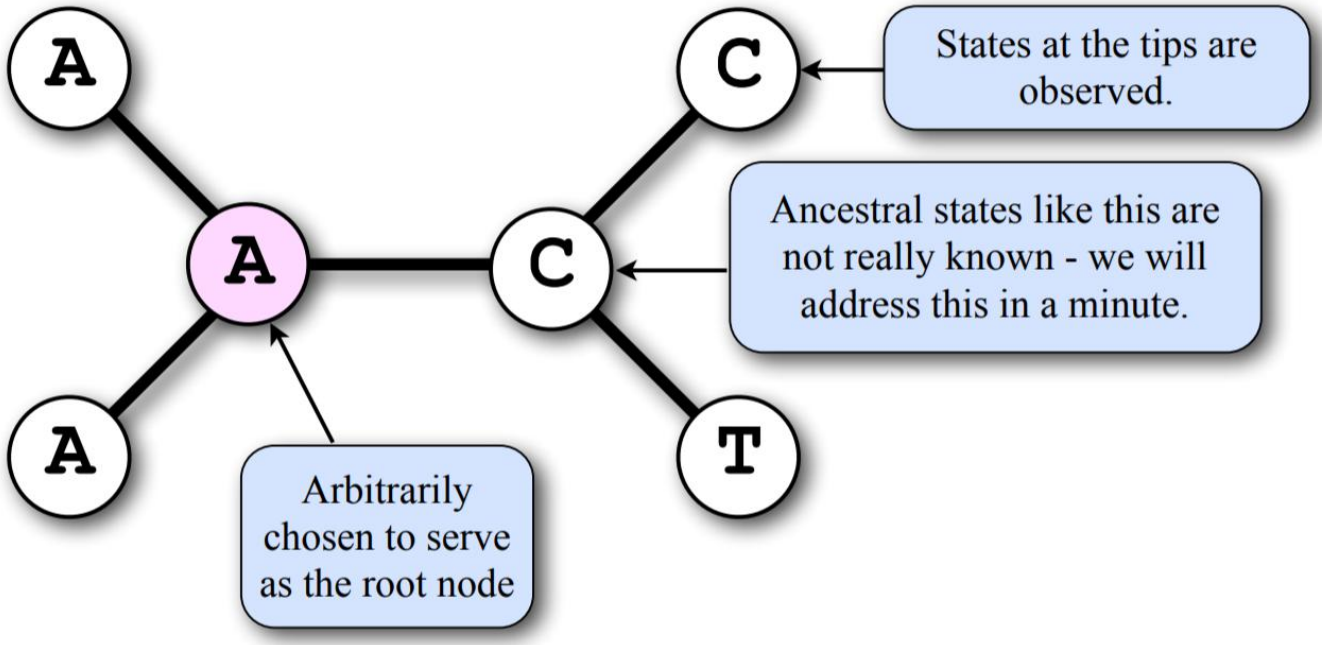
$$P_{AC} = (1/4) - (1/4)(e^{-4v/3})$$

$$P_{AT} = (1/4) - (1/4)(e^{-4v/3})$$

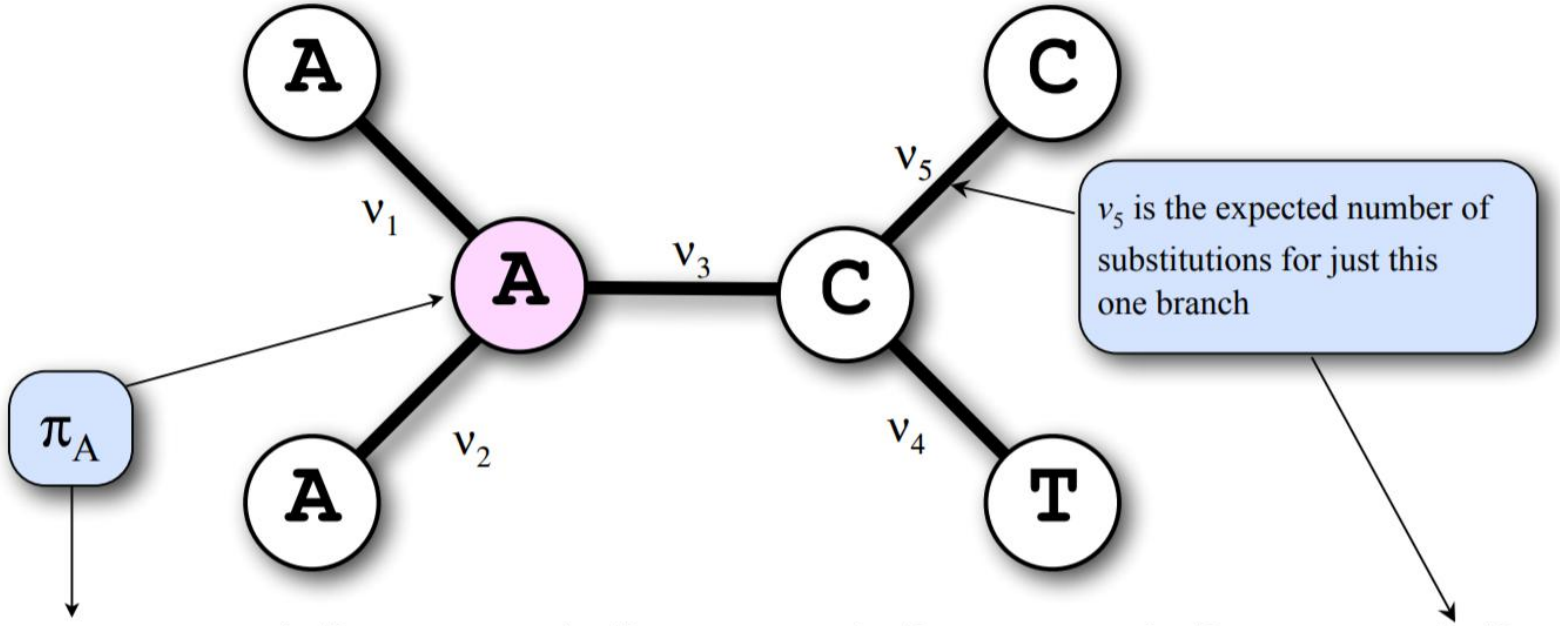
Sanity check:
Do they all add to 1?

Likelihood of an unrooted tree

(data shown for only one site)



Likelihood for site k

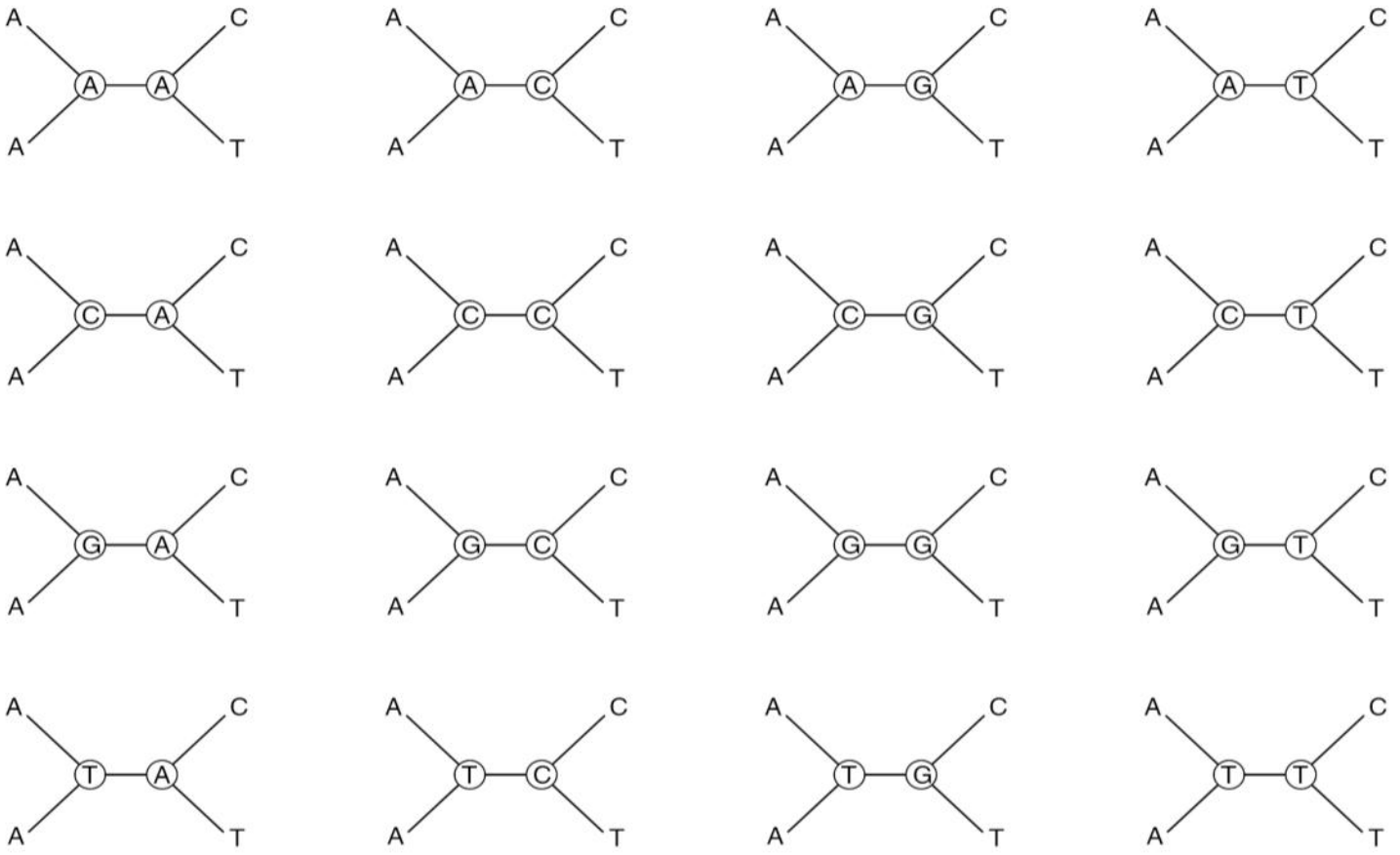


$$L_k = \tfrac{1}{4}\left[\tfrac{1}{4}+\tfrac{3}{4}e^{-4\nu_1/3}\right]\left[\tfrac{1}{4}+\tfrac{3}{4}e^{-4\nu_2/3}\right]\left[\tfrac{1}{4}-\tfrac{1}{4}e^{-4\nu_3/3}\right]\left[\tfrac{1}{4}-\tfrac{1}{4}e^{-4\nu_4/3}\right]\left[\tfrac{1}{4}+\tfrac{3}{4}e^{-4\nu_5/3}\right]$$

$P_{AA}(v_1)$ $P_{AA}(v_2)$ $P_{AC}(v_3)$ $P_{CT}(v_4)$ $P_{CC}(v_5)$

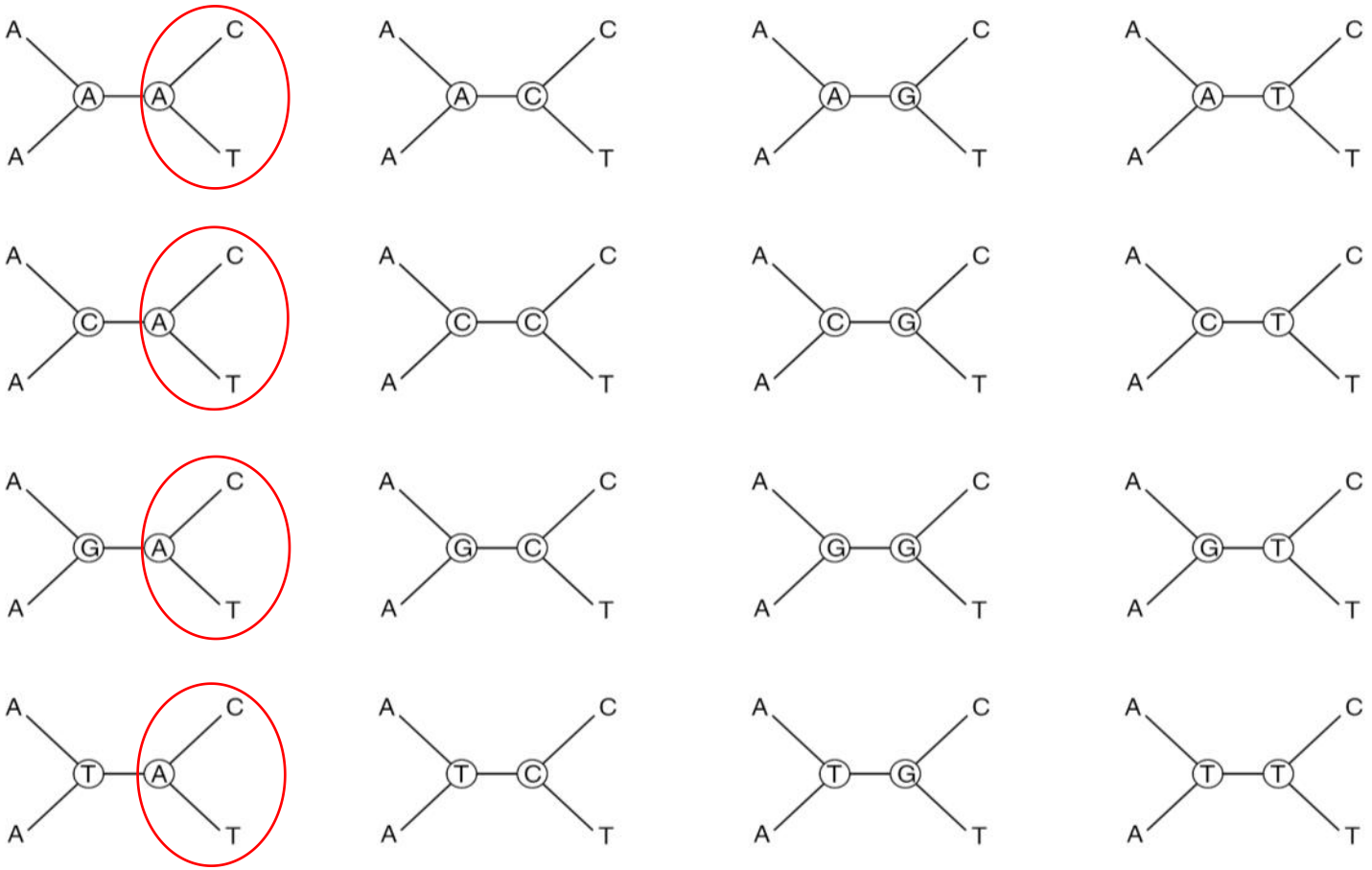
Note use of the AND probability rule

Brute force approach would be to calculate L_k for all 16 combinations of ancestral states and sum them



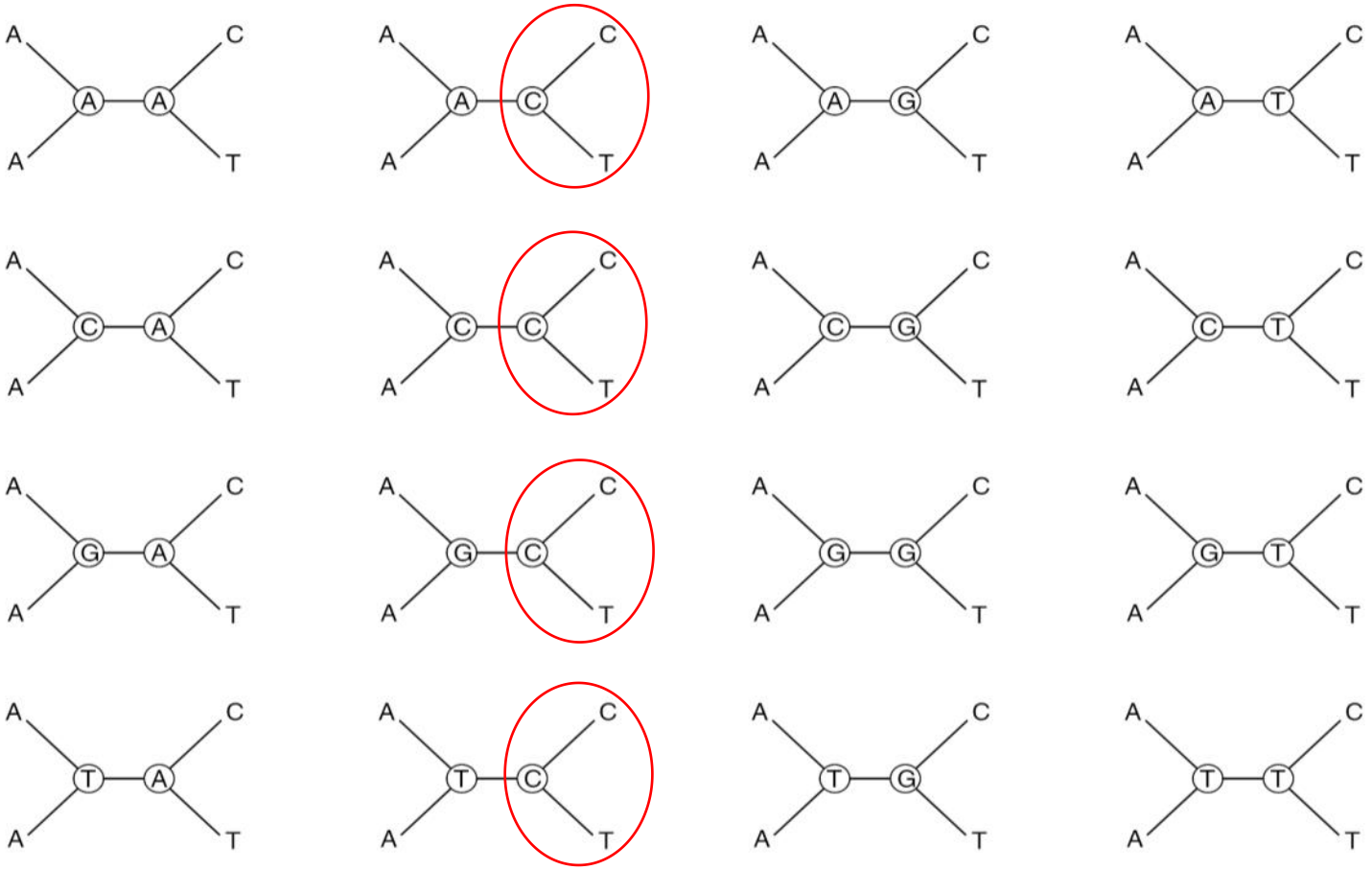
Note use of the OR probability rule

Pruning algorithm



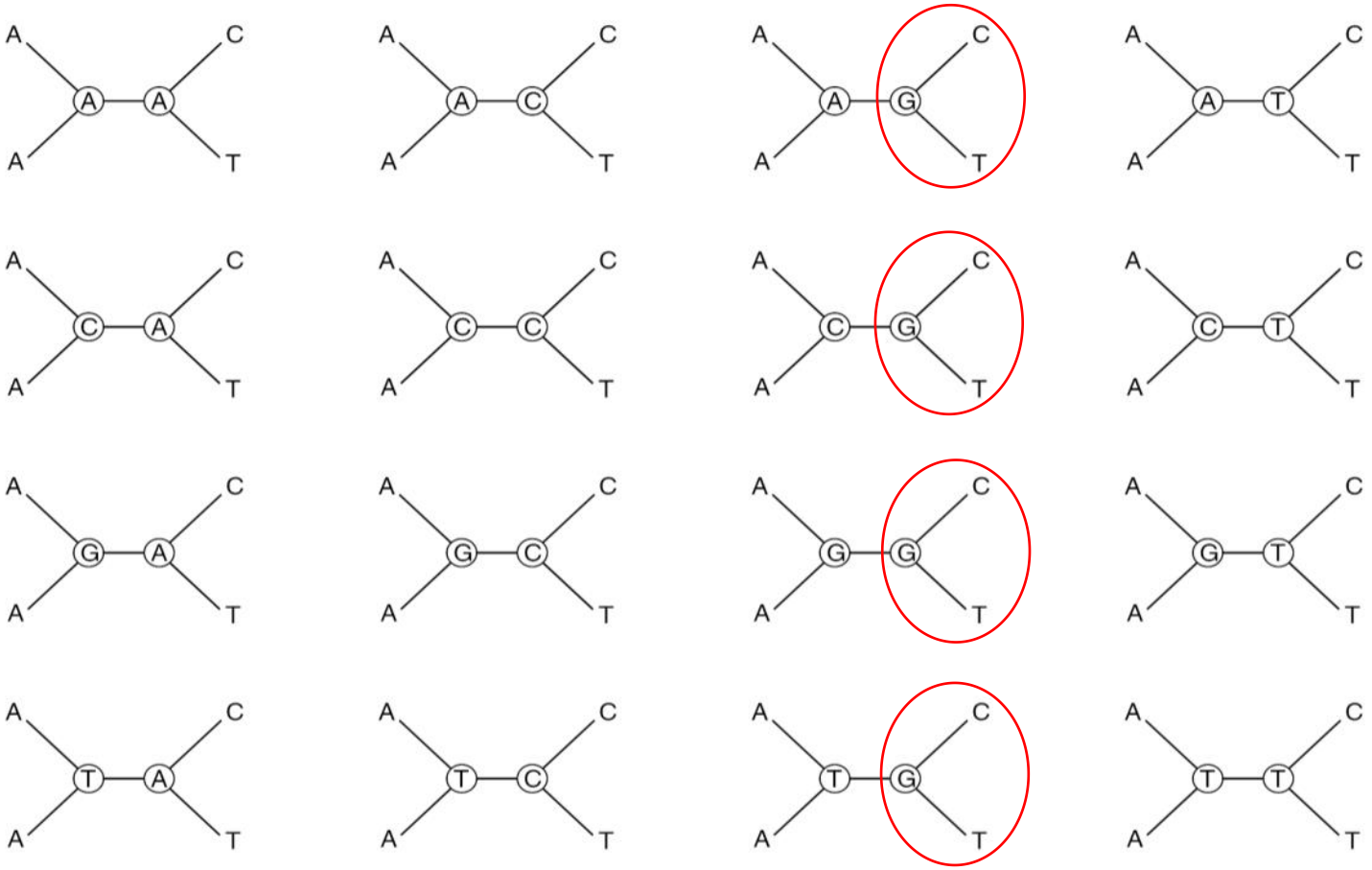
Note use of the OR probability rule

Pruning algorithm



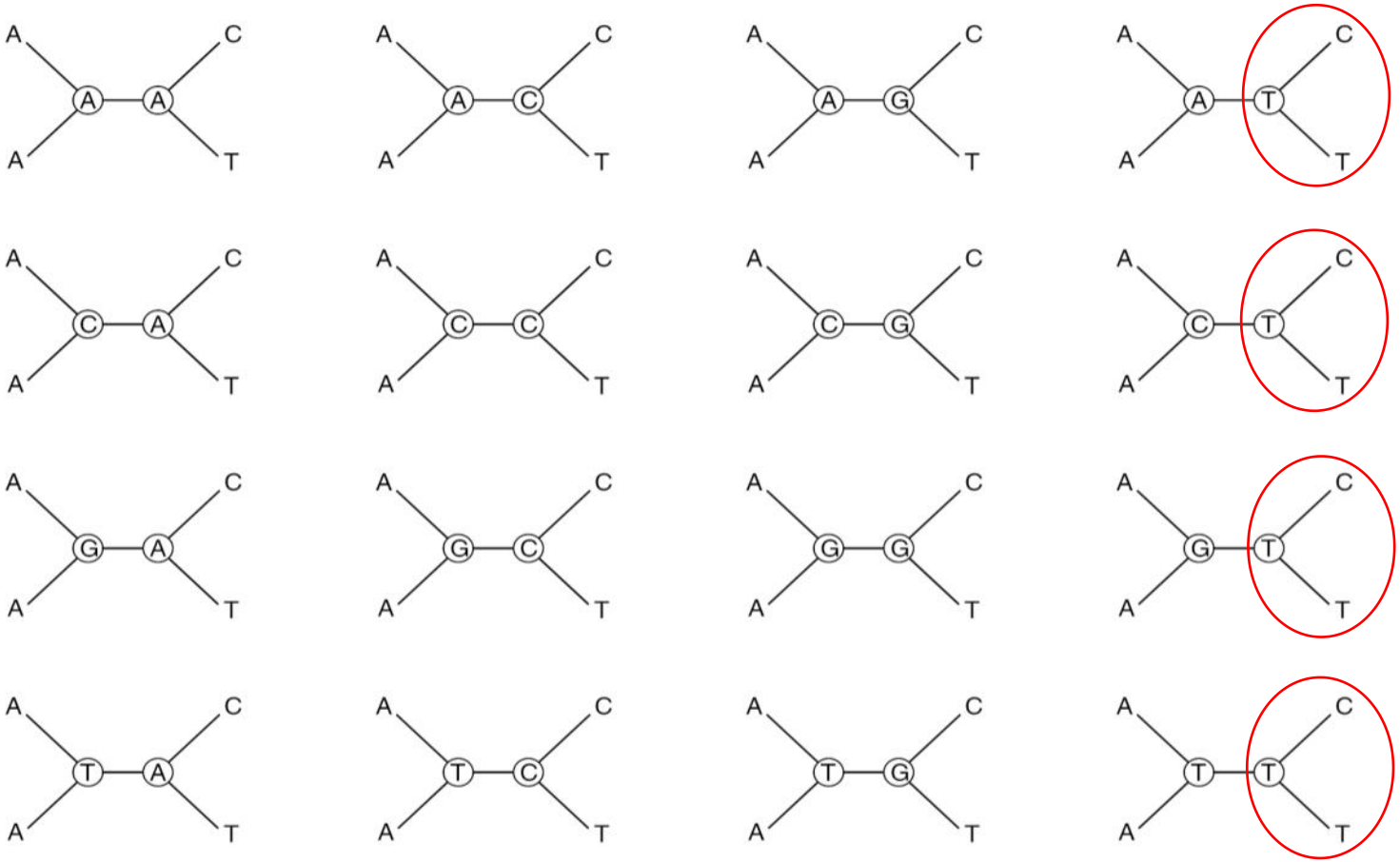
Note use of the OR probability rule

Pruning algorithm



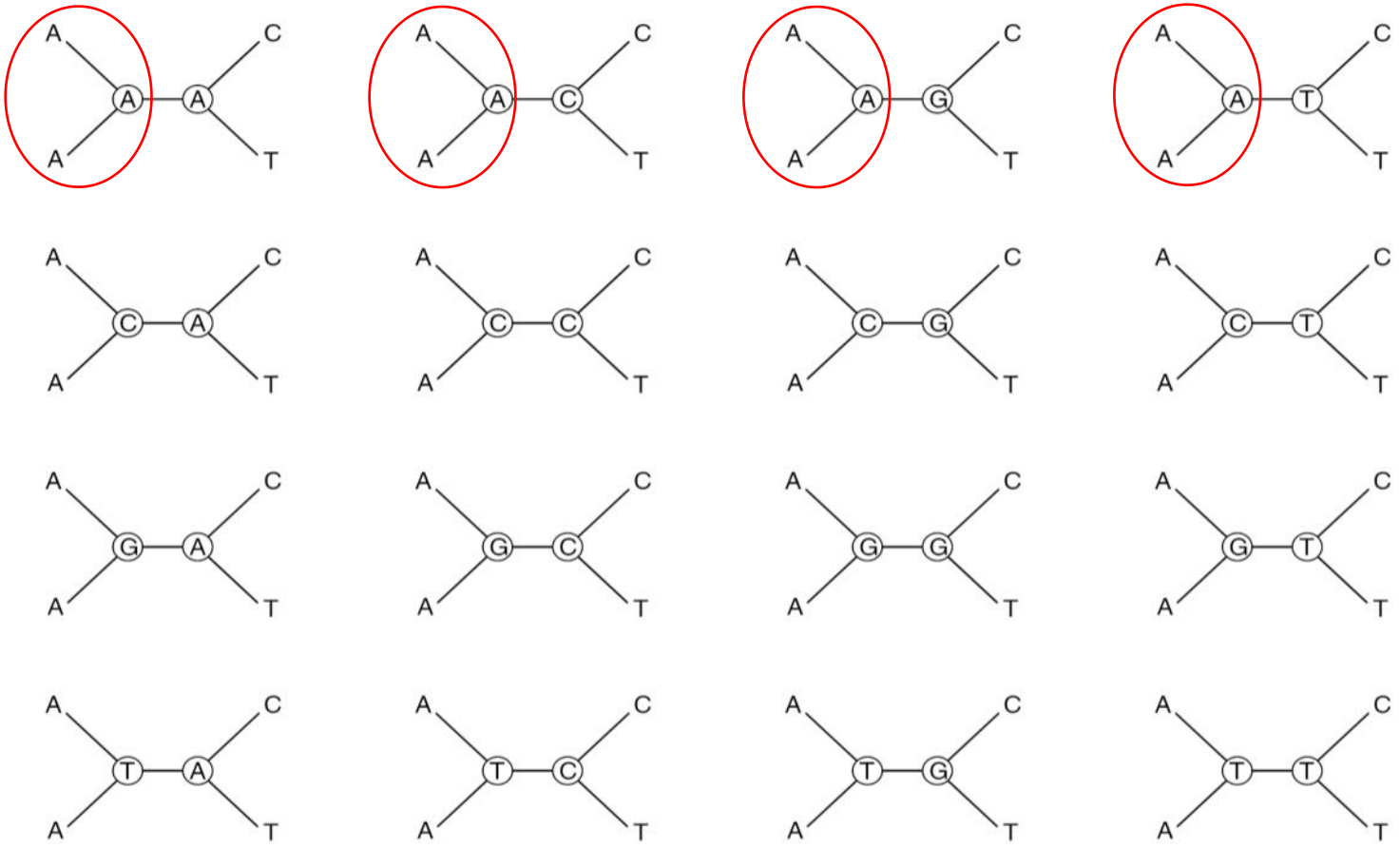
Note use of the OR probability rule

Pruning algorithm



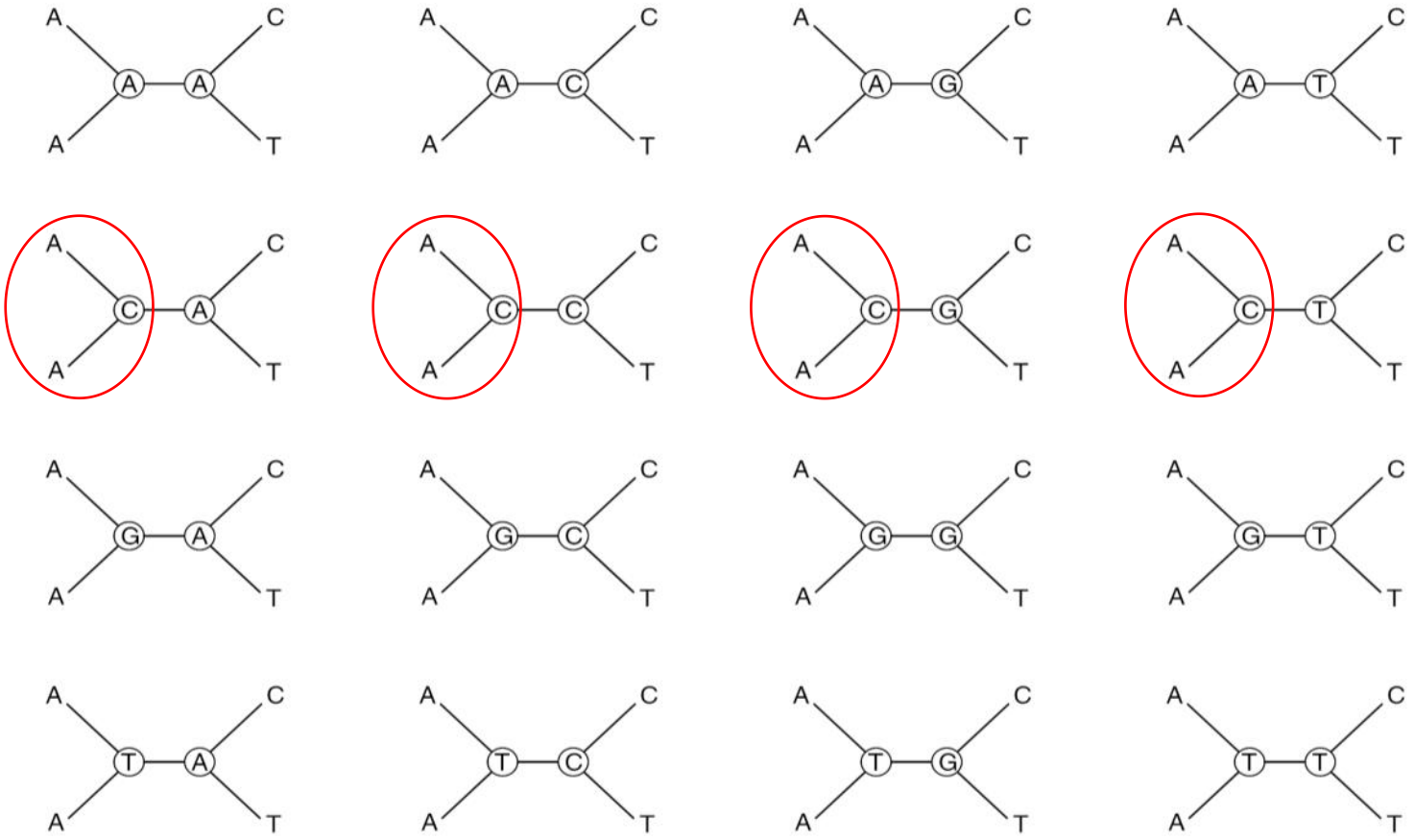
Note use of the OR probability rule

Pruning algorithm



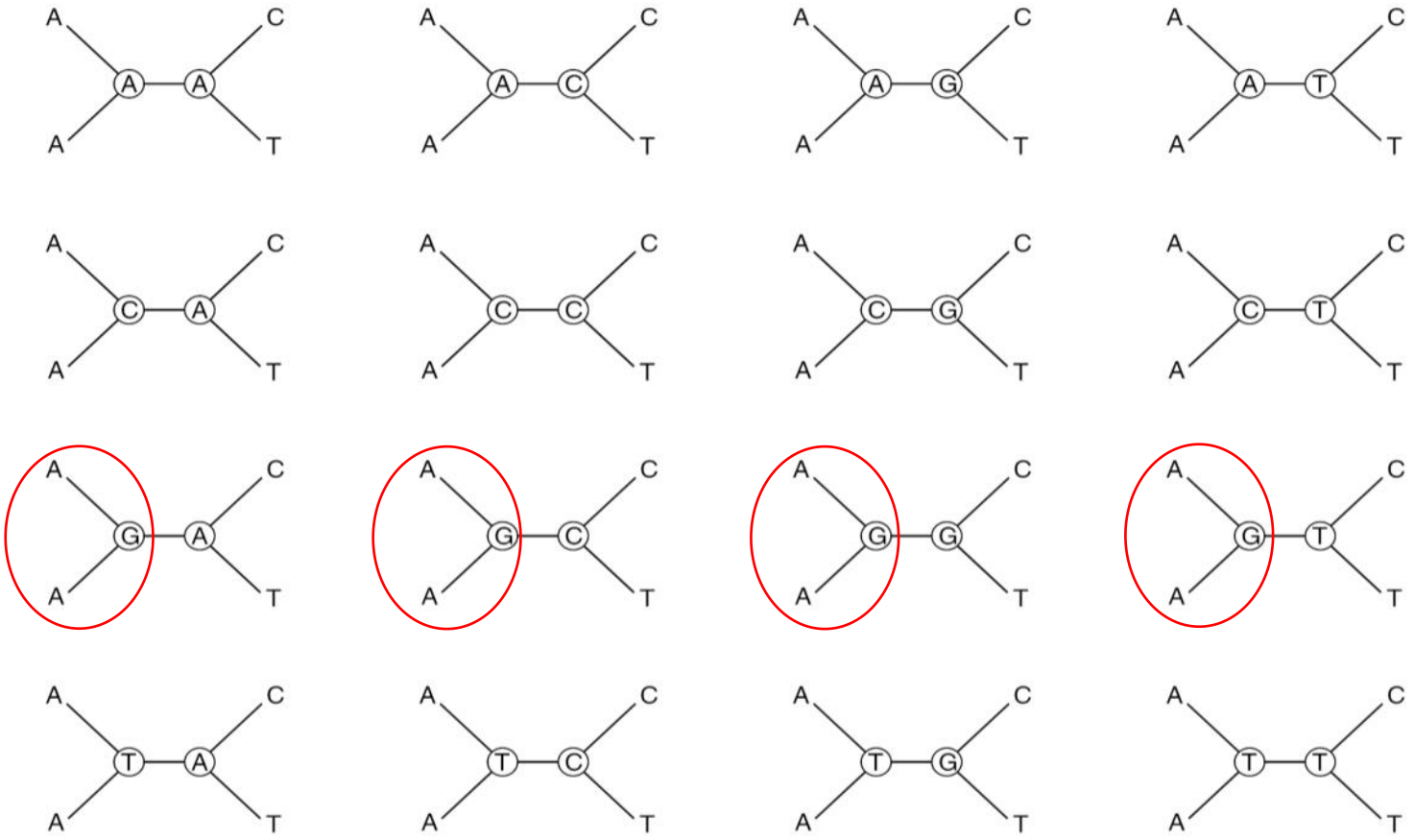
Note use of the OR probability rule

Pruning algorithm



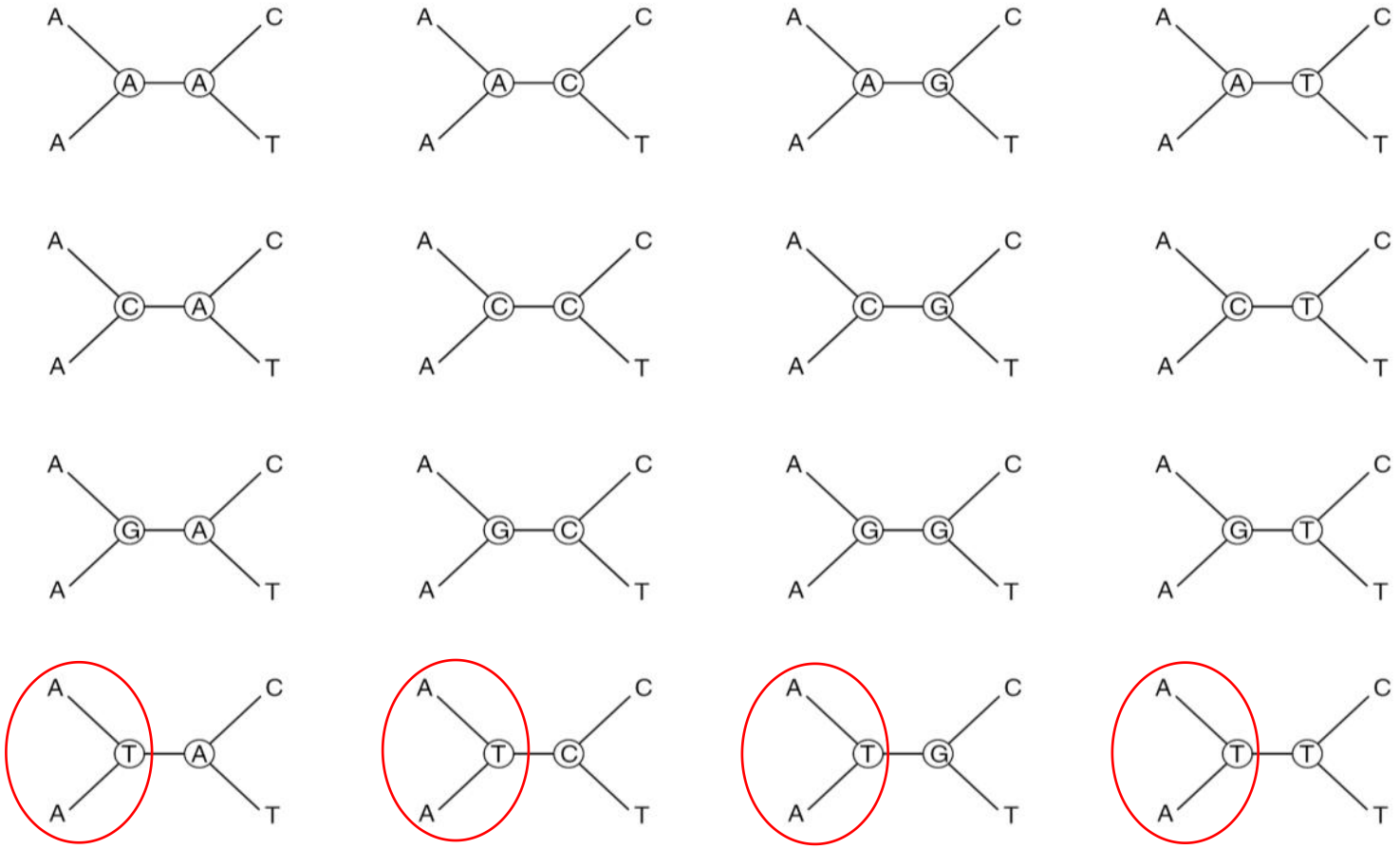
Note use of the OR probability rule

Pruning algorithm

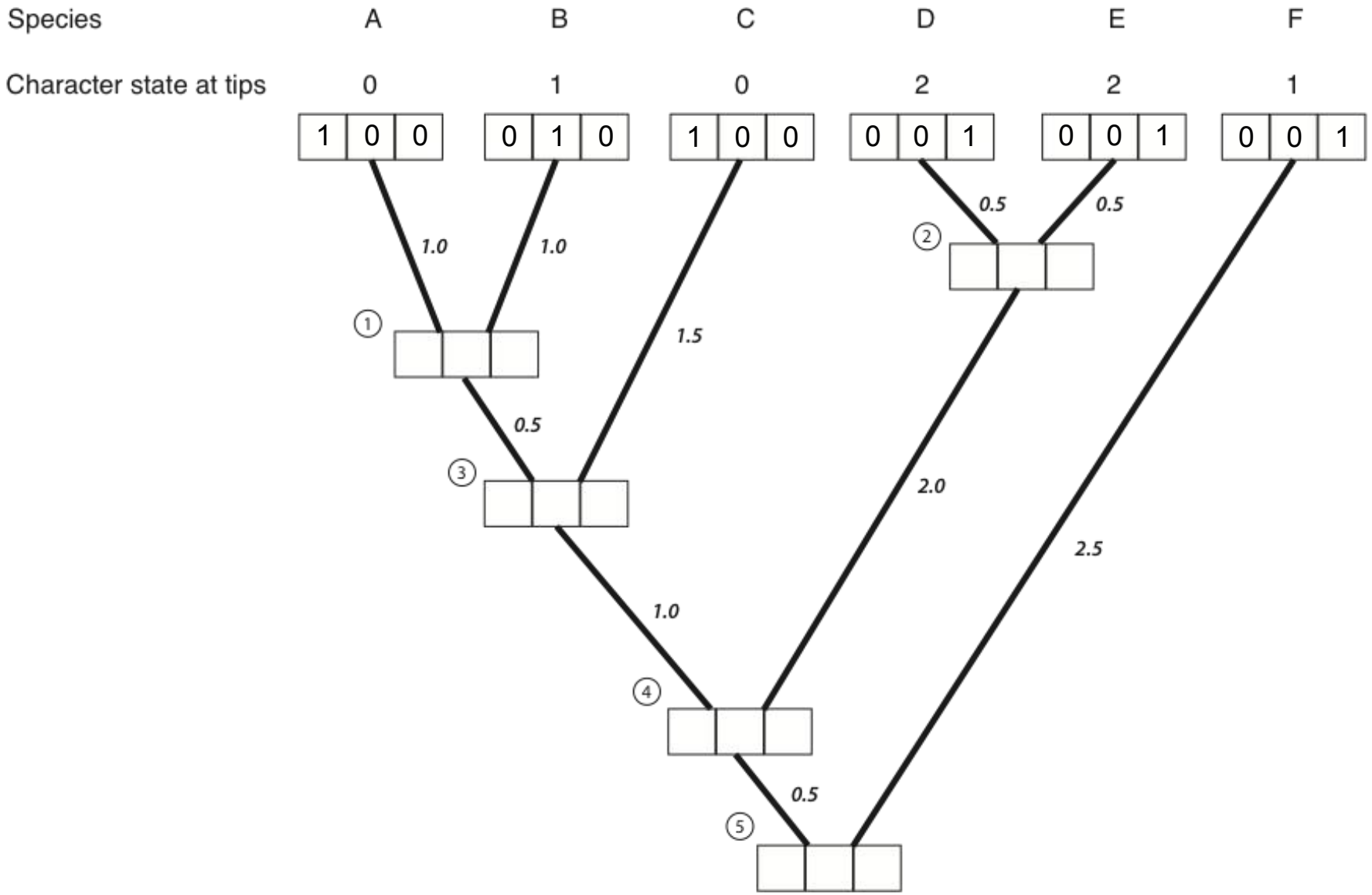


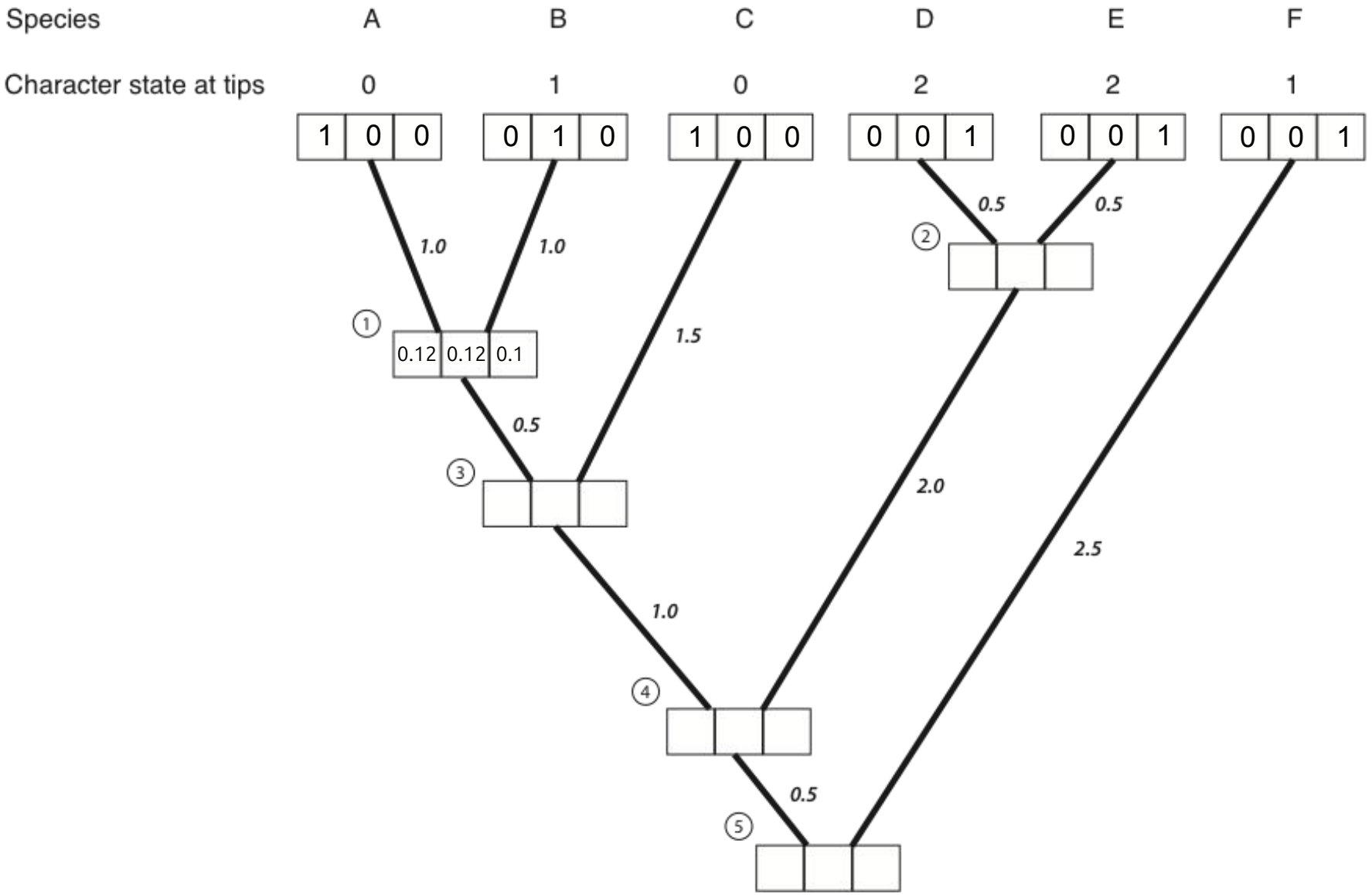
Note use of the OR probability rule

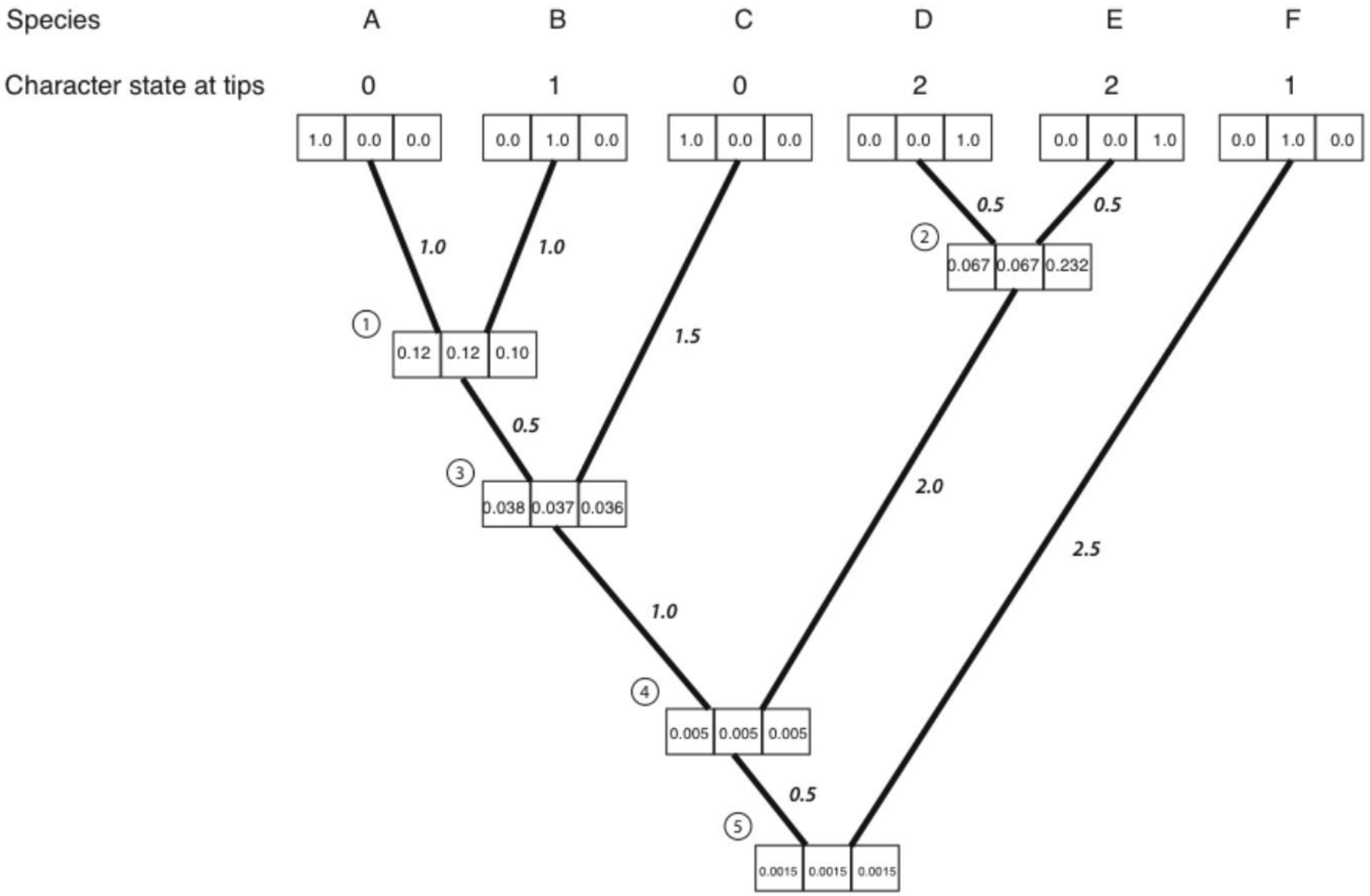
Pruning algorithm



Note use of the OR probability rule


Species
A
B
C
D
E
F
Character state at tips
0
1
0
2
2
1
1 0 0
0 1 0
1 0 0
0 0 1
0 0 1
0 0 1
1.0
1.0
0.5
0.5
1.5
0.5
2.0
2.5
1.0
0.5
1
2
3
4
5


Species
A
B
C
D
E
F
Character state at tips
0
1
0
2
2
1
1 0 0
0 1 0
1 0 0
0 0 1
0 0 1
0 0 1
1.0
1.0
0.5
0.5
1
2
0.12 0.12 0.1
1.5
0.5
3
2.0
2.5
1.0
4
0.5
5


Species
A
B
C
D
E
F
Character state at tips
0
1
0
2
2
1
1.0 0.0 0.0
0.0 1.0 0.0
1.0 0.0 0.0
0.0 0.0 1.0
0.0 0.0 1.0
0.0 1.0 0.0
1.0
1.0
0.5
0.5
1
0.12 0.12 0.10
2
0.067 0.067 0.232
1.5
0.5
3
0.038 0.037 0.036
2.0
2.5
1.0
4
0.005 0.005 0.005
0.5
5
0.0015 0.0015 0.0015

Lab Exercise in R

Goal:

Get some experience R

Understand probability, likelihoods, and Bayesian inference using non-phylogenetic models (linear regression)