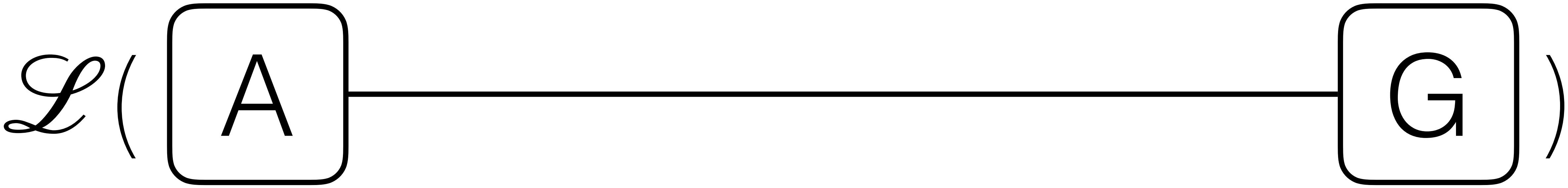
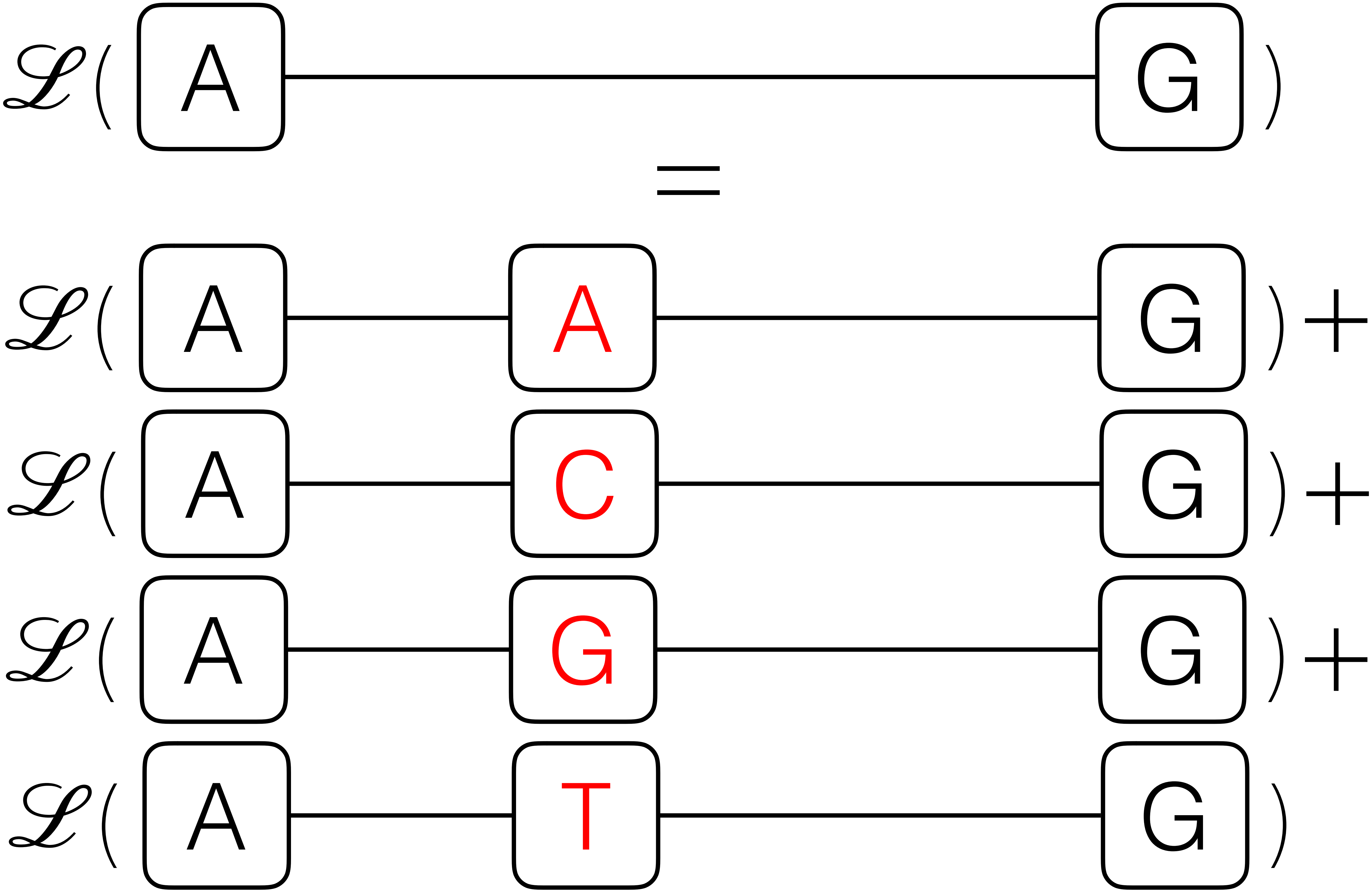
Calculating Likelihoods on Trees

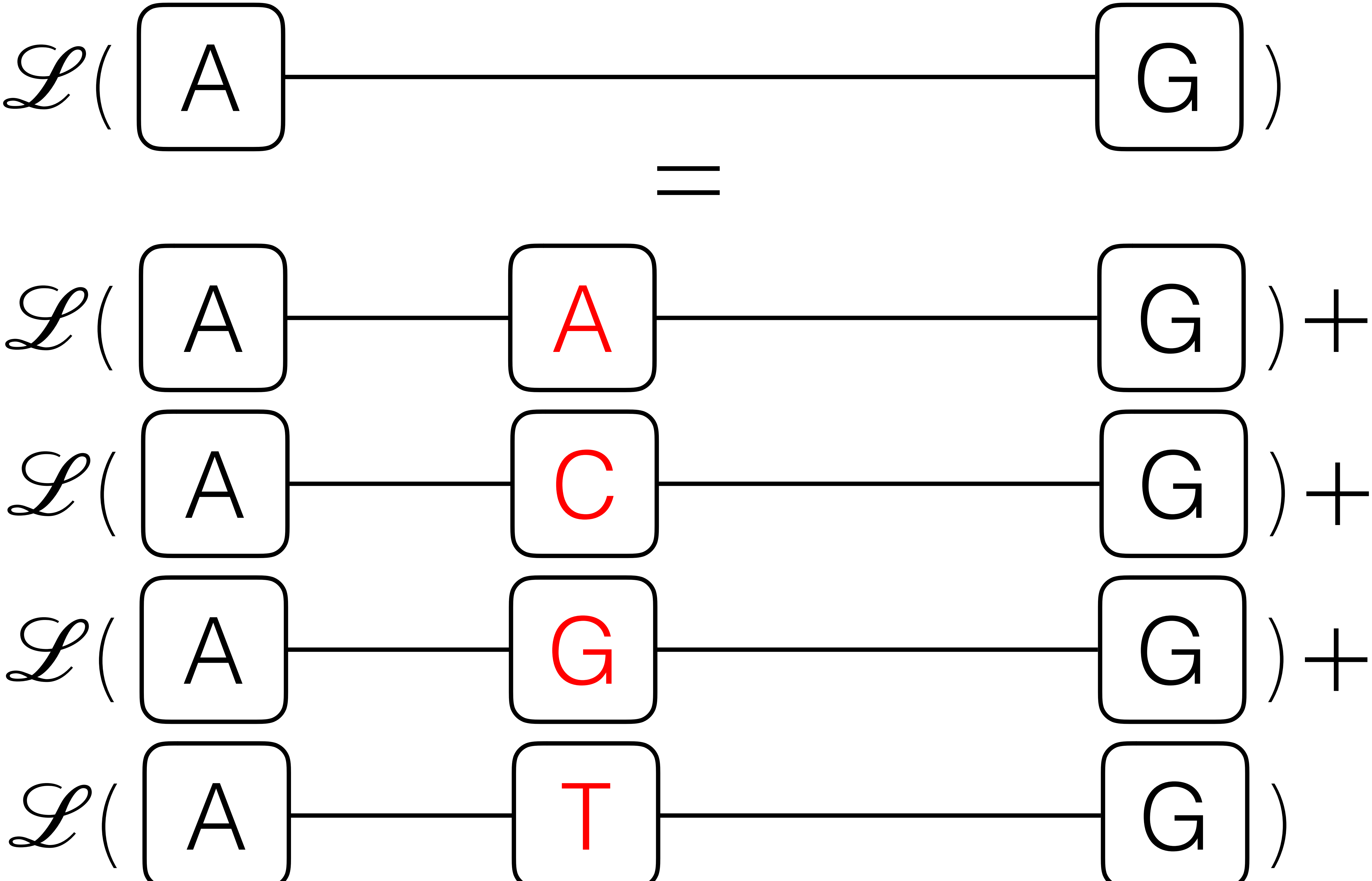
Revisiting a Single Branch



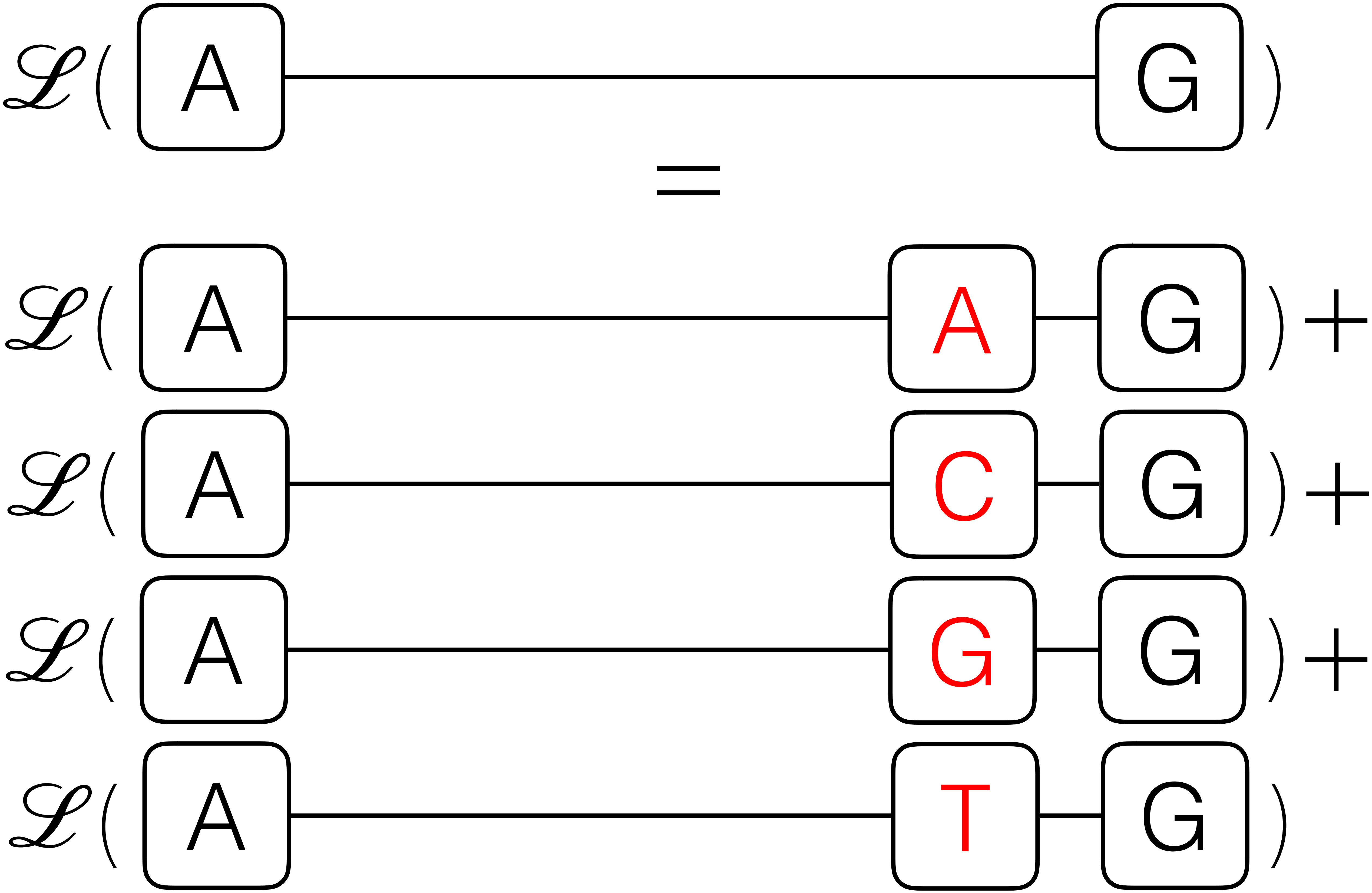
Revisiting a Single Branch



Chapman-Kolmogorov Equation



Pulley Principle (Reversibility)



Pulley Principle (Reversibility)

$$\mathscr{L}\left(\boxed{A}\longrightarrow\!\!-\!\!-\!\!-\!\!-\boxed{G}\right)$$

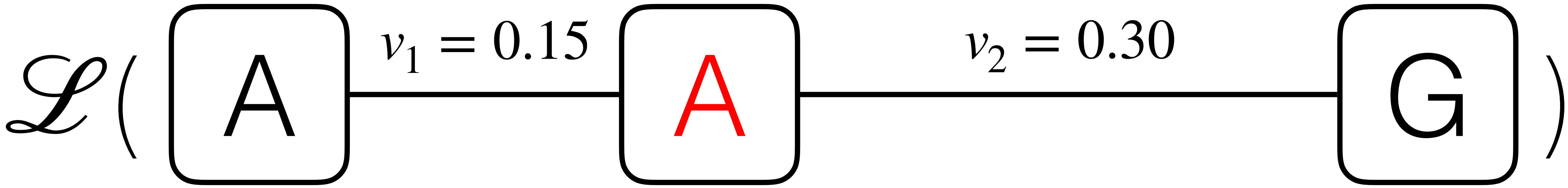
$$=$$

$$\mathscr{L}\left(\boxed{A}-\boxed{\color{red}{A}}-\!\!-\!\!-\!\!-\boxed{G}\right)+$$

$$\mathscr{L}\left(\boxed{A}-\boxed{\color{red}{C}}-\!\!-\!\!-\!\!-\boxed{G}\right)+$$

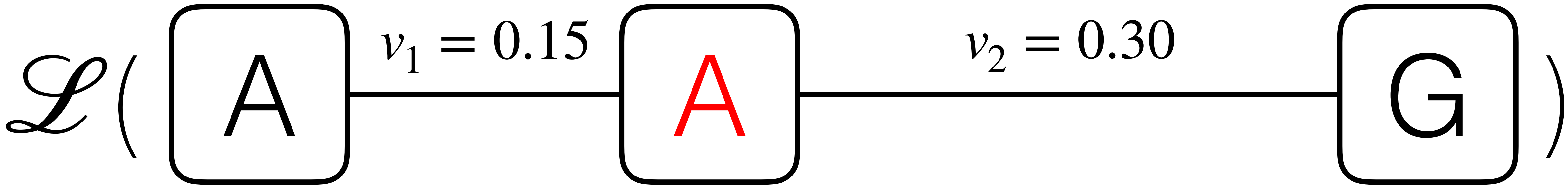
$$\mathscr{L}\left(\boxed{A}-\boxed{\color{red}{G}}-\!\!-\!\!-\!\!-\boxed{G}\right)+$$

$$\mathscr{L}\left(\boxed{A}-\boxed{\color{red}{T}}-\!\!-\!\!-\!\!-\boxed{G}\right)$$



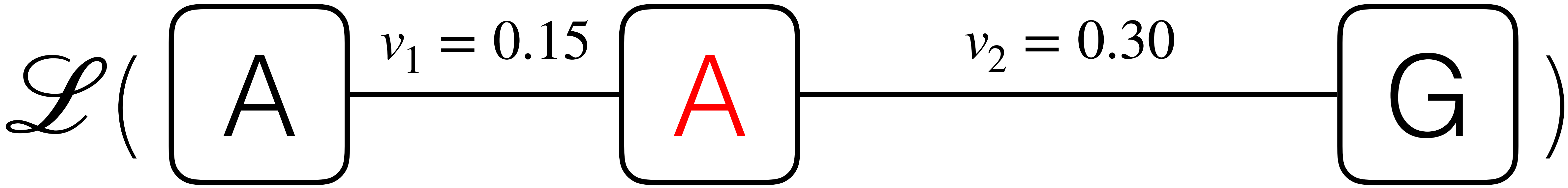
$$P_{AA}(0.15)$$

Probability of starting with an A and ending with an A on a branch of length 0.15.

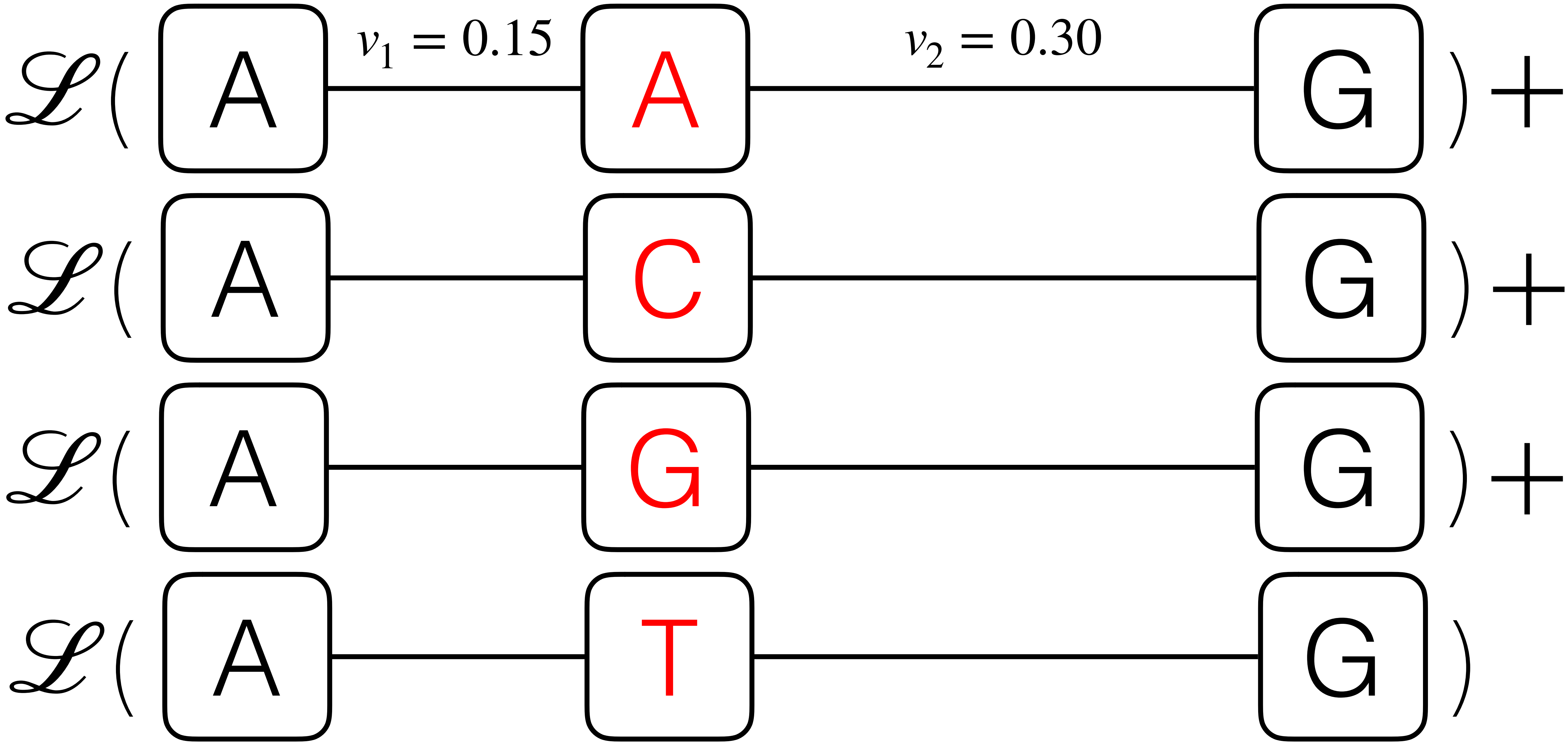


$P_{AA}(0.15)$ $P_{AG}(0.30)$

Probability of starting with an A and ending with a G on a branch of length 0.30.



$$P_{AA}(0.15) \quad * \quad P_{AG}(0.30)$$



$$
\begin{aligned}
&\big(P_{AA}(0.15) \quad * \quad P_{AG}(0.30)\big)+\\
&\big(P_{AC}(0.15) \quad * \quad P_{CG}(0.30)\big)+\\
&\big(P_{AG}(0.15) \quad * \quad P_{GG}(0.30)\big)+\\
&\big(P_{AT}(0.15) \quad * \quad P_{TG}(0.30)\big)
\end{aligned}
$$

$$\mathscr{L}\left(\boxed{A} \overset{v_2 = 0.45}{\rule[0.5ex]{12em}{0.4pt}} \boxed{G}\right)$$

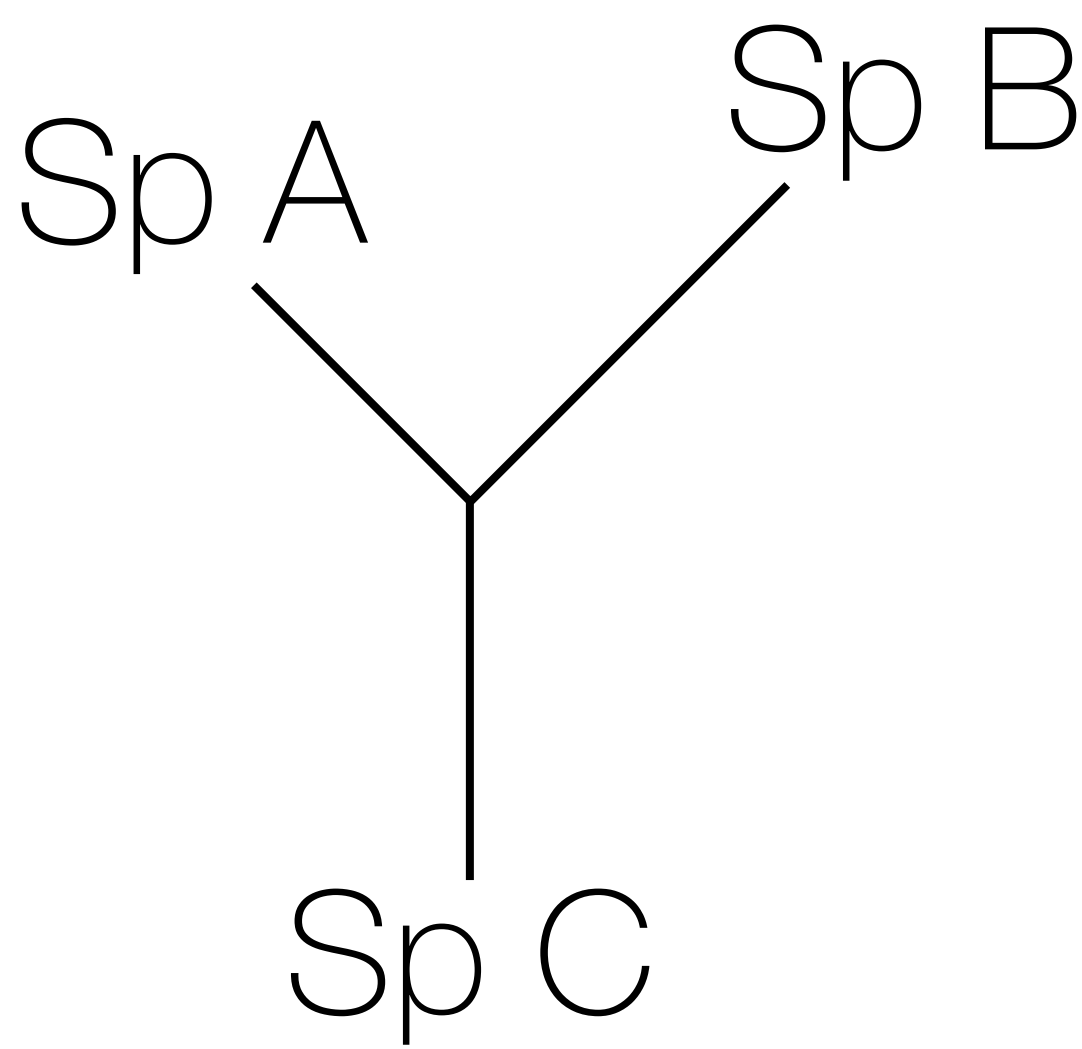
$$=$$

$$P_{AG}(0.45)$$

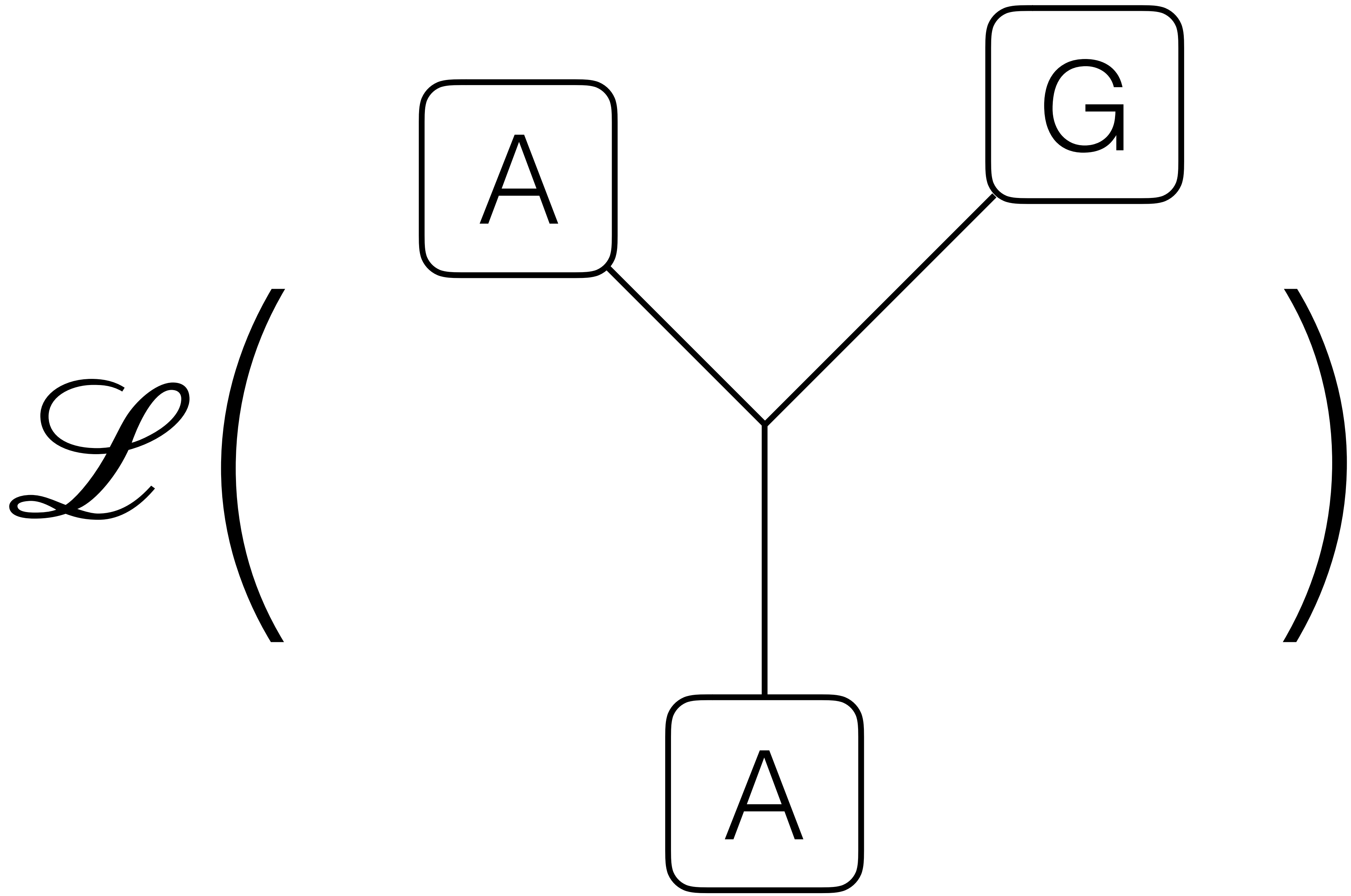
$$=$$

$$\begin{aligned}
&\Big(P_{AA}(0.15) \quad * \quad P_{AG}(0.30)\Big) + \\
&\Big(P_{AC}(0.15) \quad * \quad P_{CG}(0.30)\Big) + \\
&\Big(P_{AG}(0.15) \quad * \quad P_{GG}(0.30)\Big) + \\
&\Big(P_{AT}(0.15) \quad * \quad P_{TG}(0.30)\Big)
\end{aligned}$$

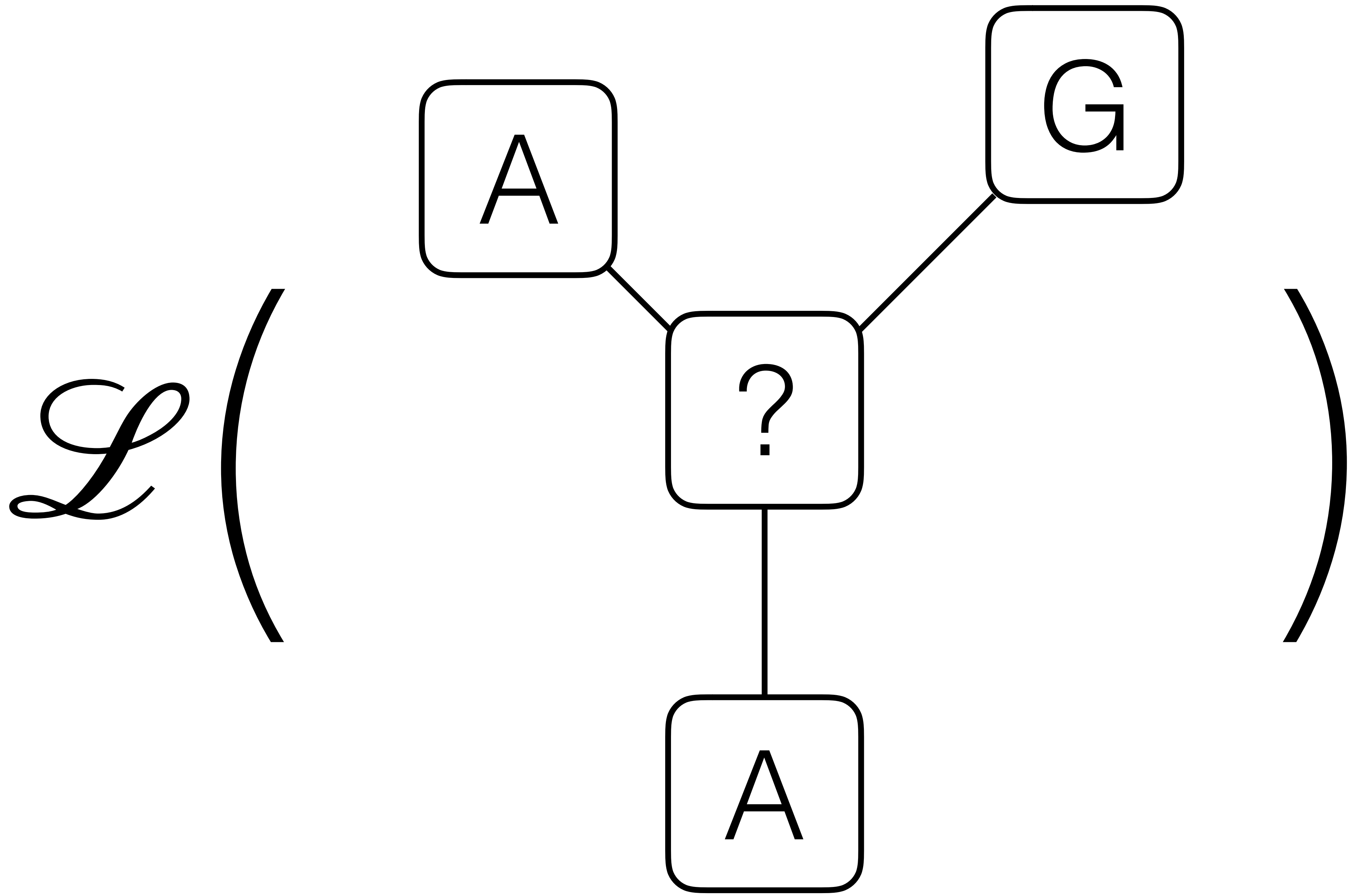
Three-Taxon Unrooted Tree



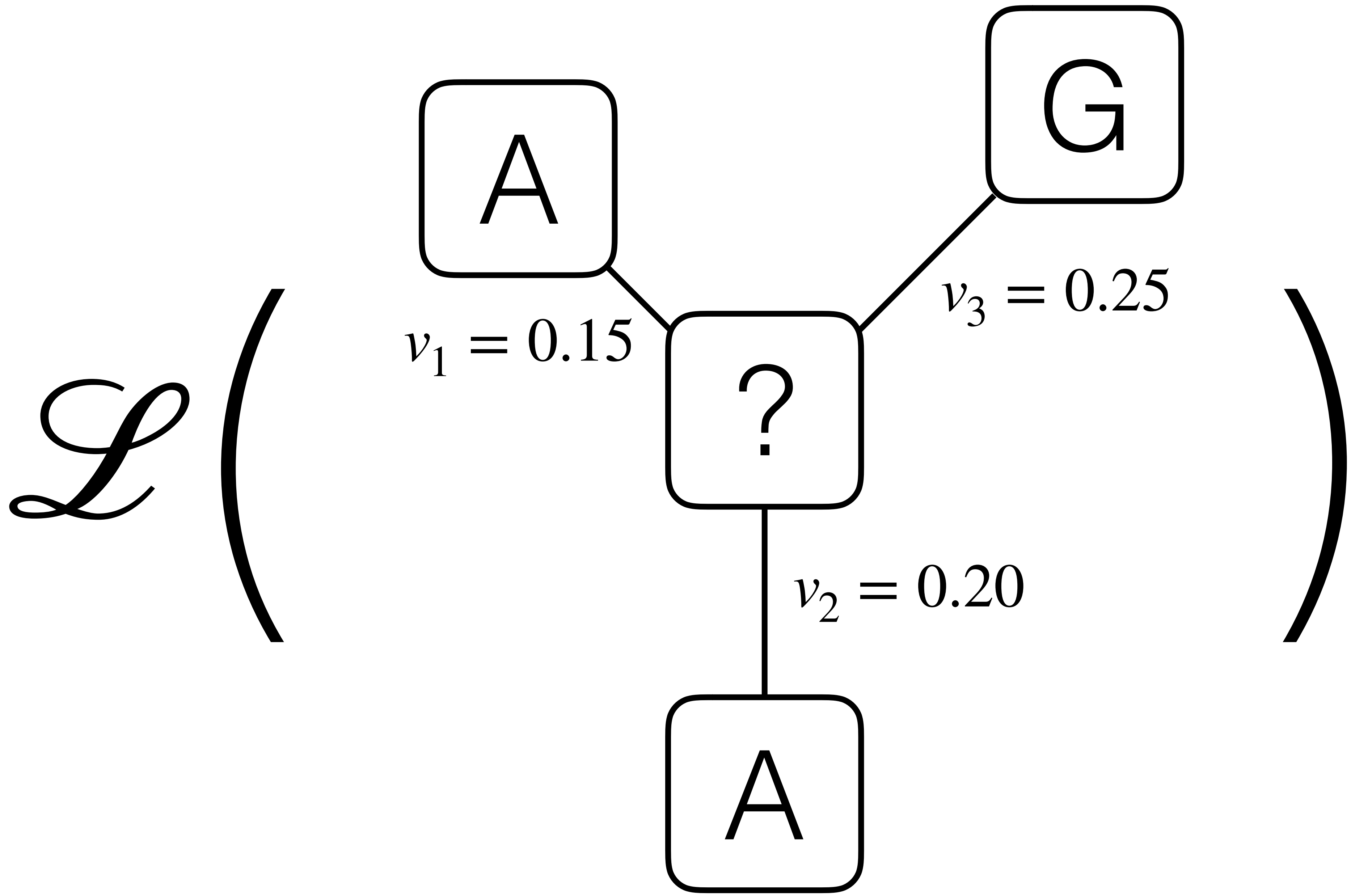
Three-Taxon Unrooted Tree

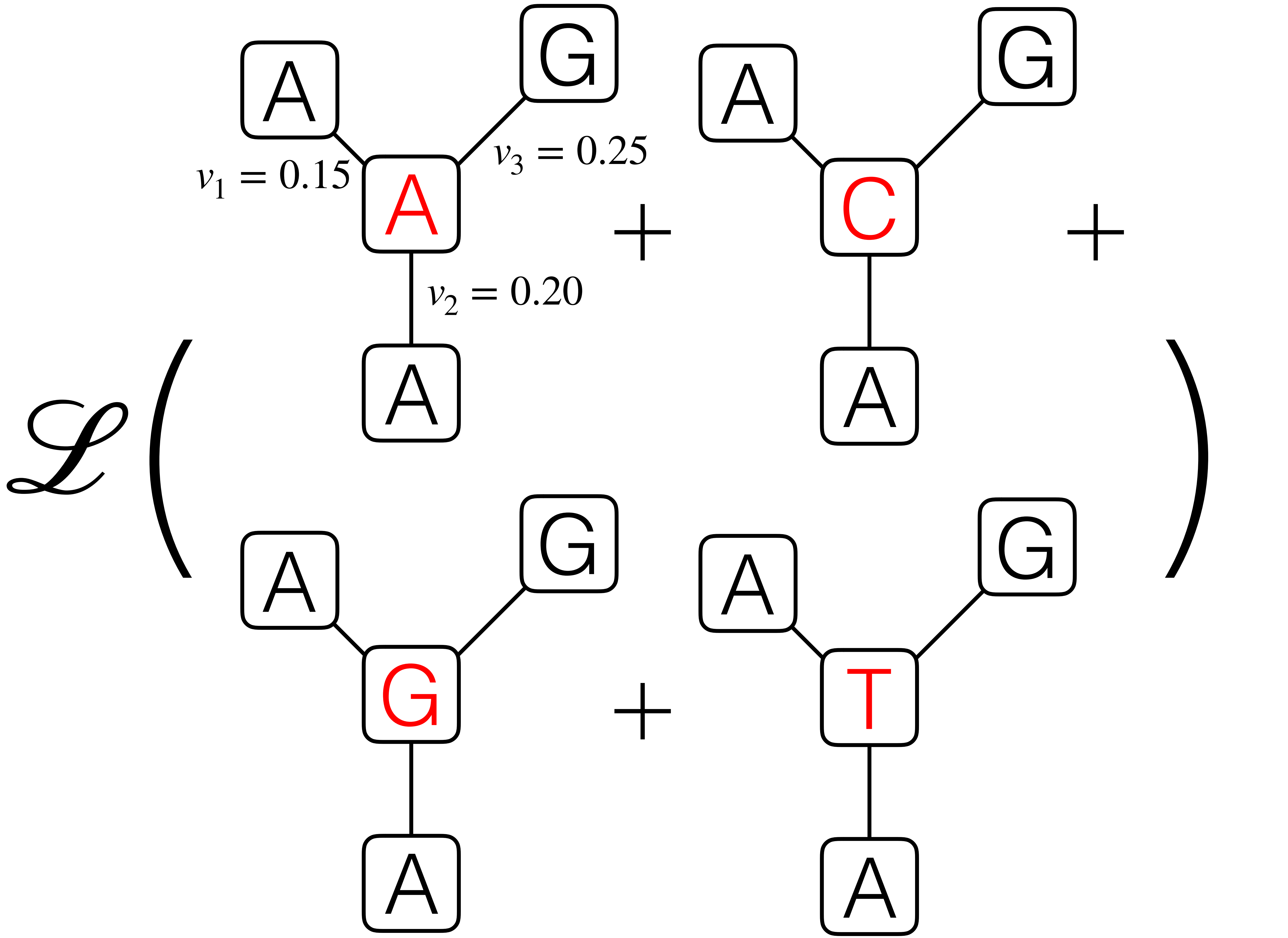


Three-Taxon Unrooted Tree

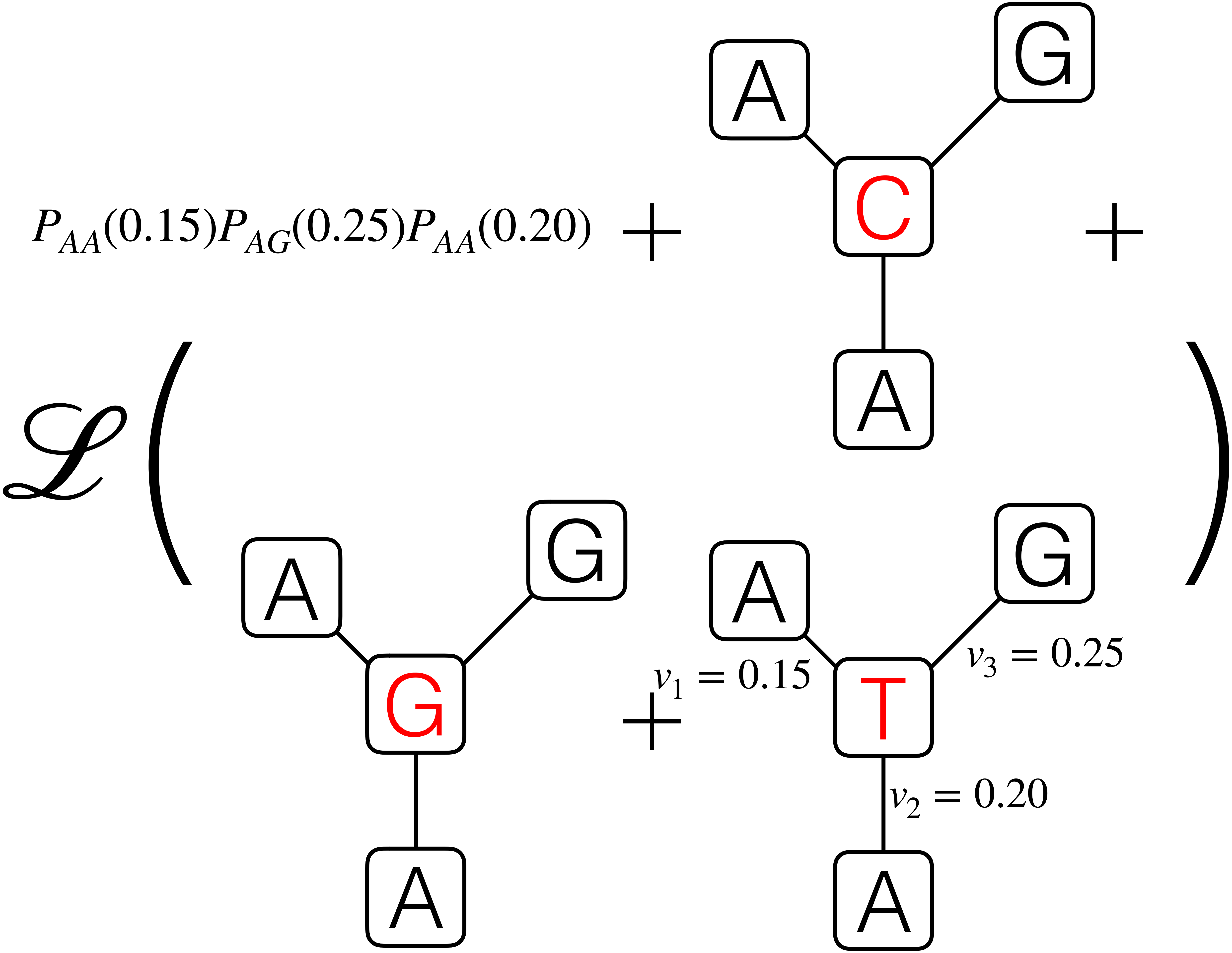


Three-Taxon Unrooted Tree



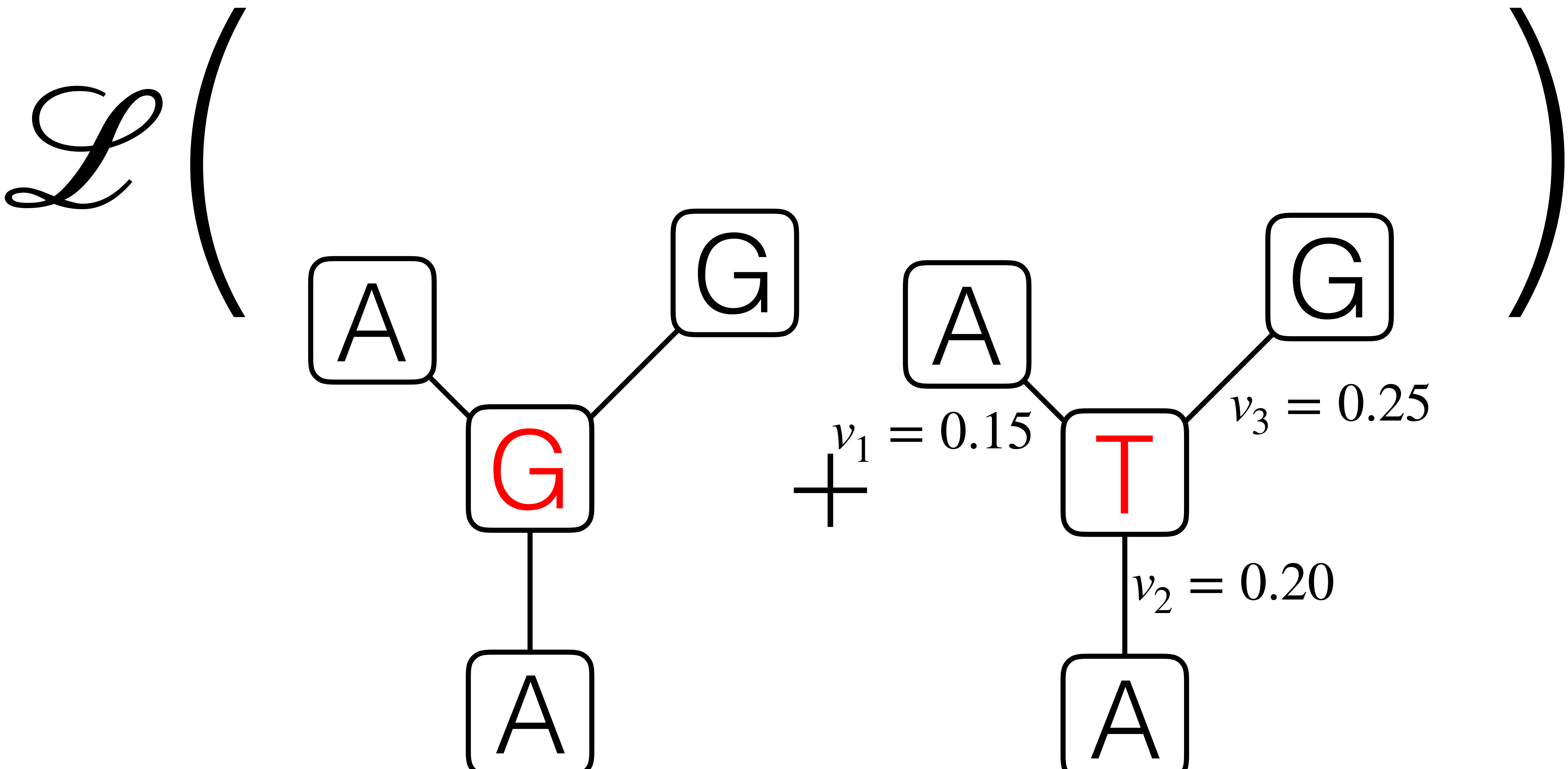


$\mathscr{L}$
A
G
A
A
$v_1 = 0.15$
$v_2 = 0.20$
$v_3 = 0.25$
A
G
C
A
A
G
G
A
A
G
T
A

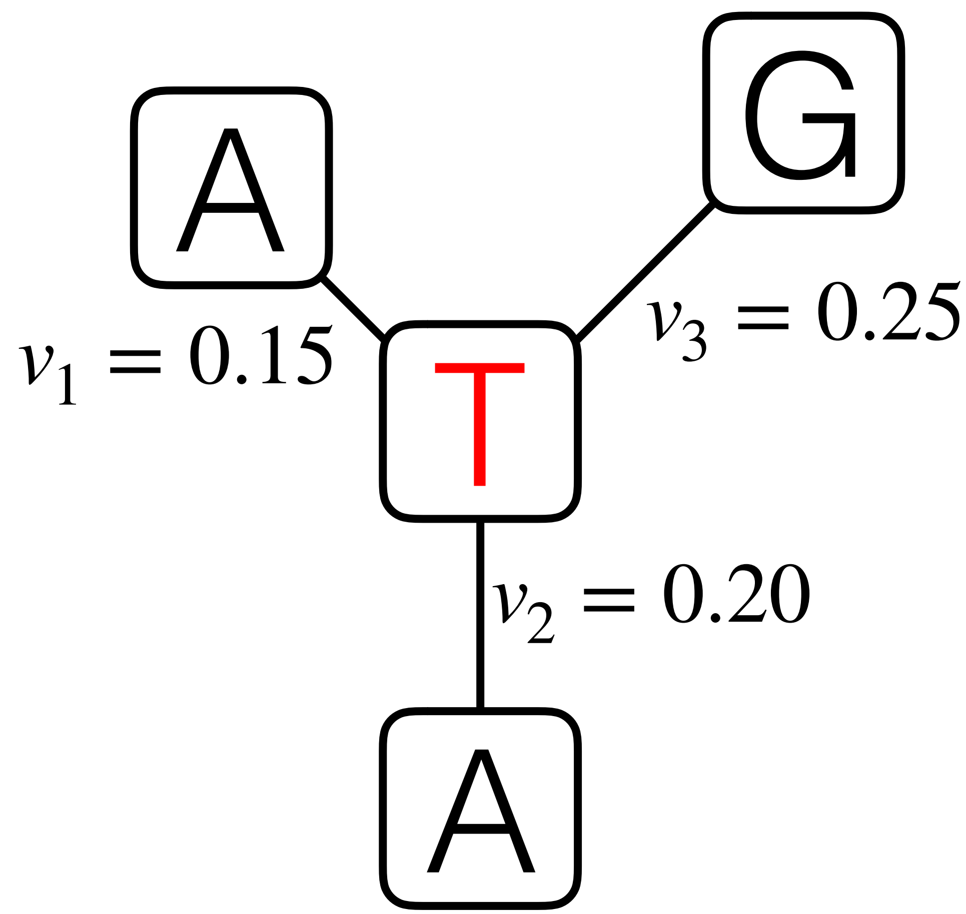


$P_{AA}(0.15)P_{AG}(0.25)P_{AA}(0.20)$
+
A
G
C
A
+
$\mathscr{L}$
A
G
G
A
+
A
G
$v_1 = 0.15$
$v_3 = 0.25$
T
$v_2 = 0.20$
A

$$P_{AA}(0.15)P_{AG}(0.25)P_{AA}(0.20) + P_{CA}(0.15)P_{CG}(0.25)P_{CA}(0.20) +$$



$$\mathscr{L}\Big(P_{AA}(0.15)P_{AG}(0.25)P_{AA}(0.20) + P_{CA}(0.15)P_{CG}(0.25)P_{CA}(0.20) + P_{GA}(0.15)P_{GG}(0.25)P_{GA}(0.20) + \Big)$$

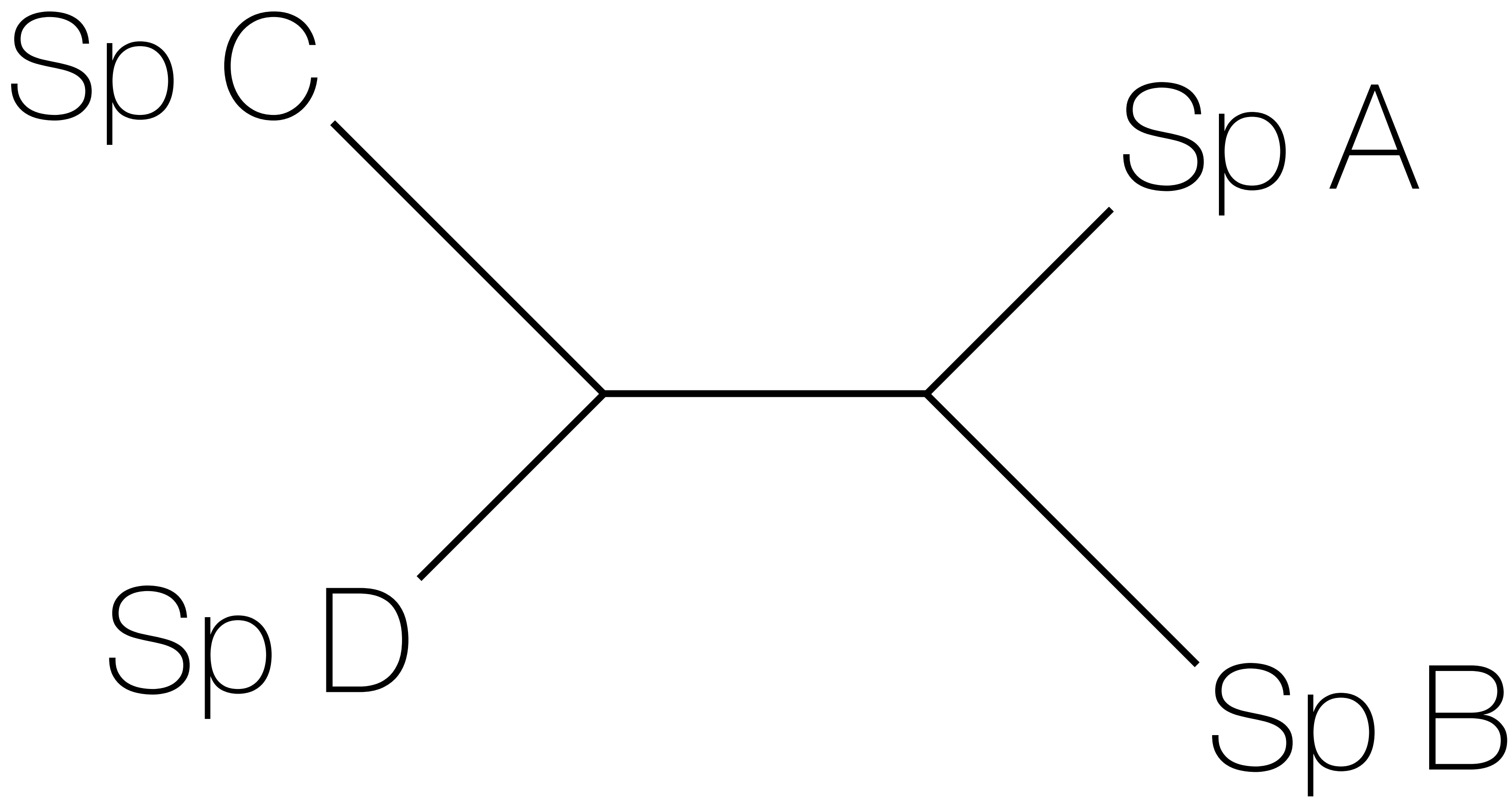


$$P_{AA}(0.15)P_{AG}(0.25)P_{AA}(0.20) + P_{CA}(0.15)P_{CG}(0.25)P_{CA}(0.20) +$$

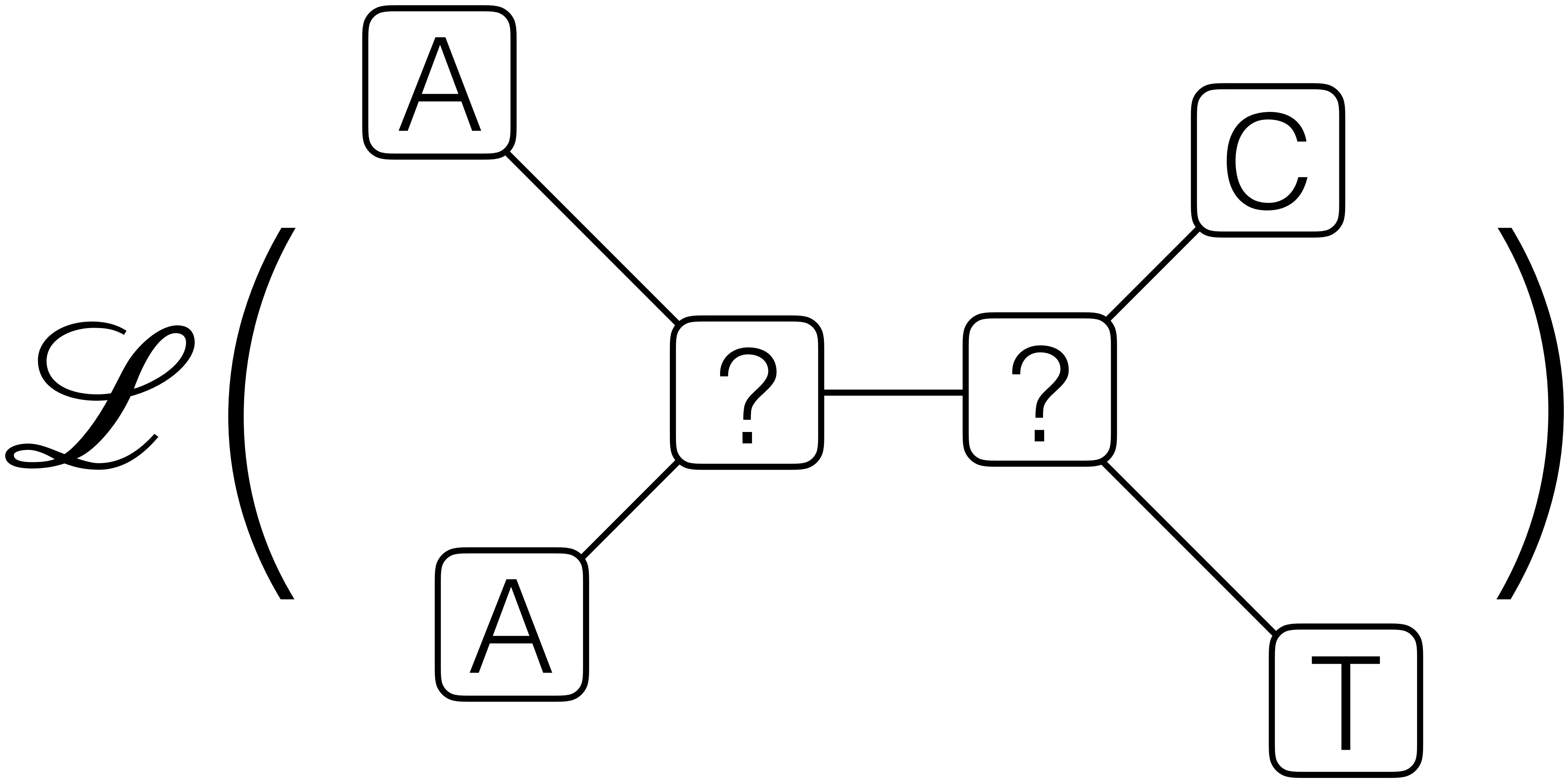
$$\mathscr{L}\Bigg(\qquad\qquad\qquad\qquad\qquad\qquad \Bigg)$$

$$P_{GA}(0.15)P_{GG}(0.25)P_{GA}(0.20) + P_{TA}(0.15)P_{TG}(0.25)P_{TA}(0.20)$$

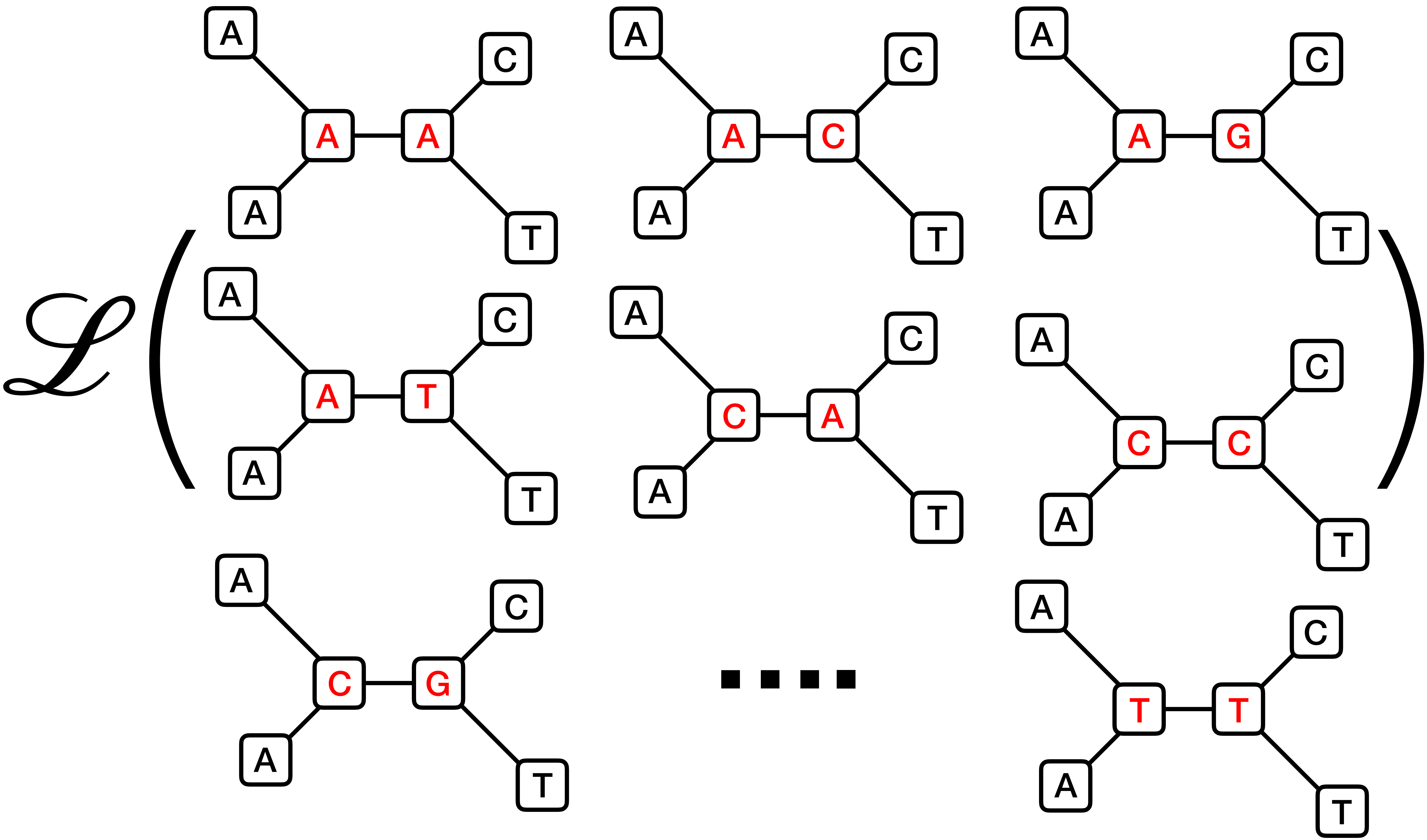
Four-Taxon Unrooted Tree



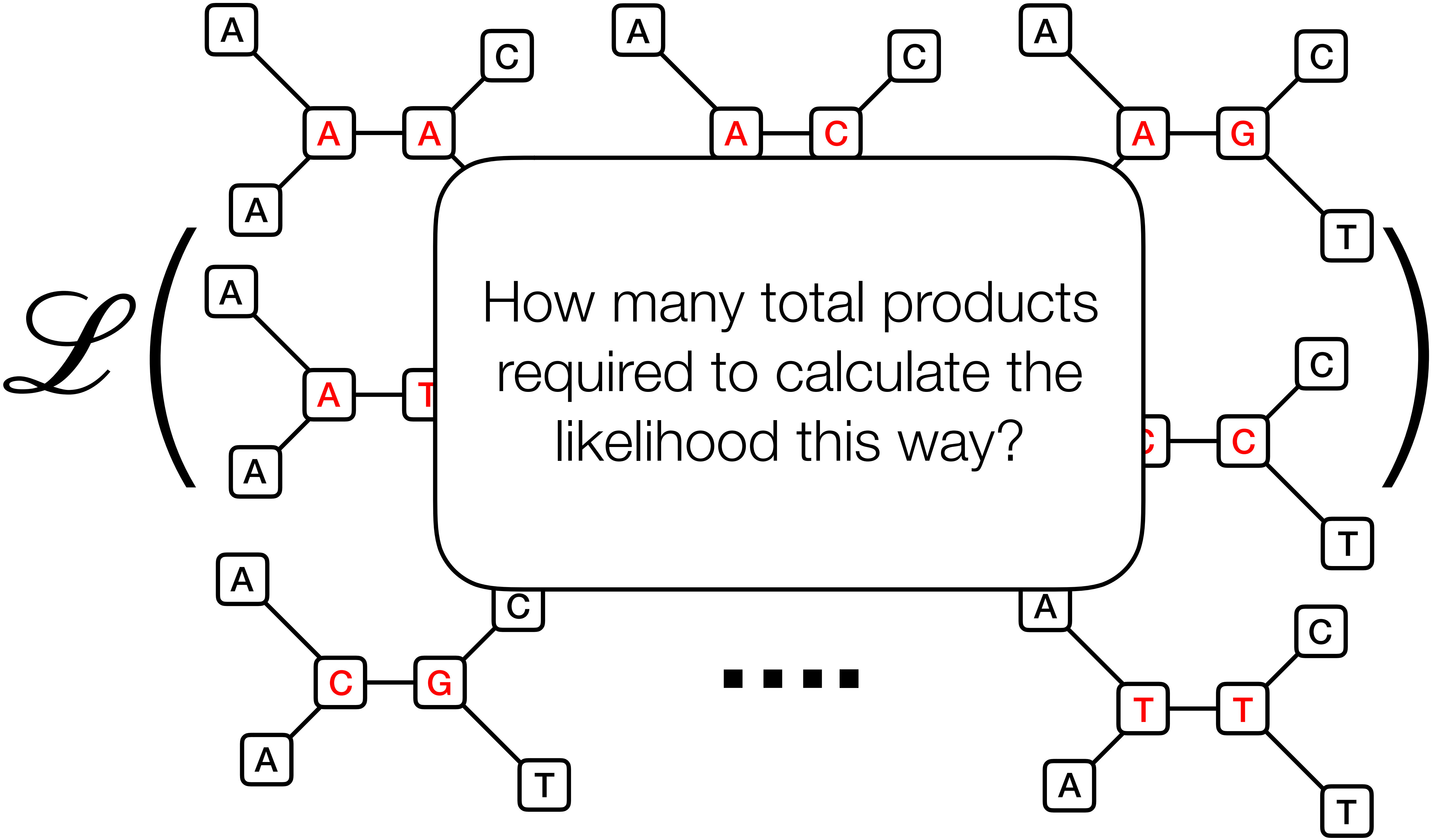
Four-Taxon Unrooted Tree



Four-Taxon Unrooted Tree

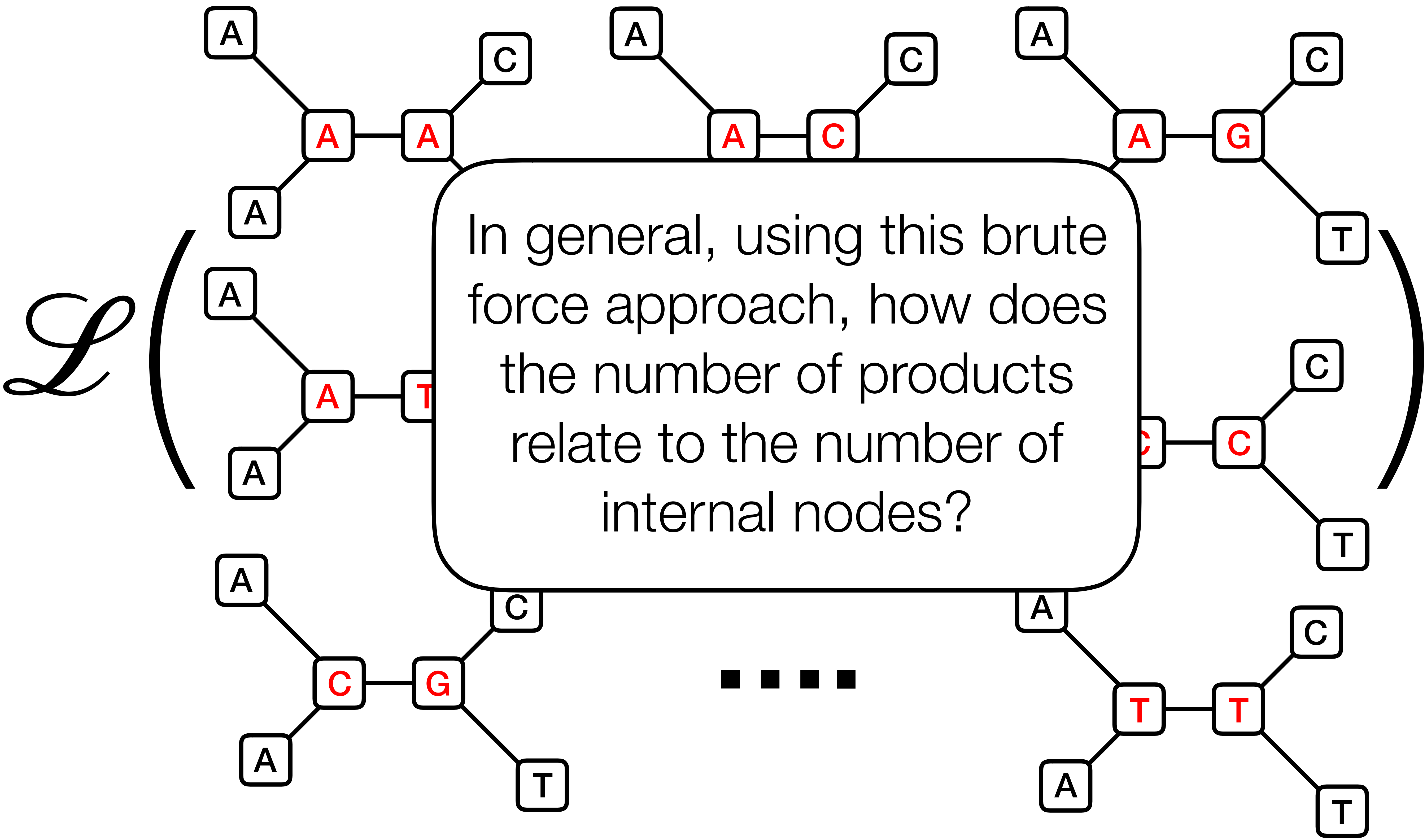


Four-Taxon Unrooted Tree



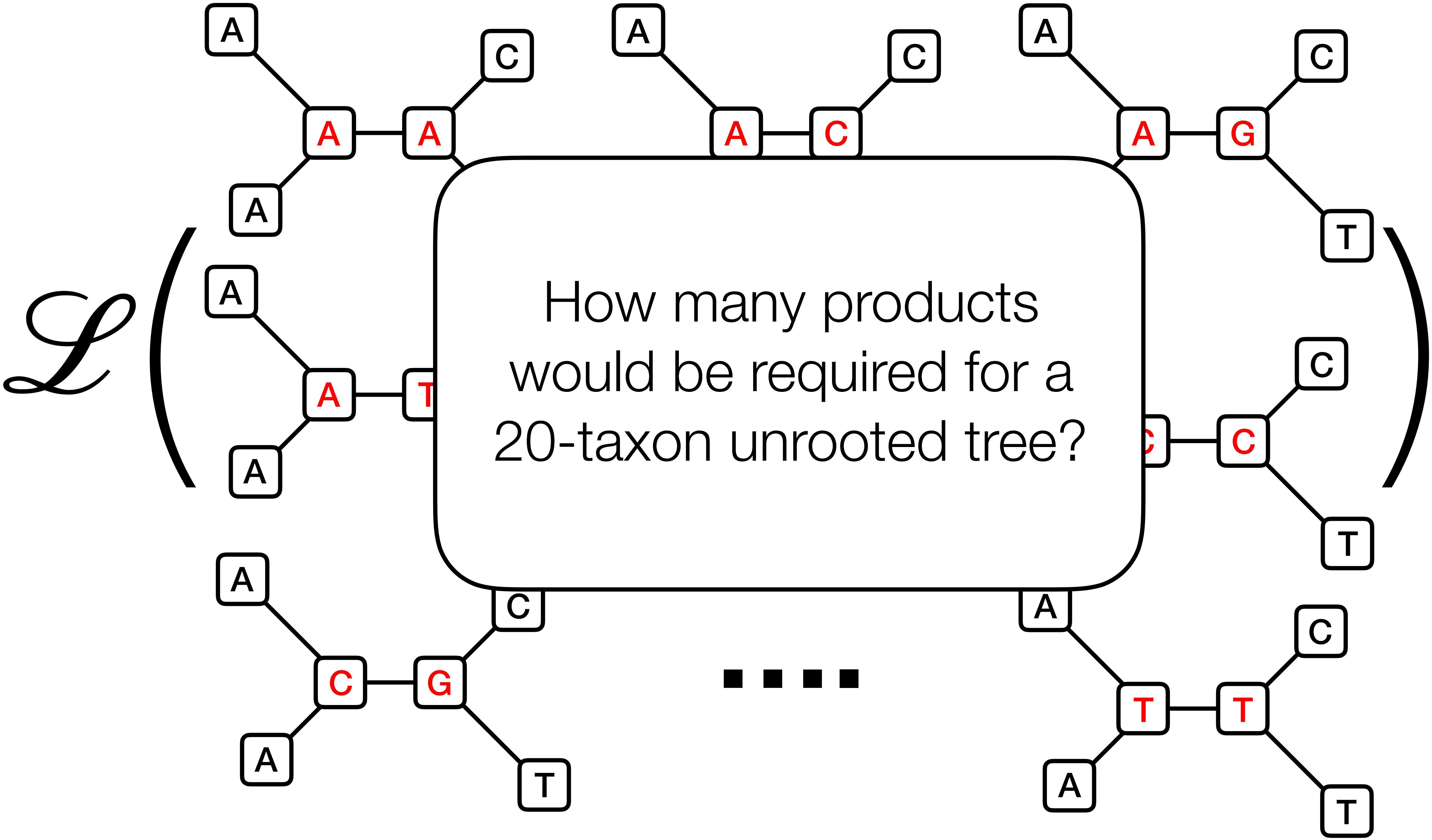
How many total products required to calculate the likelihood this way?

Four-Taxon Unrooted Tree



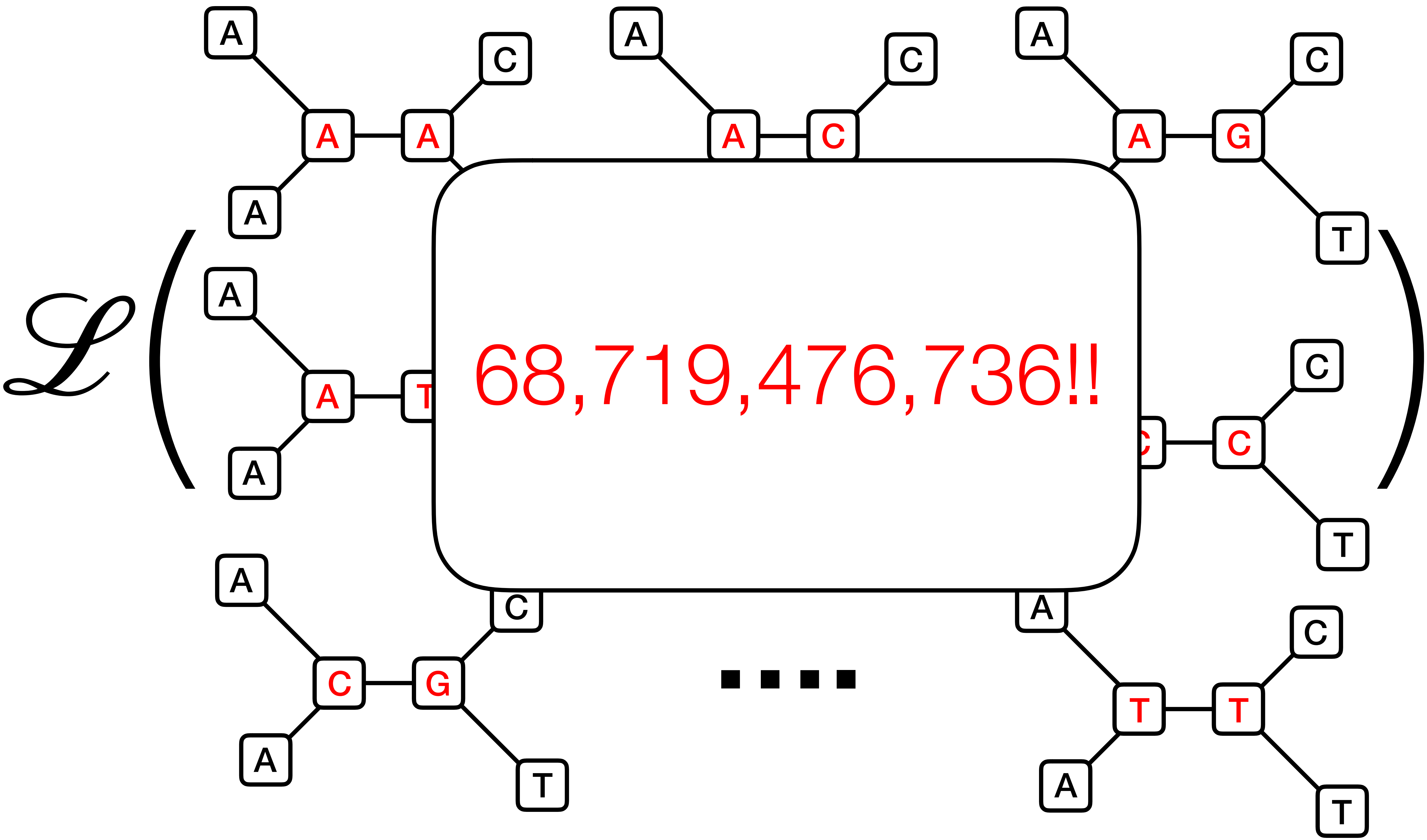
In general, using this brute force approach, how does the number of products relate to the number of internal nodes?

Four-Taxon Unrooted Tree



How many products would be required for a 20-taxon unrooted tree?

Four-Taxon Unrooted Tree



Felsenstein's Pruning Algorithm

J Mol Evol (1981) 17:368–376

Journal of Molecular Evolution

Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach

Joseph Felsenstein

Department of Genetics, University of Washington, Seattle, Washington 98195, USA

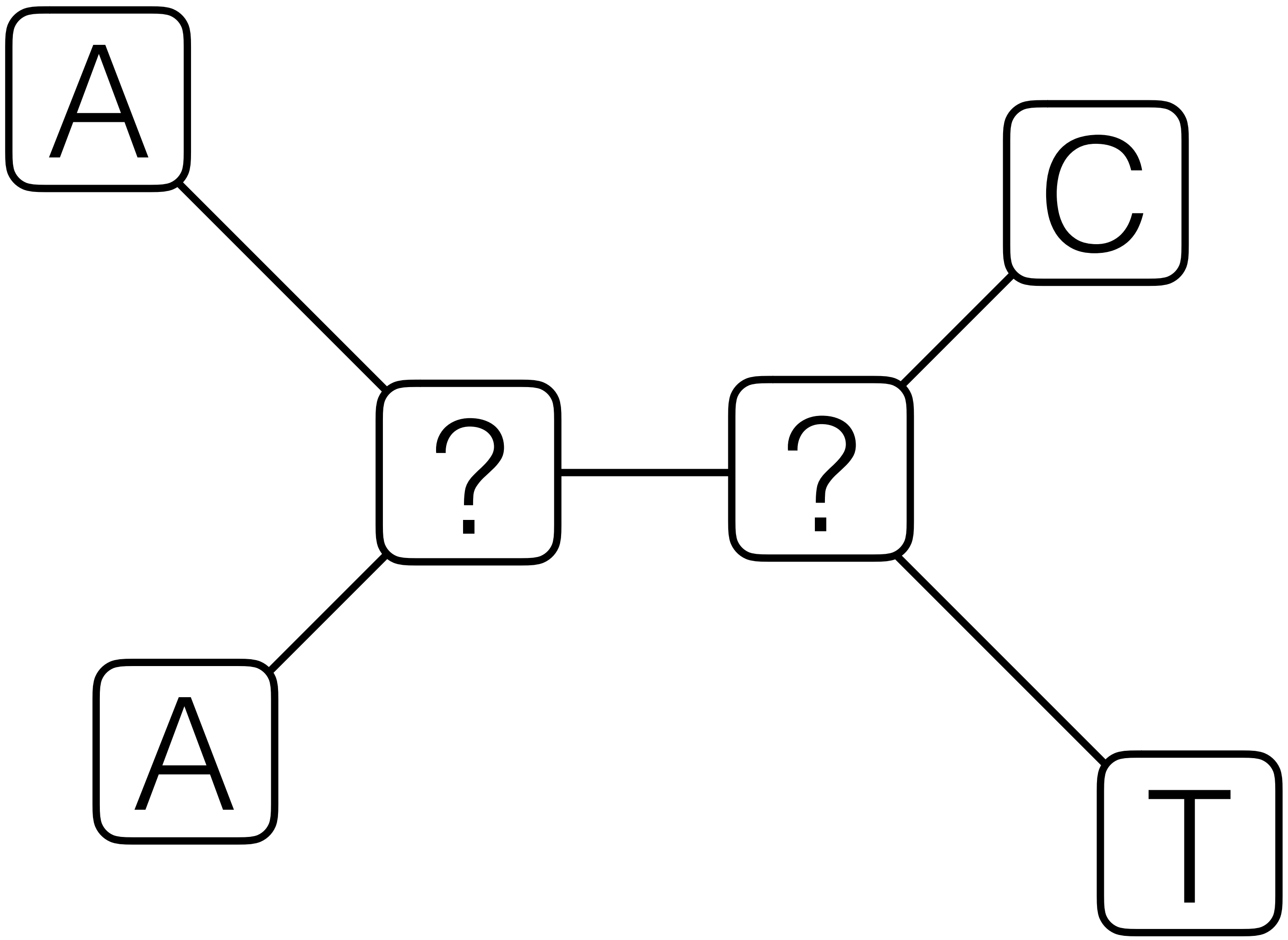
Summary. The application of maximum likelihood techniques to the estimation of evolutionary trees from nucleic acid sequence data is discussed. A computationally feasible method for finding such maximum likelihood estimates is developed, and a computer program is available. This method has advantages over the traditional parsimony algorithms, which can give misleading results if rates of evolution differ in different lineages. It also allows the testing of hypotheses about the constancy of evolutionary rates by likelihood ratio tests, and gives rough indication of the error of the estimate of the tree.

Key words: Evolution – Phylogeny – Maximum likelihood – Parsimony – Estimation – DNA sequences

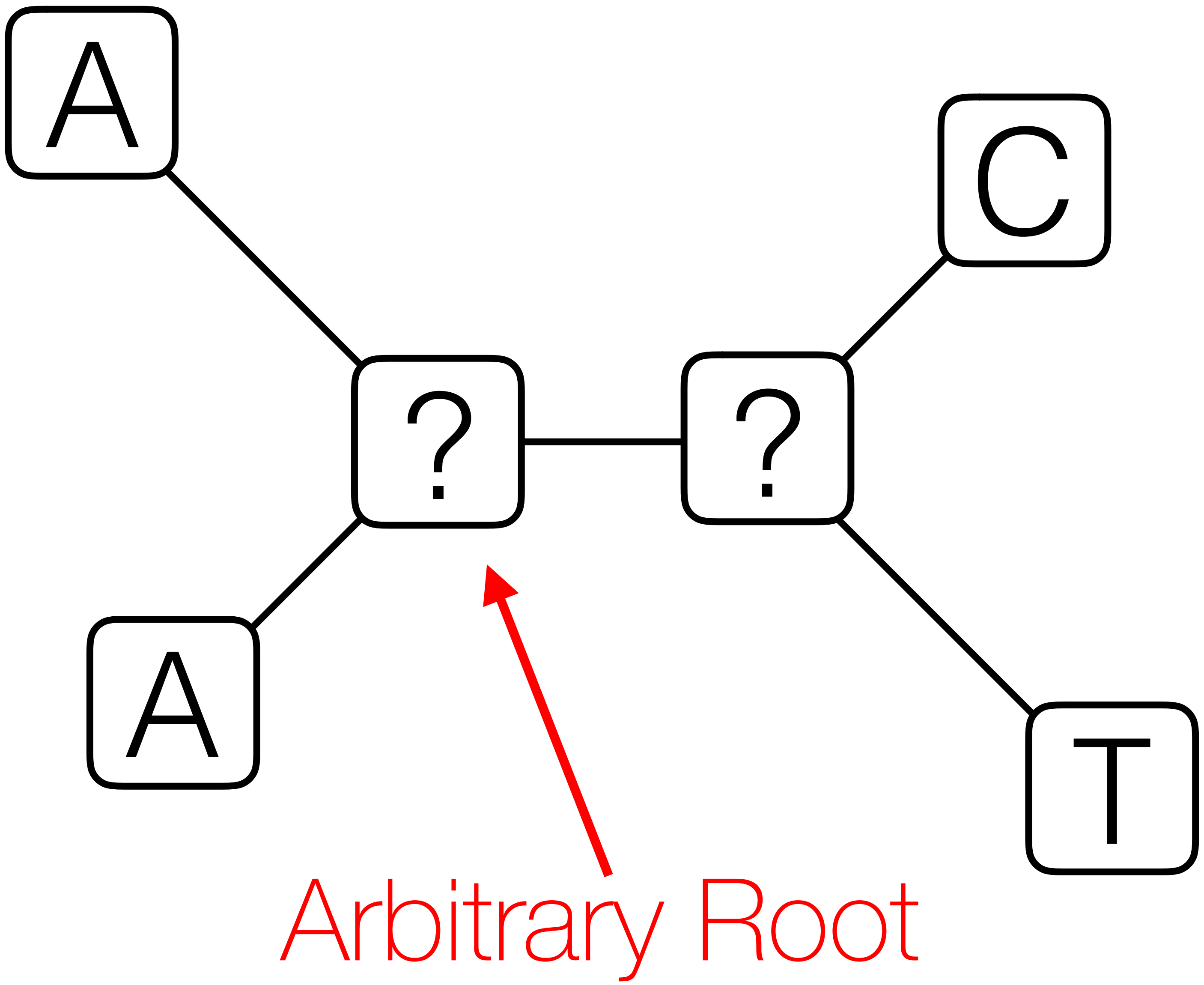
produced by parsimony methods (Edwards 1963; Edwards and Cavalli-Sforza 1964; Camin and Sokal 1965). These methods implicitly assume that change is improbable a priori (Felsenstein 1973, 1979). If the amount of change is small over the evolutionary times being considered, parsimony methods will be well-justified statistical methods.

Most data involve moderate to large amounts of change, and it is in such cases that parsimony methods can fail. When amounts of evolutionary change in different lineages are sufficiently unequal, it can be shown (Felsenstein 1978b) that parsimony methods make an inconsistent estimate of the evolutionary tree, converging to the wrong tree with increasing certainty as more sequences are considered for the same set of species. The compatibility approach to estimating evolutionary trees (Le Quesne 1969; Sneath et al. 1975;

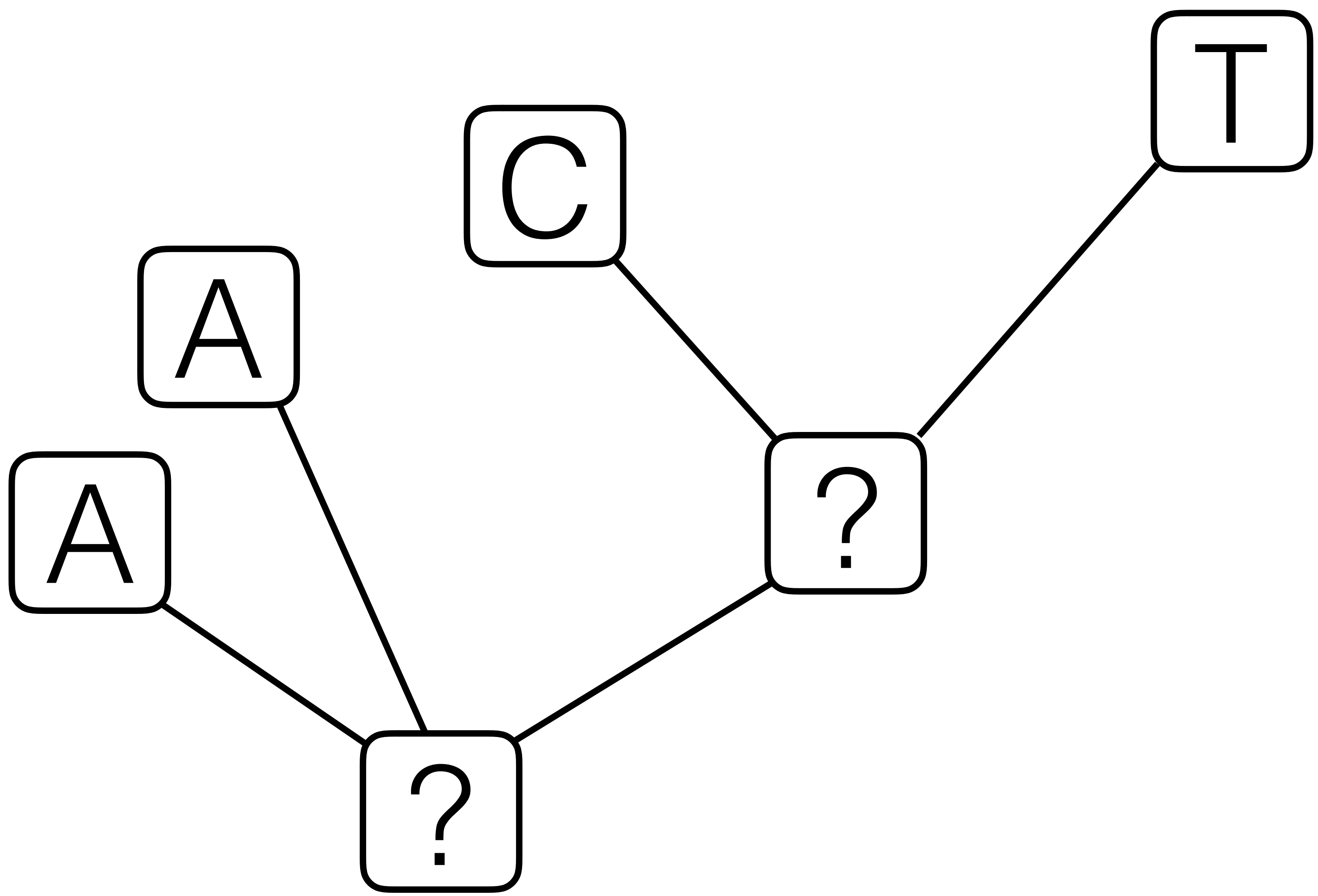
Felsenstein’s Pruning Algorithm



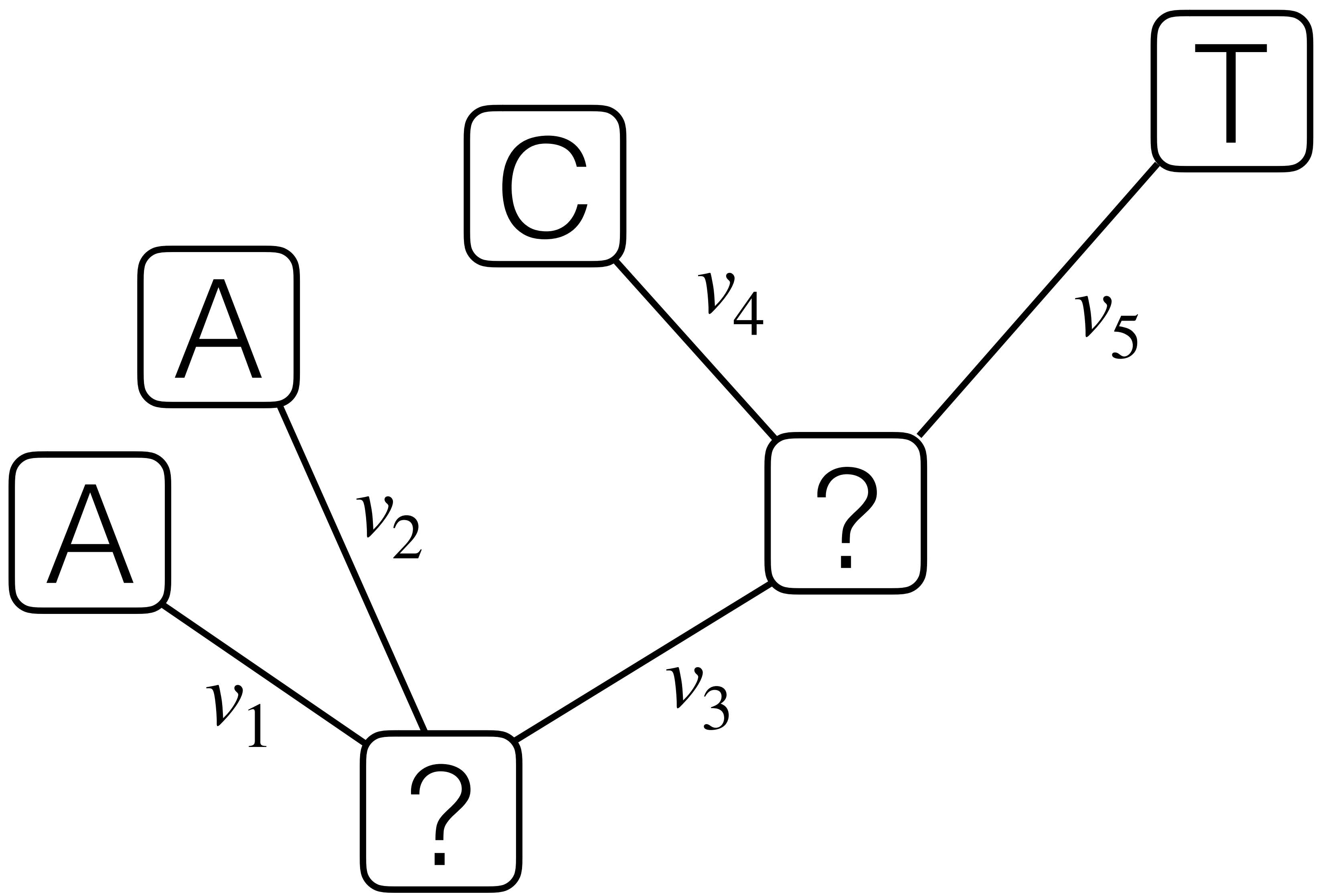
Felsenstein’s Pruning Algorithm



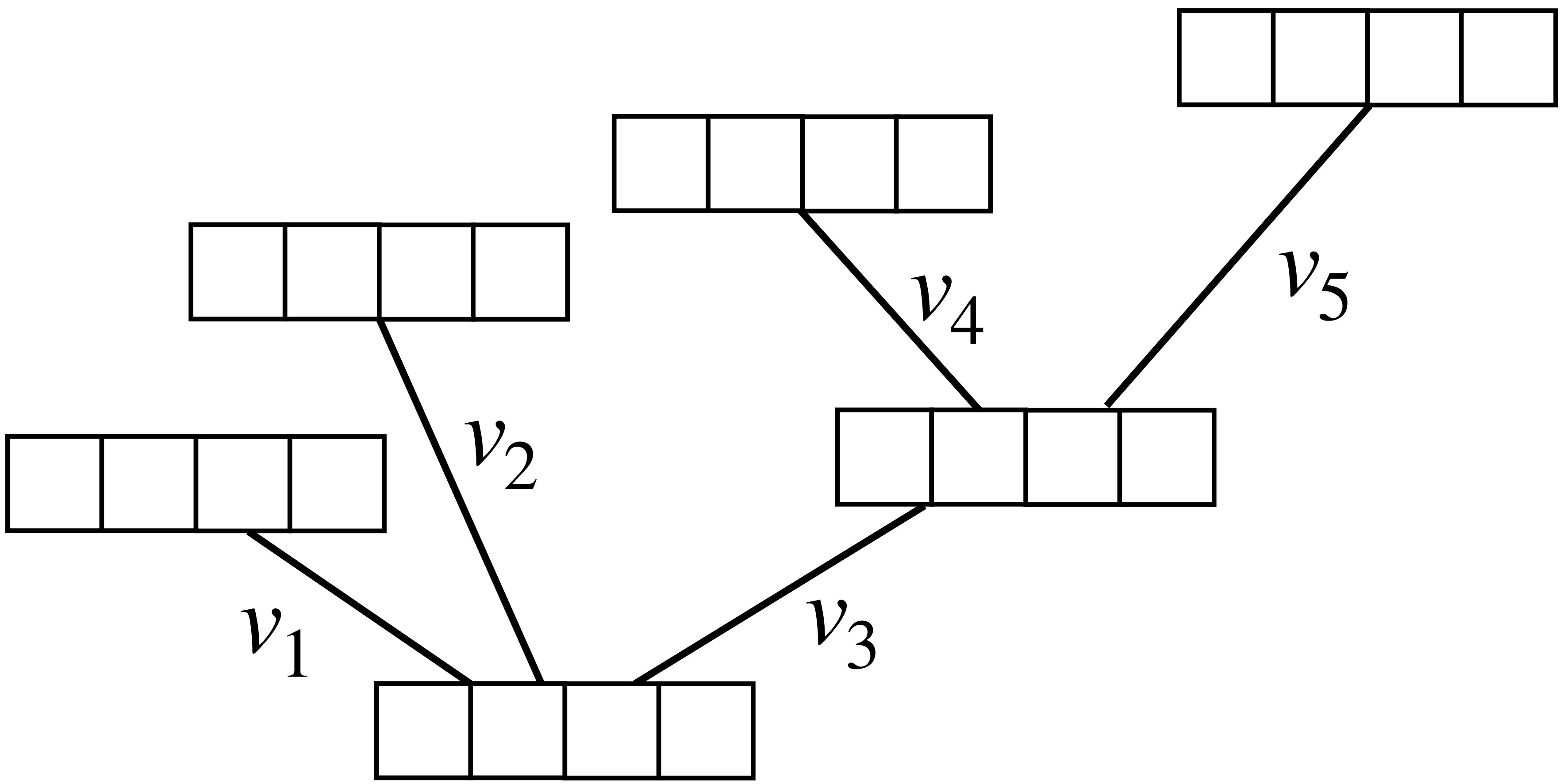
Felsenstein's Pruning Algorithm



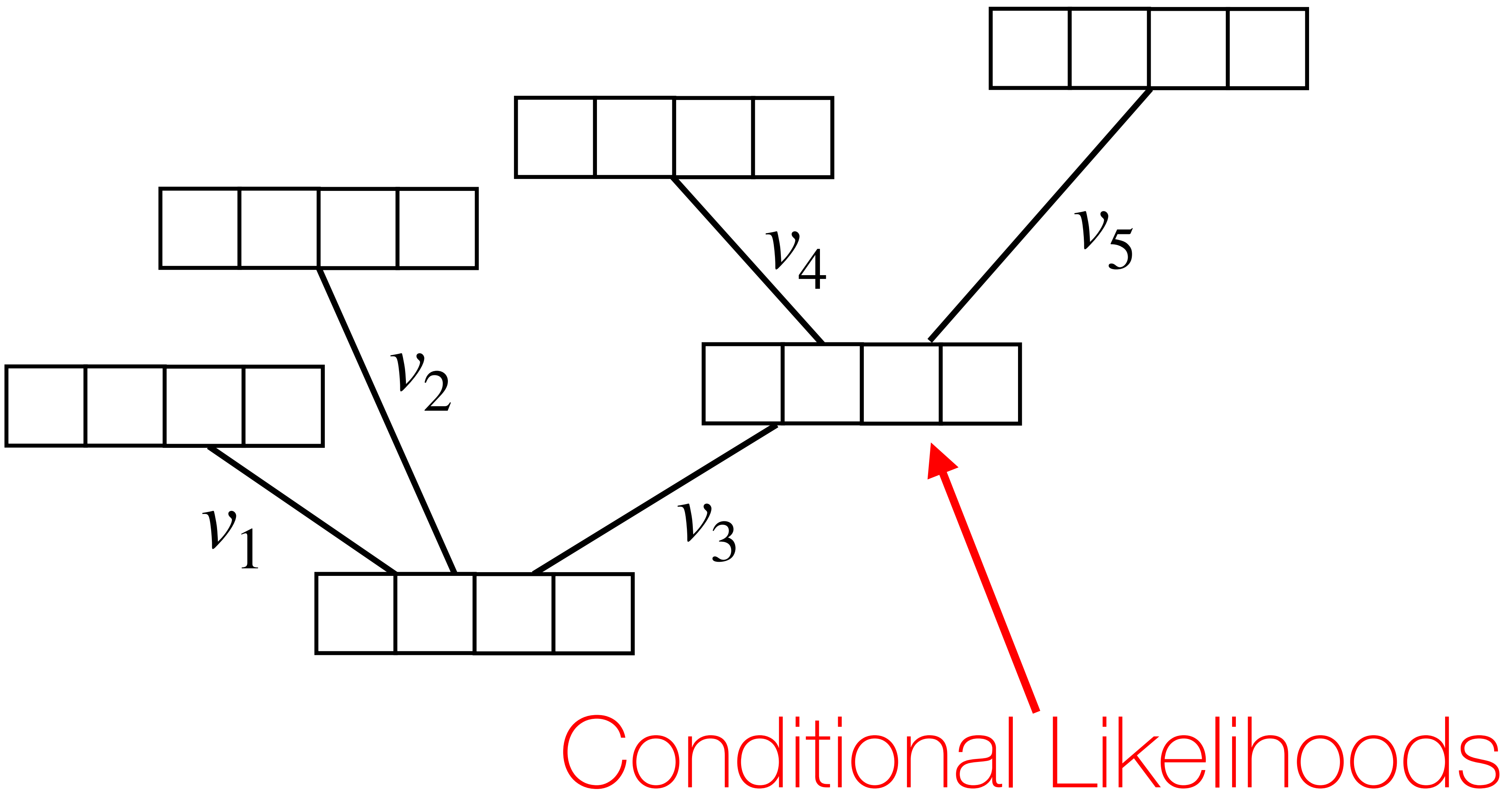
Felsenstein's Pruning Algorithm



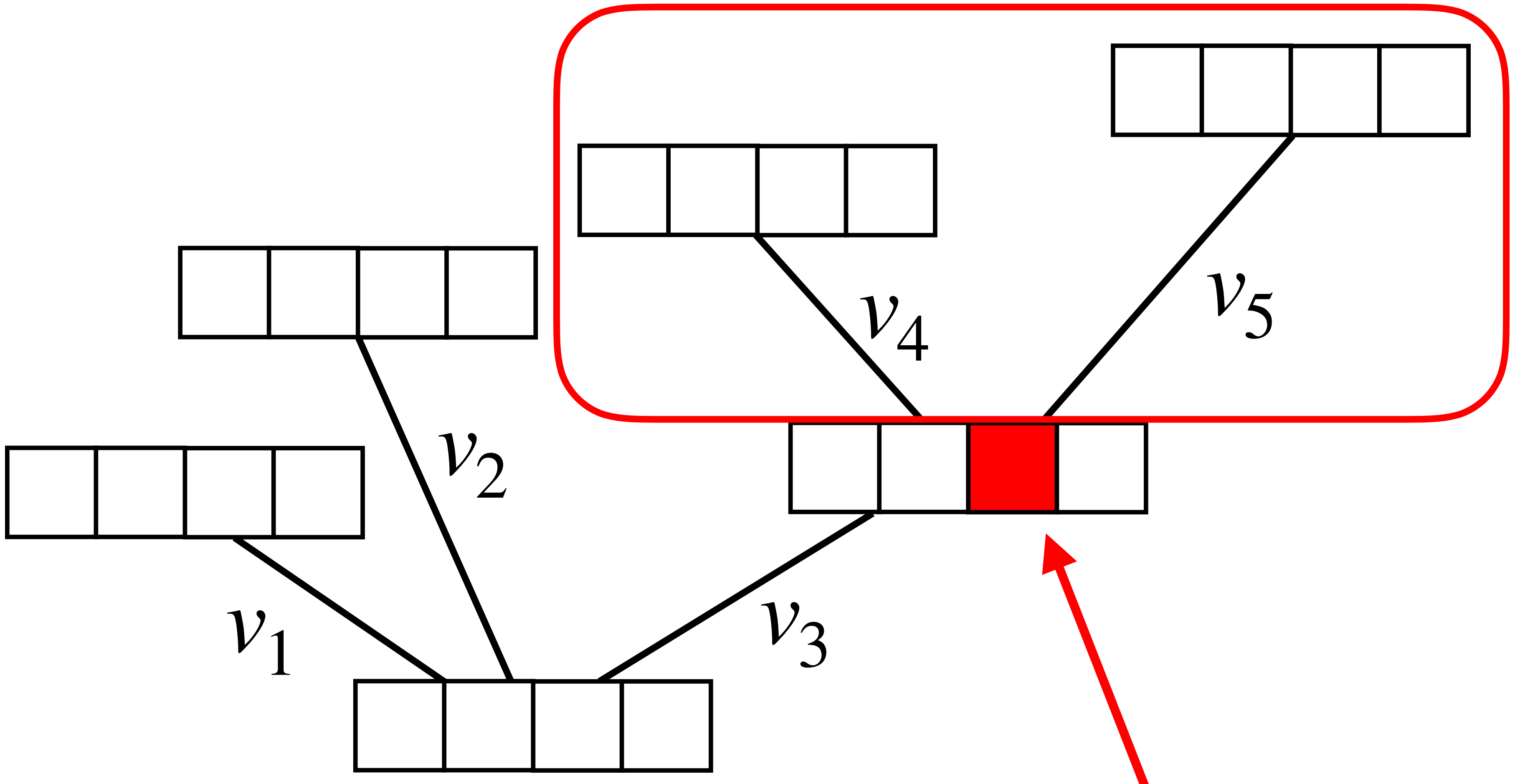
Felsenstein's Pruning Algorithm



Felsenstein's Pruning Algorithm



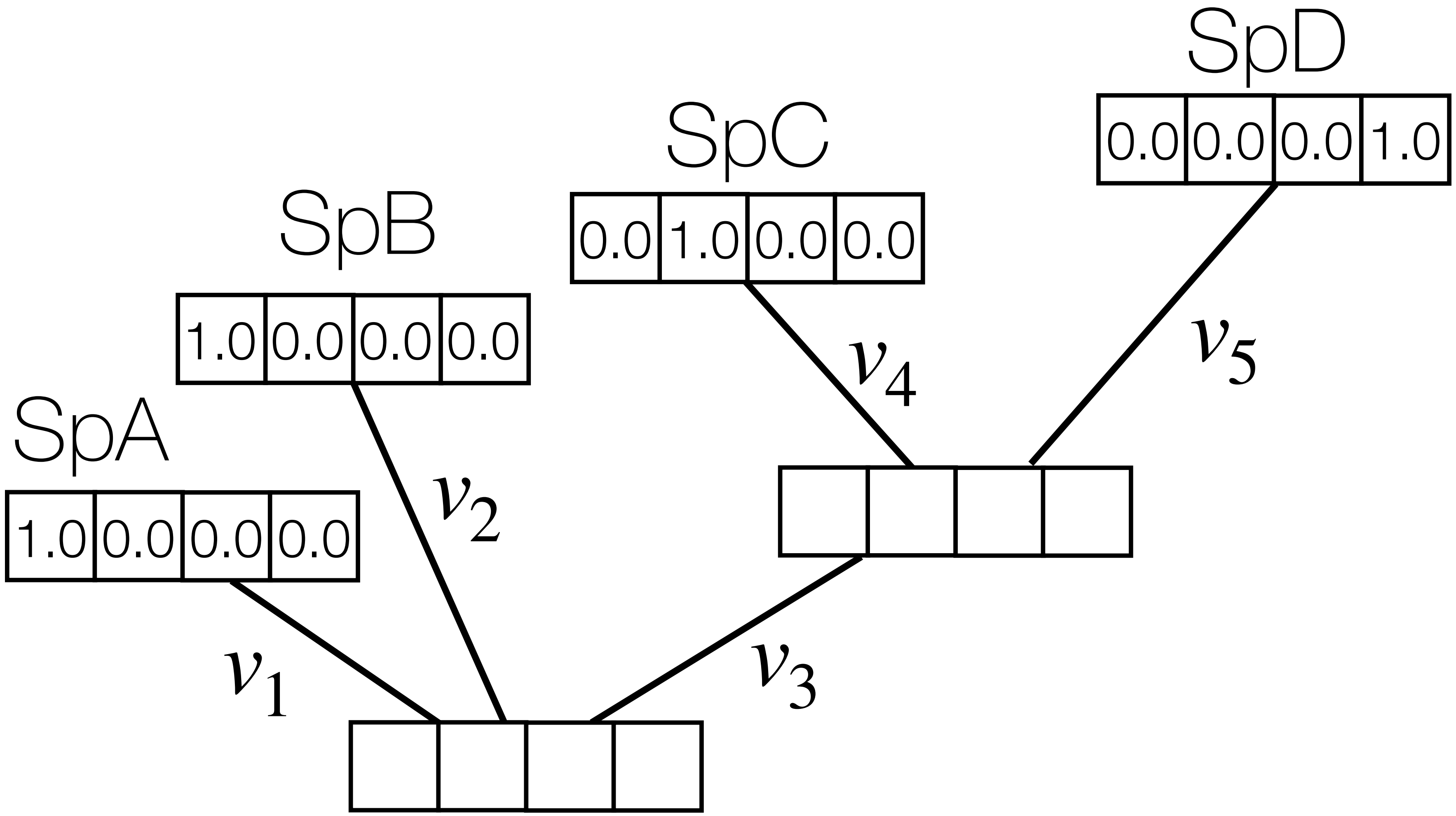
Felsenstein's Pruning Algorithm



Conditional Likelihoods

What's probability of everything
above this node if this was a G?

Felsenstein's Pruning Algorithm

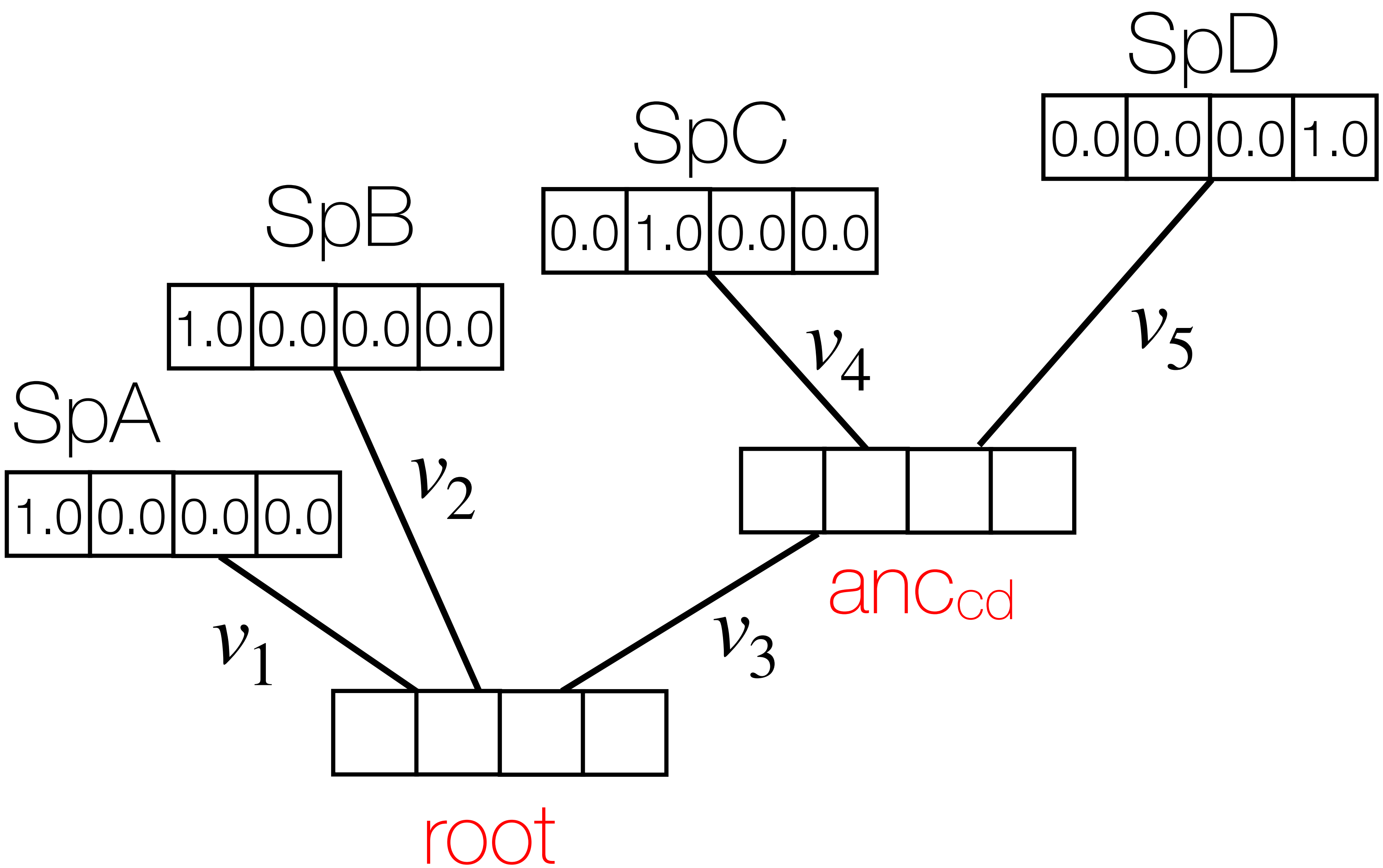


Step 1. Fill in Observations

What bases are consistent with your observation?

If no error or ambiguity, 1.0 for one base, 0.0 for others.

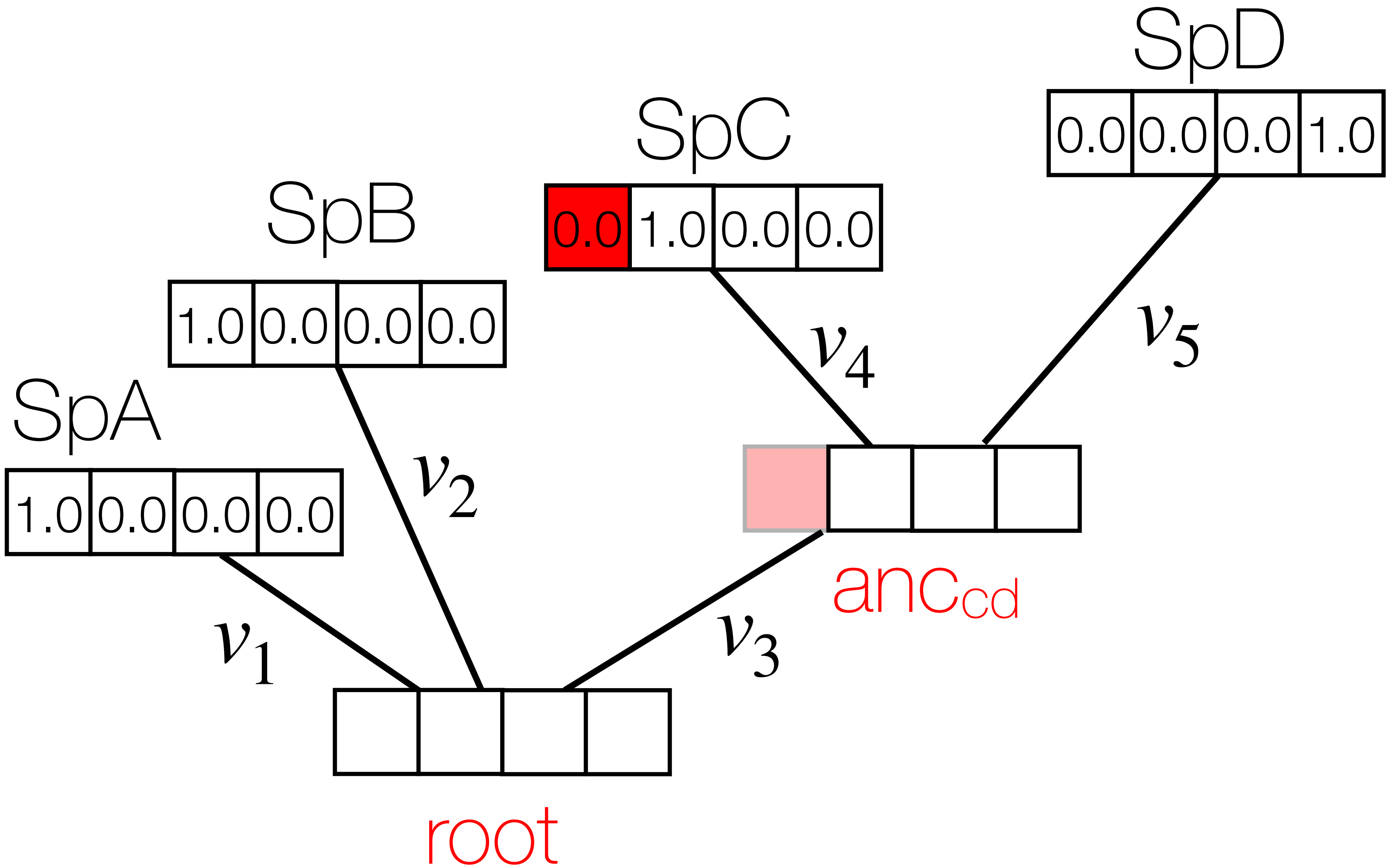
Felsenstein's Pruning Algorithm



Step 2. Work your way down the tree

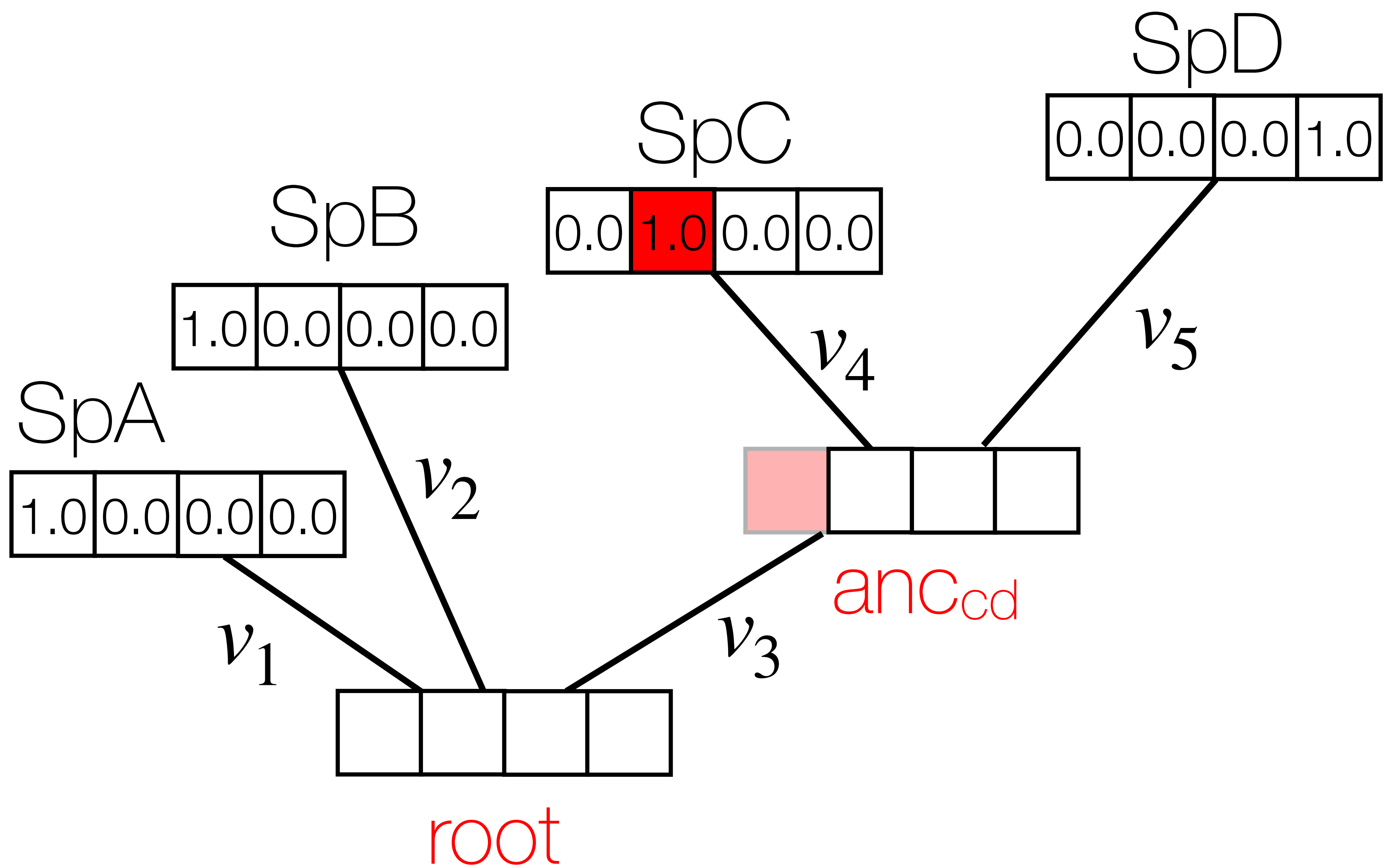
Post-order traversal

Felsenstein's Pruning Algorithm



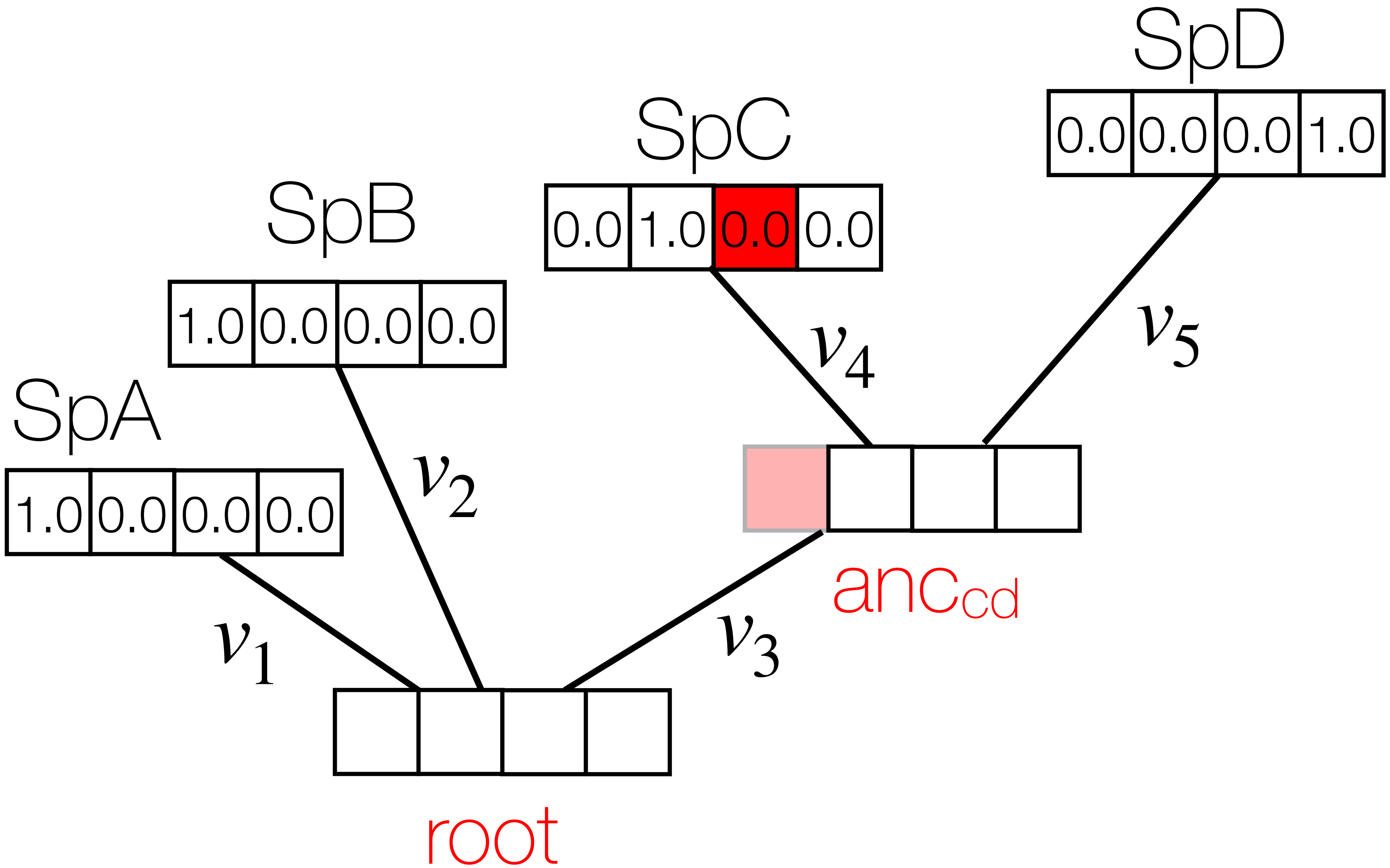
$$\ell_A^{ancCD} = P_{AA}(v_4)\ell_A^{spC}$$

Felsenstein's Pruning Algorithm



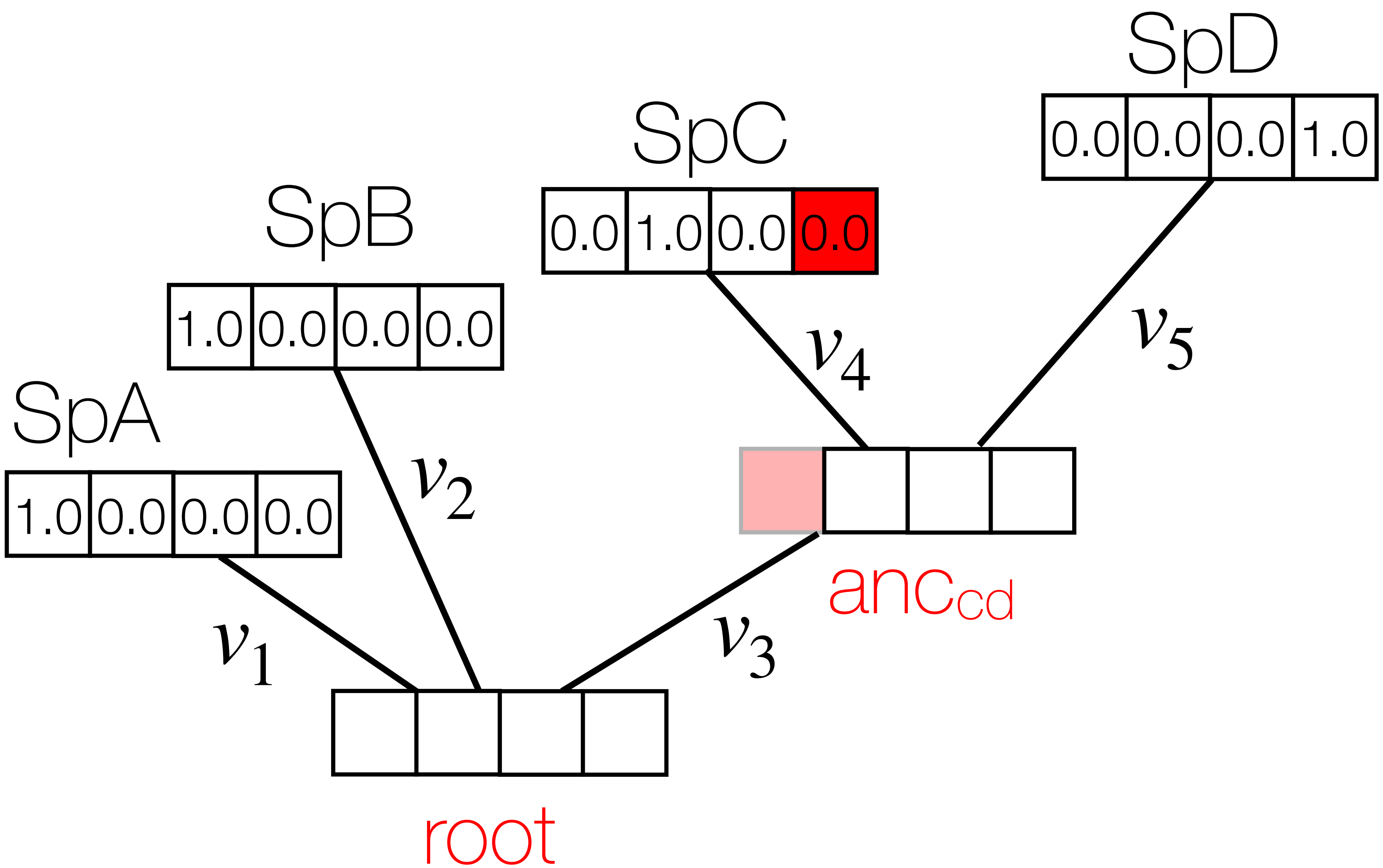
$$\ell_A^{ancCD} = P_{AA}(v_4)\ell_A^{spC} + P_{AC}(v_4)\ell_C^{spC}$$

Felsenstein's Pruning Algorithm



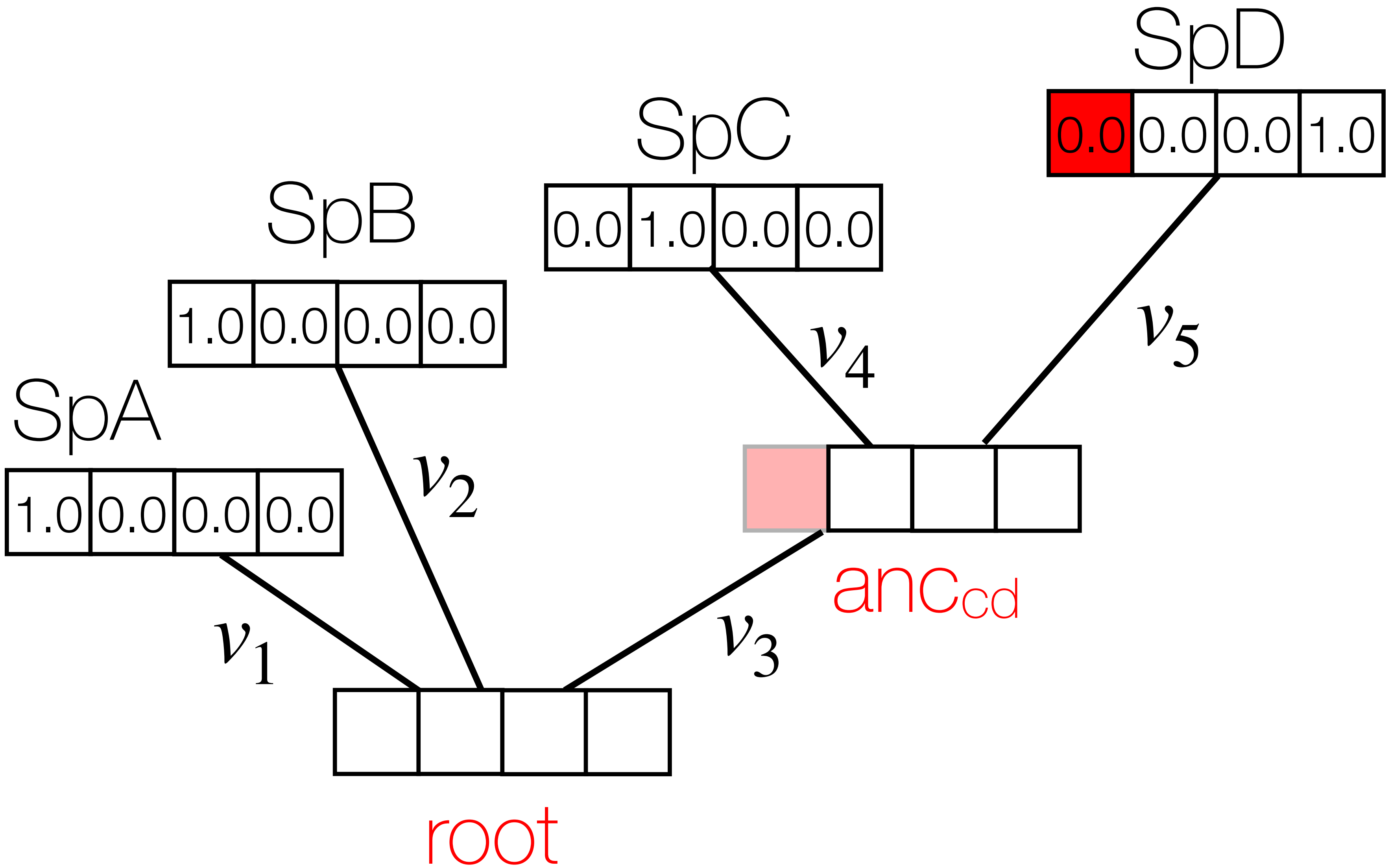
$$\ell_A^{ancCD} = P_{AA}(v_4)\ell_A^{spC} + P_{AC}(v_4)\ell_C^{spC} + P_{AG}(v_4)\ell_G^{spC}$$

Felsenstein's Pruning Algorithm



$$\ell_A^{ancCD} = P_{AA}(v_4)\ell_A^{spC} + P_{AC}(v_4)\ell_C^{spC} + P_{AG}(v_4)\ell_G^{spC} + P_{AT}(v_4)\ell_T^{spC}$$

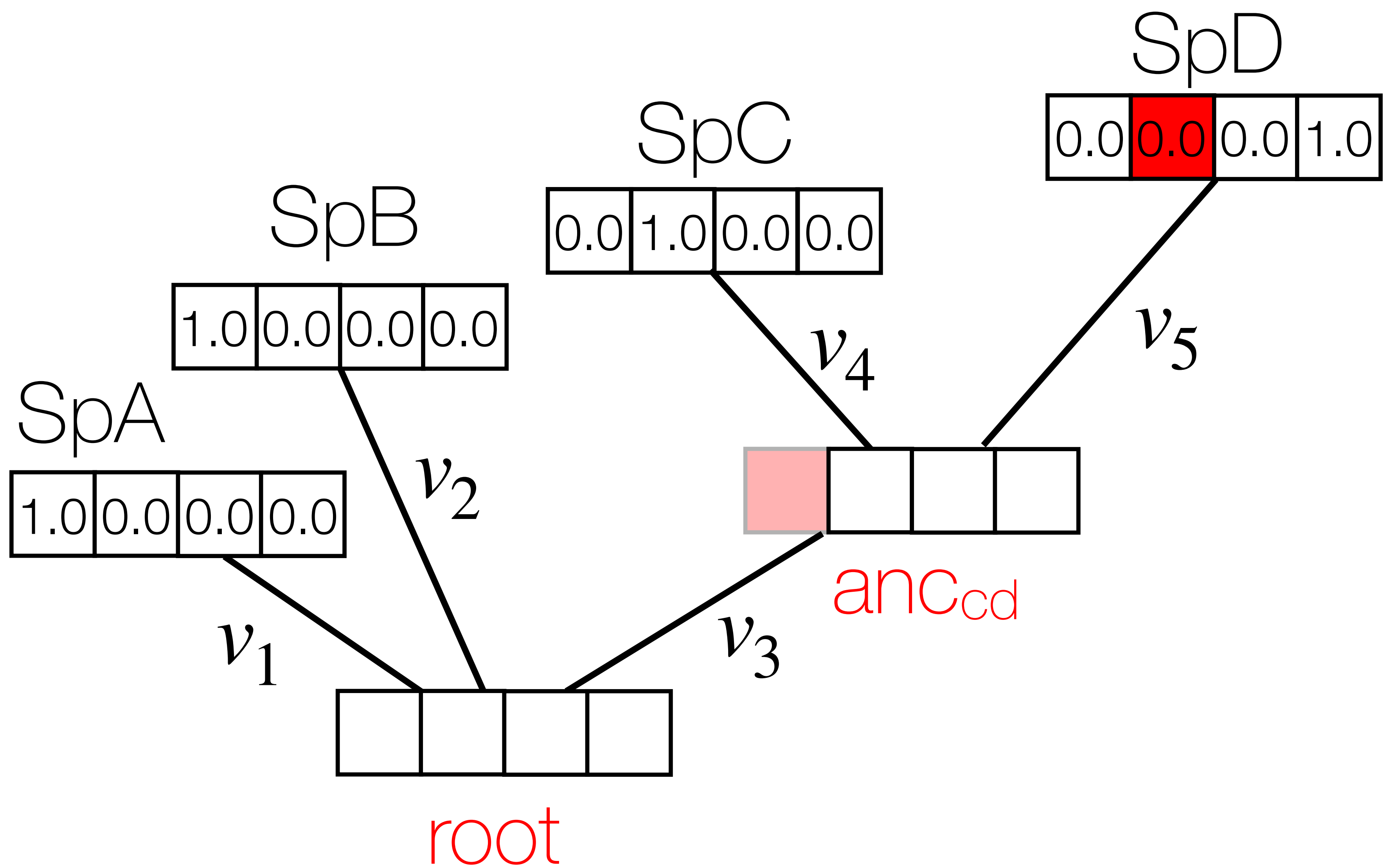
Felsenstein's Pruning Algorithm



$$\ell_A^{ancCD} = P_{AA}(v_4)\ell_A^{spC} + P_{AC}(v_4)\ell_C^{spC} + P_{AG}(v_4)\ell_G^{spC} + P_{AT}(v_4)\ell_T^{spC}$$

$$+P_{AA}(v_5)\ell_A^{spD}$$

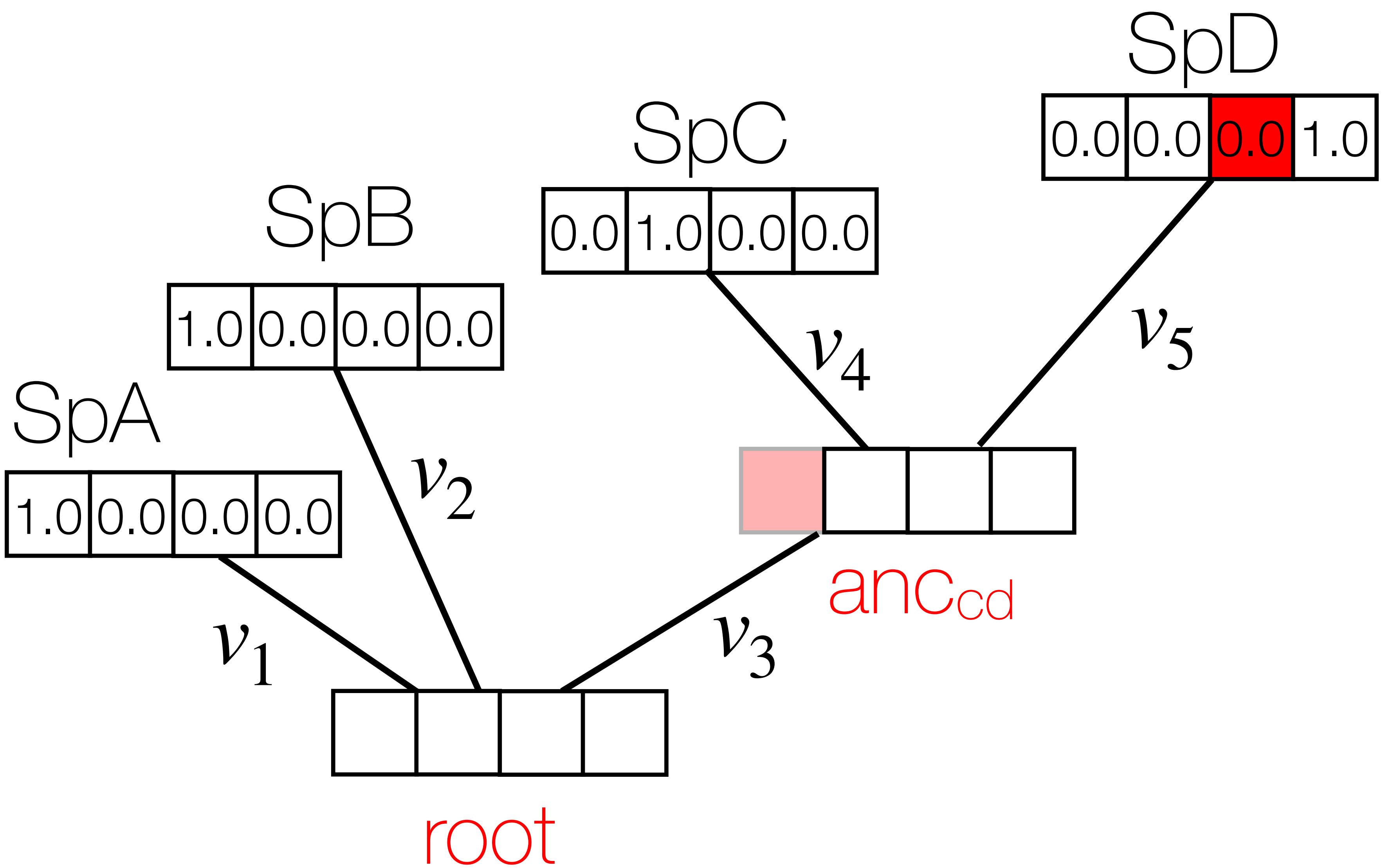
Felsenstein's Pruning Algorithm



$$\ell_A^{ancCD} = P_{AA}(v_4)\ell_A^{spC} + P_{AC}(v_4)\ell_C^{spC} + P_{AG}(v_4)\ell_G^{spC} + P_{AT}(v_4)\ell_T^{spC}$$

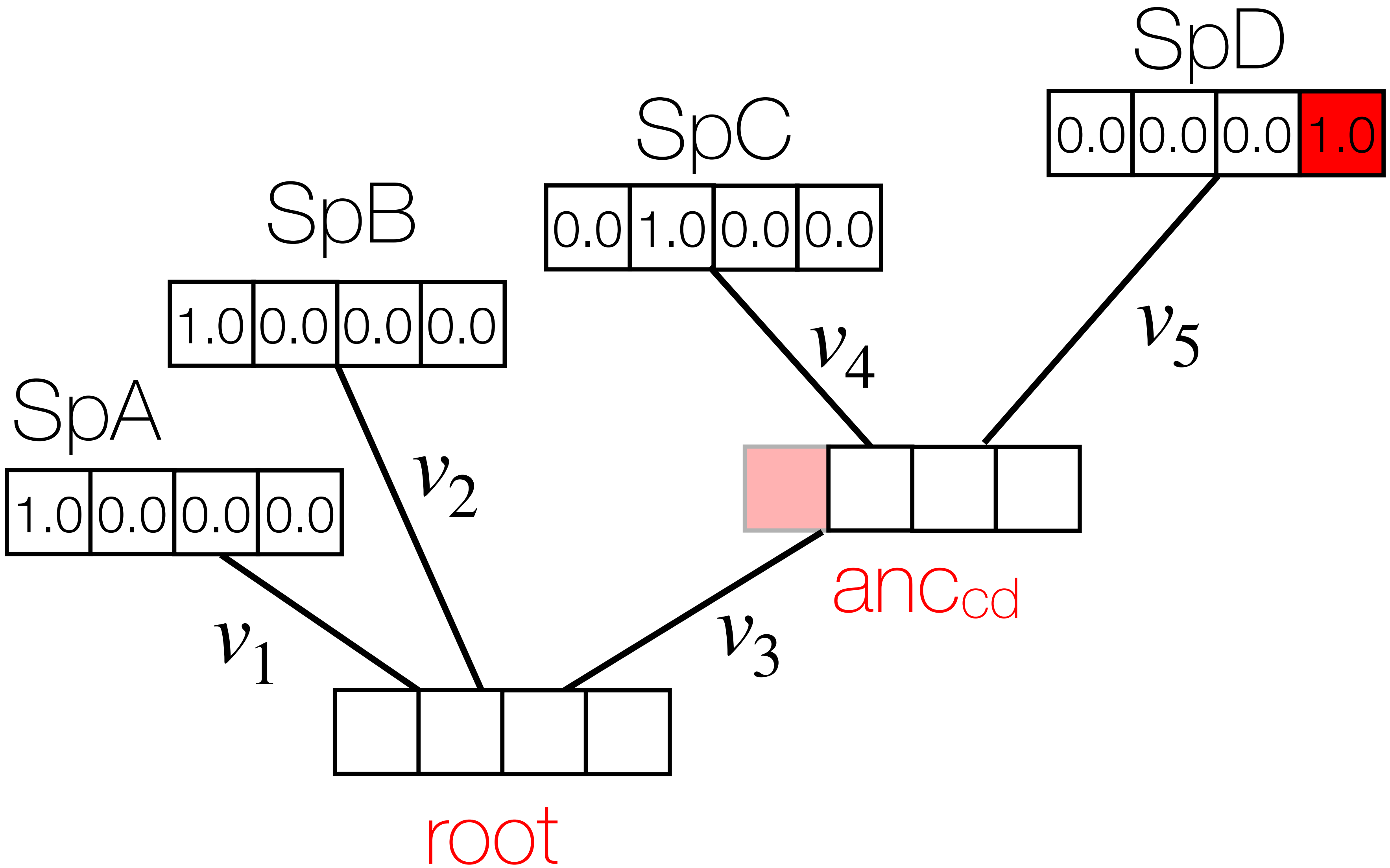
$$+P_{AA}(v_5)\ell_A^{spD} + P_{AC}(v_5)\ell_C^{spD}$$

Felsenstein's Pruning Algorithm



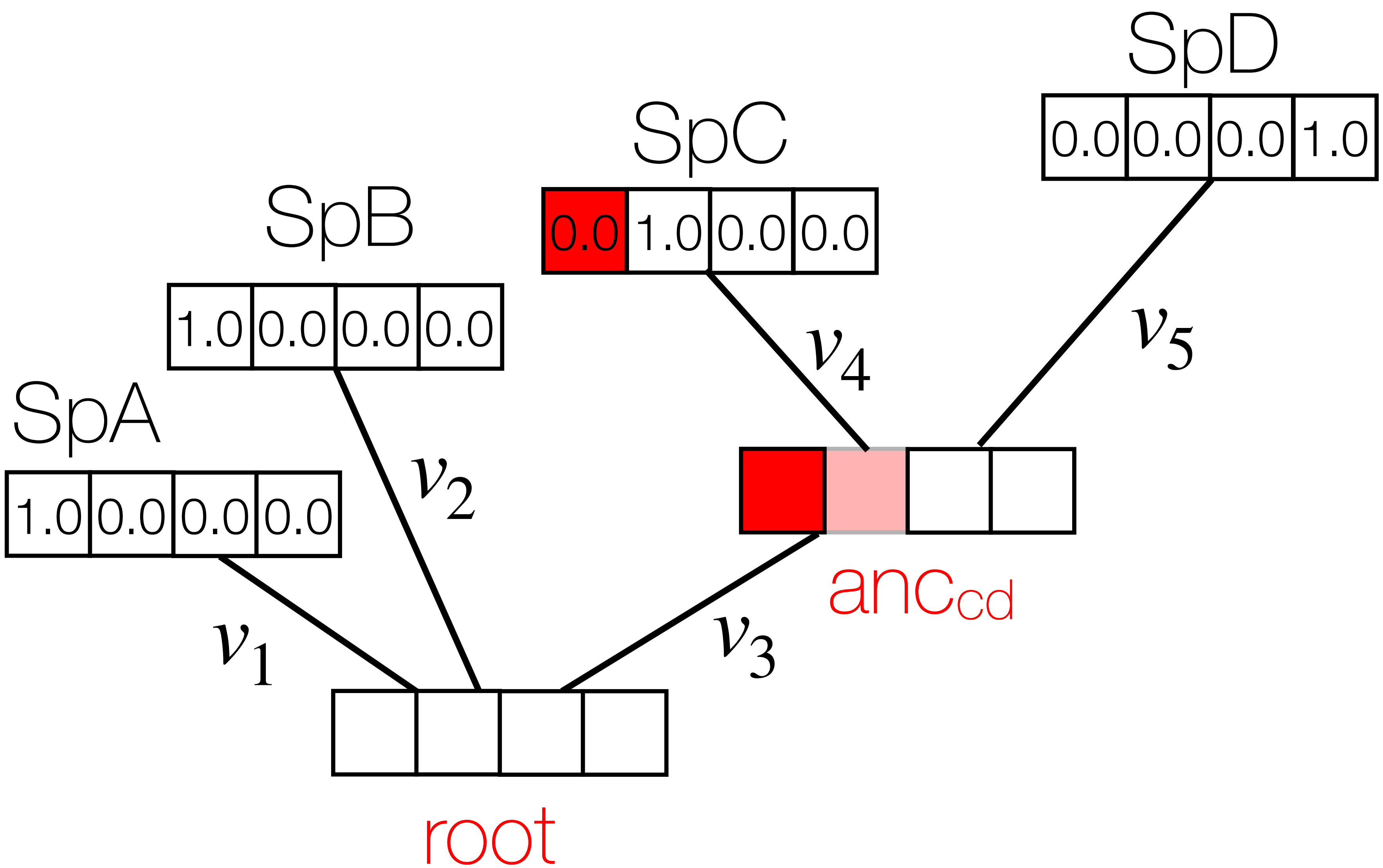
$$\ell_A^{ancCD} = P_{AA}(v_4)\ell_A^{spC} + P_{AC}(v_4)\ell_C^{spC} + P_{AG}(v_4)\ell_G^{spC} + P_{AT}(v_4)\ell_T^{spC}$$
$$+P_{AA}(v_5)\ell_A^{spD} + P_{AC}(v_5)\ell_C^{spD} + P_{AG}(v_5)\ell_G^{spD}$$

Felsenstein's Pruning Algorithm



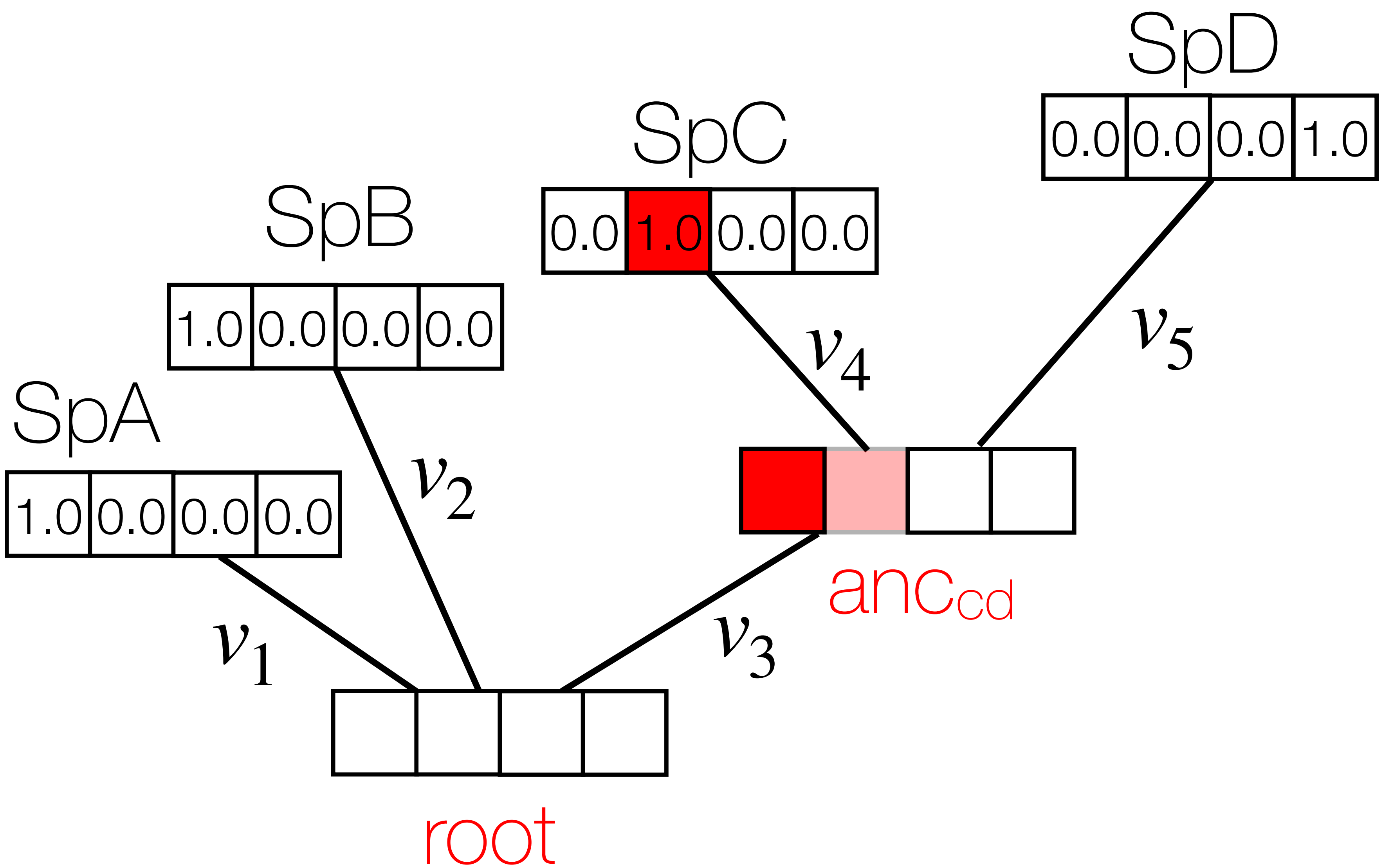
$$\ell_A^{ancCD} = P_{AA}(v_4)\ell_A^{spC} + P_{AC}(v_4)\ell_C^{spC} + P_{AG}(v_4)\ell_G^{spC} + P_{AT}(v_4)\ell_T^{spC}$$
$$+P_{AA}(v_5)\ell_A^{spD} + P_{AC}(v_5)\ell_C^{spD} + P_{AG}(v_5)\ell_G^{spD} + P_{AT}(v_5)\ell_T^{spD}$$

Felsenstein's Pruning Algorithm



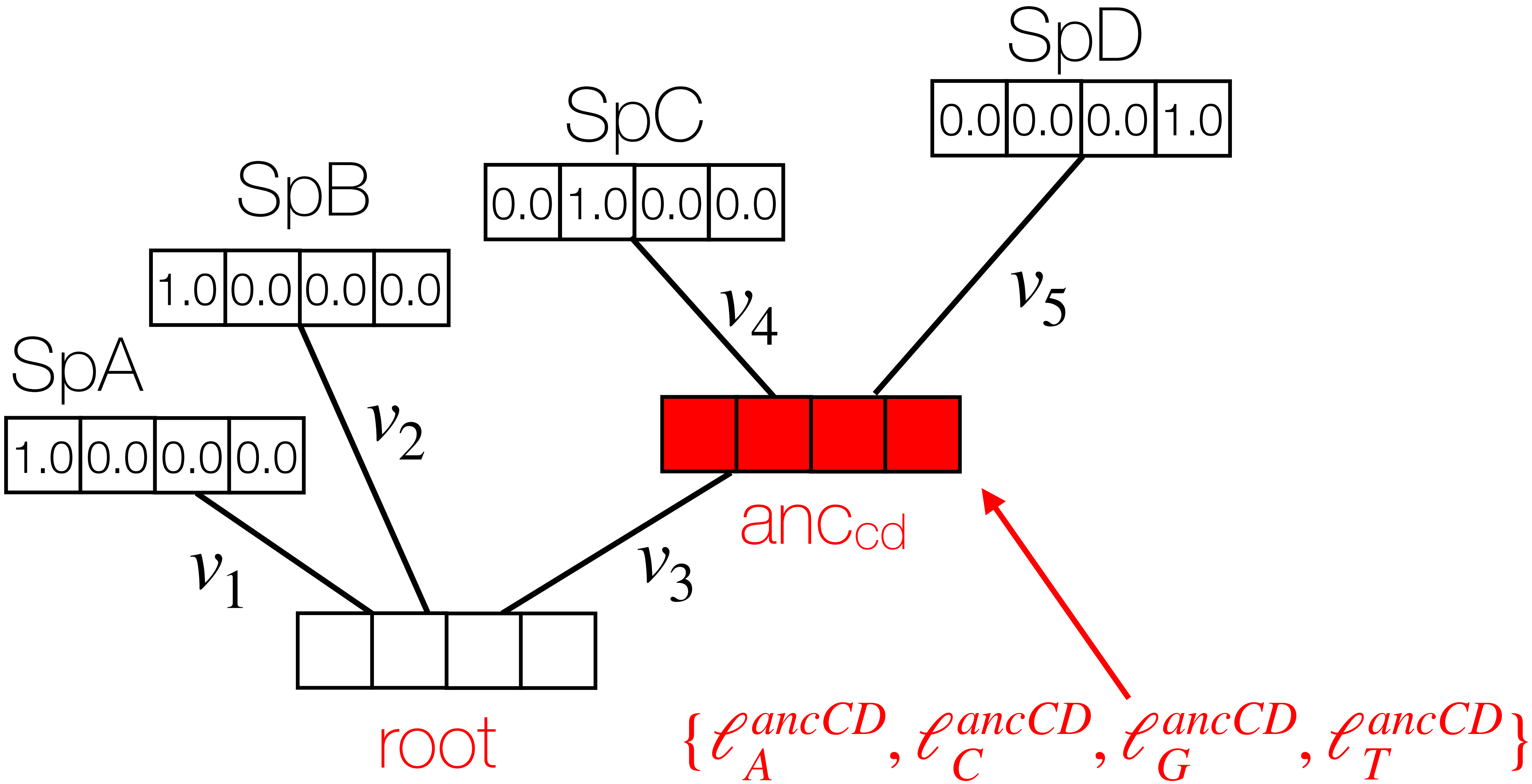
$$\ell_C^{ancCD} = P_{CA}(v_4)\ell_A^{spC}$$

Felsenstein's Pruning Algorithm

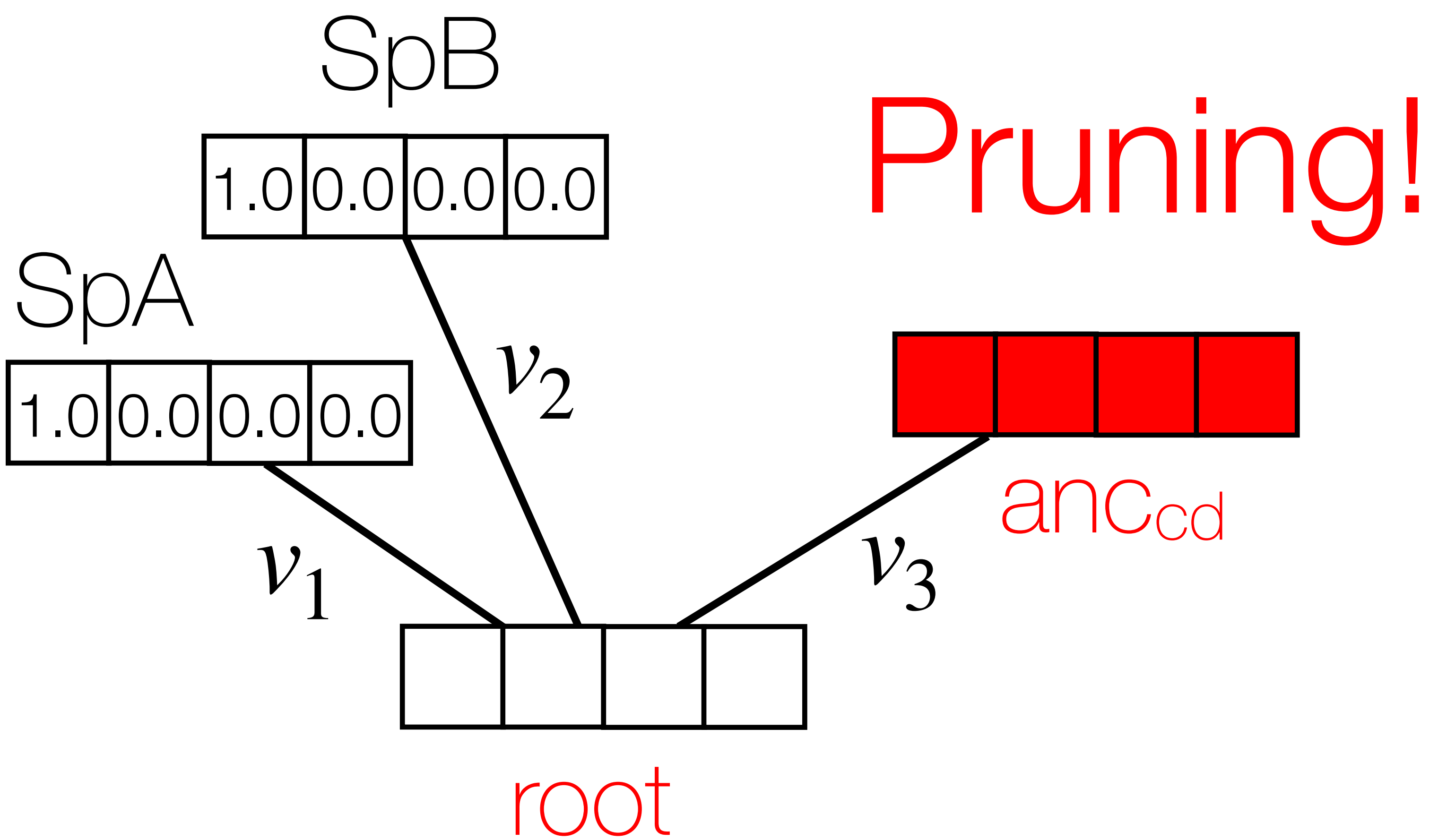


$$\ell_C^{ancCD} = P_{CA}(v_4)\ell_A^{spC} + P_{CC}(v_4)\ell_C^{spC} + \ldots$$

Felsenstein's Pruning Algorithm



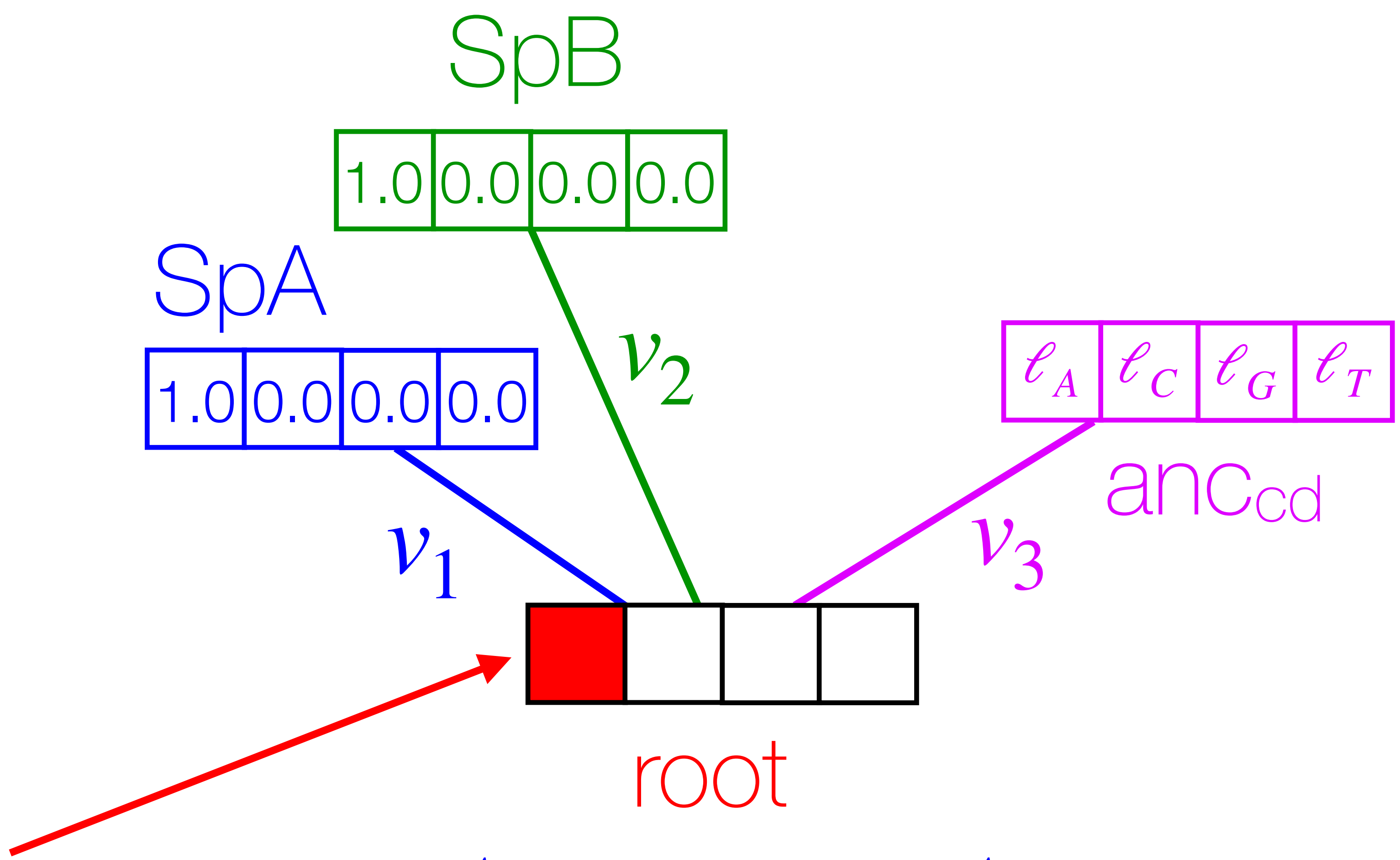
Felsenstein's Pruning Algorithm



Step 2. Work your way down the tree

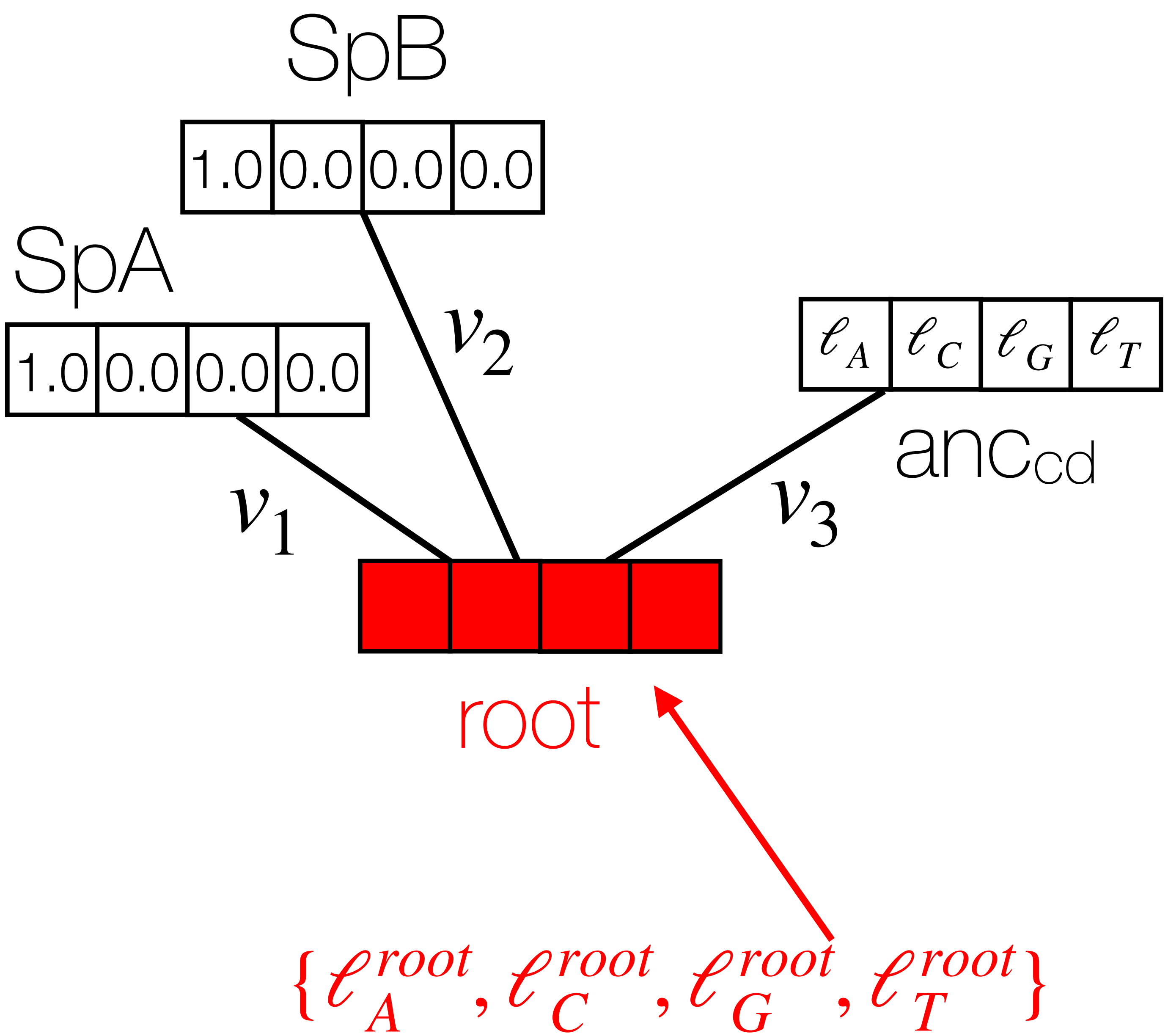
Post-order traversal

Felsenstein's Pruning Algorithm



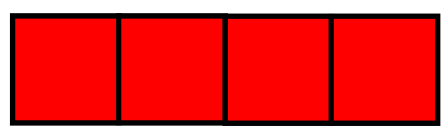
$$\ell_A^{root} = P_{AA}(v_1)\ell_A^{spA} + P_{AC}(v_1)\ell_C^{spA} + P_{AG}(v_1)\ell_G^{spA} + P_{AT}(v_1)\ell_T^{spA}$$
$$+P_{AA}(v_2)\ell_A^{spB} + P_{AC}(v_2)\ell_C^{spB} + P_{AG}(v_2)\ell_G^{spB} + P_{AT}(v_2)\ell_T^{spB}$$
$$+P_{AA}(v_3)\ell_A^{ancCD} + P_{AC}(v_3)\ell_C^{ancCD} + P_{AG}(v_3)\ell_G^{ancCD} + P_{AT}(v_3)\ell_T^{ancCD}$$

Felsenstein's Pruning Algorithm



Felsenstein's Pruning Algorithm

$$\mathscr{L} = (\pi_A \ell_A^{root}) + (\pi_C \ell_C^{root}) + (\pi_G \ell_G^{root}) + (\pi_T \ell_T^{root})$$



Step 3. Weight partial likelihoods at root
by base frequencies.