

Mixture Models

Classic Mixture Models

- Proportion of invariable sites model

$$\hat{L} = p_{\text{invar}} L_{\text{invar}} + (1 - p_{\text{invar}}) L_{\text{variable}}$$

- Discrete gamma among-site rate het. model

$$\hat{L} = \left(\frac{1}{4}\right) L_{r_1} + \left(\frac{1}{4}\right) L_{r_2} + \left(\frac{1}{4}\right) L_{r_3} + \left(\frac{1}{4}\right) L_{r_4}$$

(Number of categories $k=4$ in this case)

Pagel and Meade's 2004 mixture model

$$L = w_1 L_{Q_1} + w_2 L_{Q_2} + w_3 L_{Q_3}$$

ζC

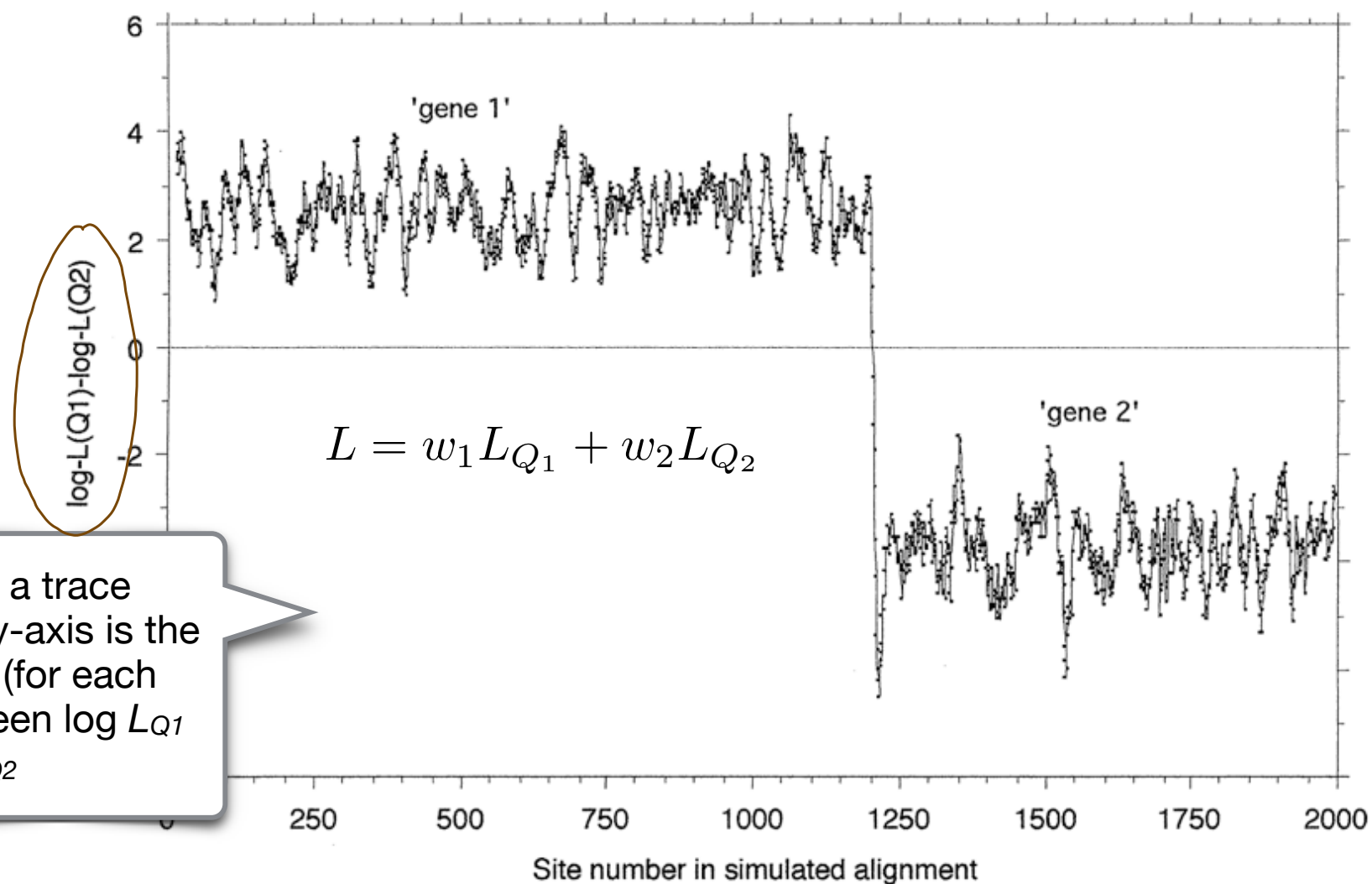
 $15KY$

$$\begin{pmatrix} - & a_1\pi_C & b_1\pi_G & c_1\pi_T \\ a_1\pi_A & - & d_1\pi_G & e_1\pi_T \\ b_1\pi_A & d_1\pi_C & - & f_1\pi_T \\ c_1\pi_A & e_1\pi_C & f_1\pi_G & - \end{pmatrix}
 \quad
 \begin{pmatrix} - & a_2\pi_C & b_2\pi_G & c_2\pi_T \\ a_2\pi_A & - & d_2\pi_G & e_2\pi_T \\ b_2\pi_A & d_2\pi_C & - & f_2\pi_T \\ c_2\pi_A & e_2\pi_C & f_2\pi_G & - \end{pmatrix}
 \quad
 \begin{pmatrix} - & a_3\pi_C & b_3\pi_G & c_3\pi_T \\ a_3\pi_A & - & d_3\pi_G & e_3\pi_T \\ b_3\pi_A & d_3\pi_C & - & f_3\pi_T \\ c_3\pi_A & e_3\pi_C & f_3\pi_G & - \end{pmatrix}$$

Base frequencies are *common* to all Q matrices.

(BayesPhylogenies software at <http://www.evolution.reading.ac.uk/SoftwareMain.html>)

Pagel-Meade 2004 mixture model

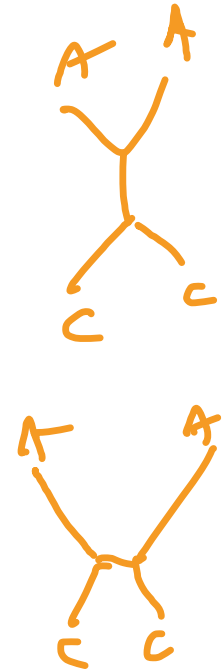
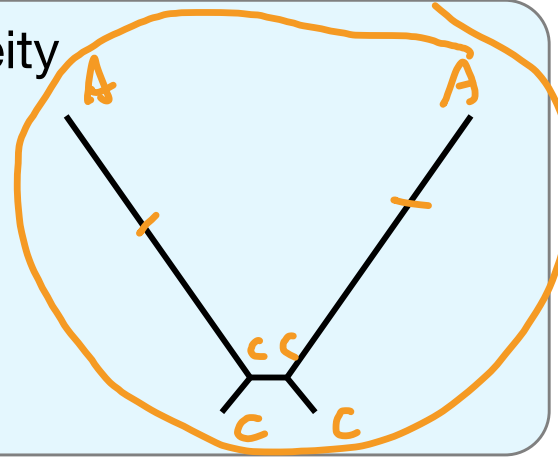
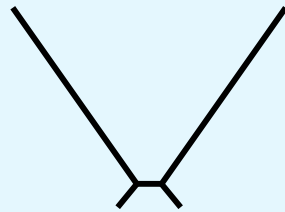
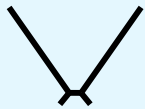


This is *not* a trace plot - the y-axis is the difference (for each site) between $\log L_{Q_1}$ and $\log L_{Q_2}$

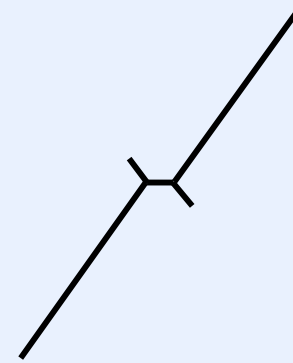
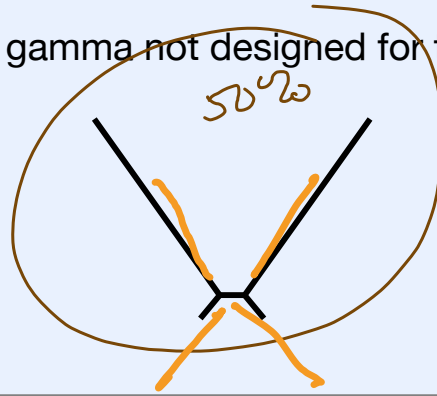
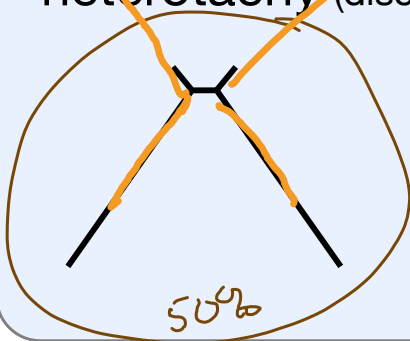
Heterotachy

Heterotachy

classical among-site rate heterogeneity



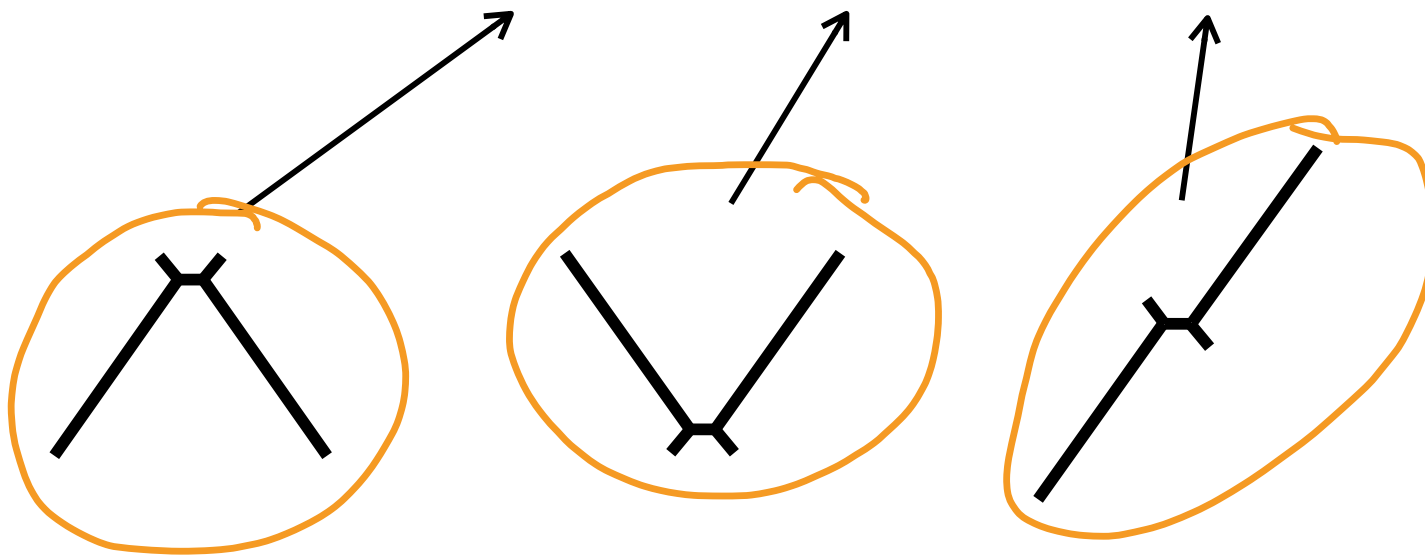
heterotachy (discrete gamma not designed for this)



Kolaczkowski and Thornton (2004)

Kolaczkowski & Thornton's (2008) heterotachy model

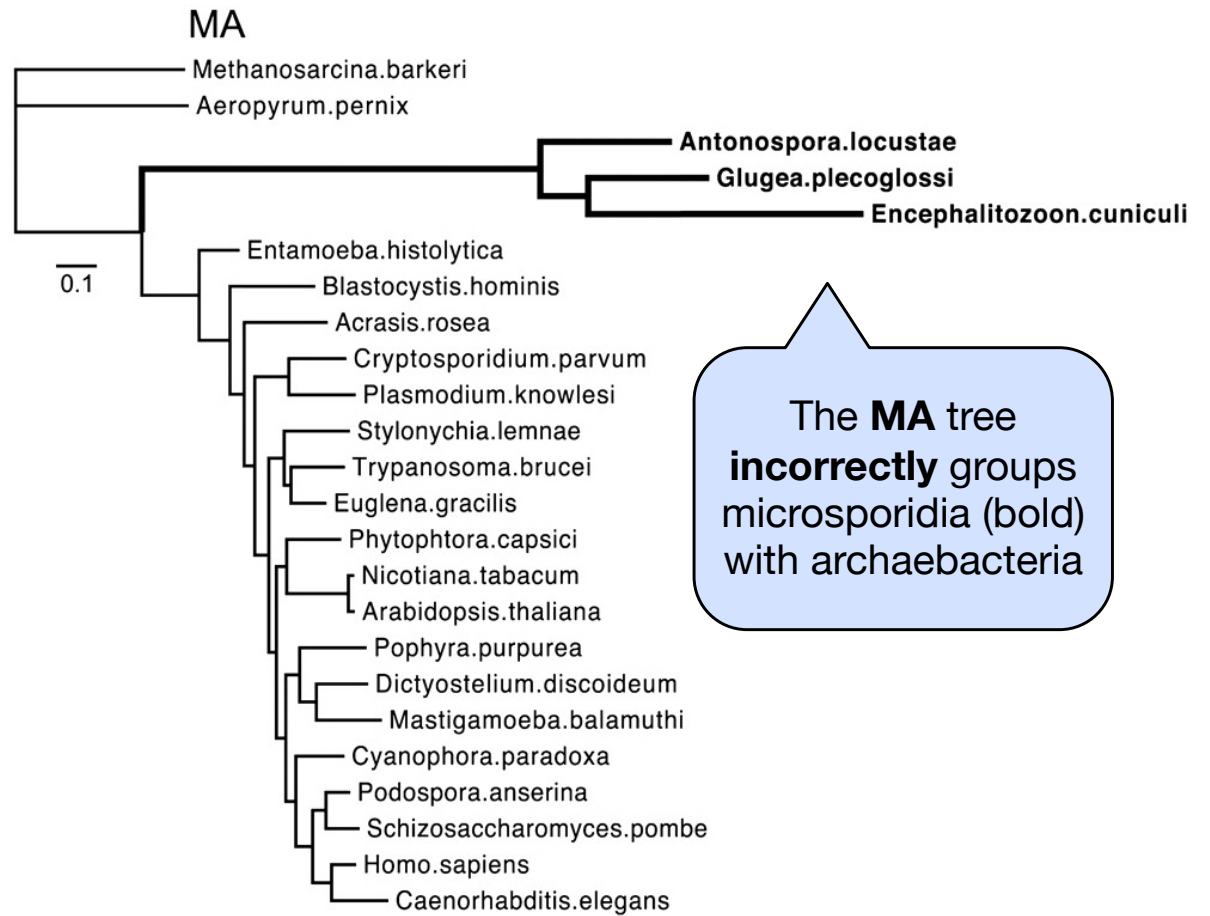
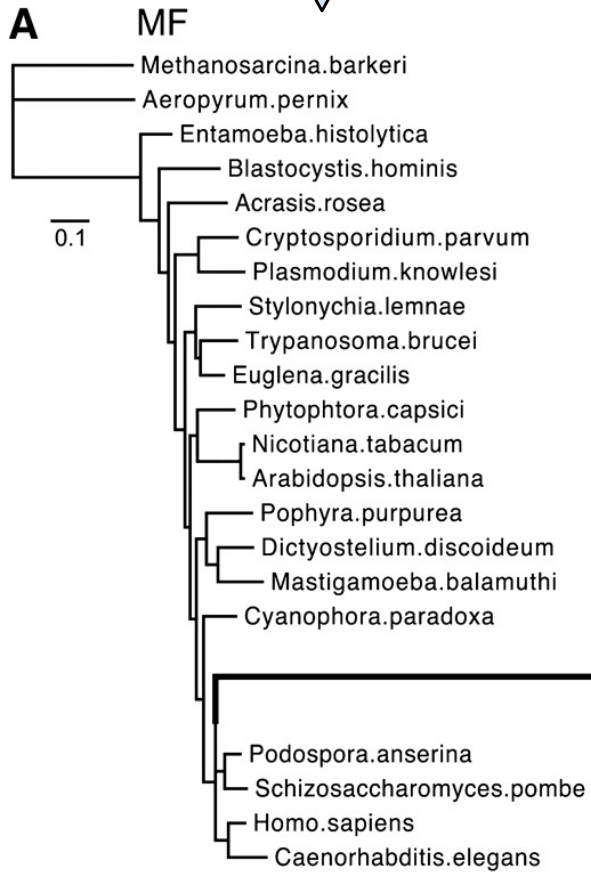
$$L = w_1 L_{b_1} + w_2 L_{b_2} + w_3 L_{b_3}$$



Each b_i represents an *entire set* of 5 branch lengths.
This model potentially adds a *lot* of parameters (if the tree is large)

Kolaczkowski and Thornton (2008)

The **MF** tree correctly groups microsporidia (bold) with fungi



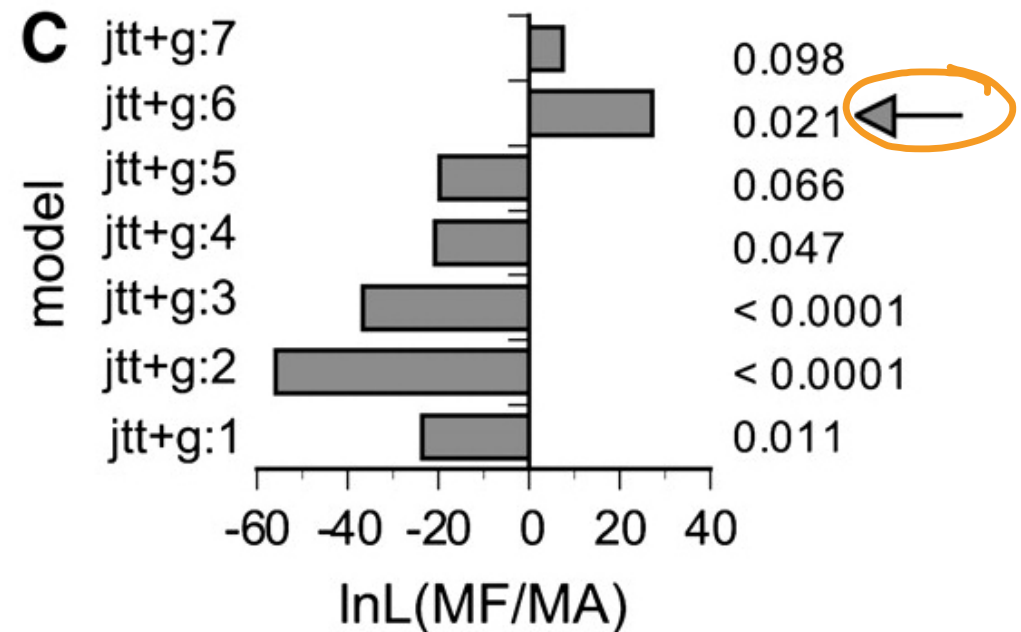
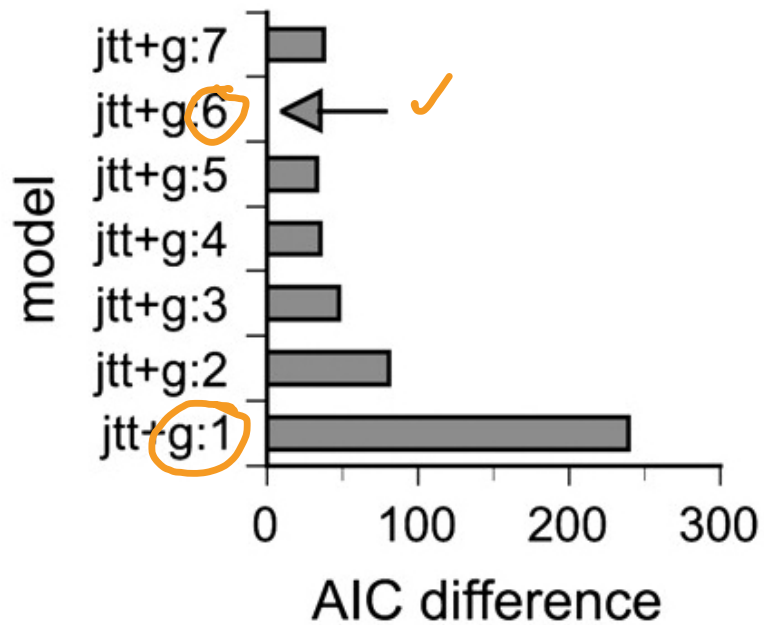
The **MA** tree incorrectly groups microsporidia (bold) with archaeobacteria

Heterotachy

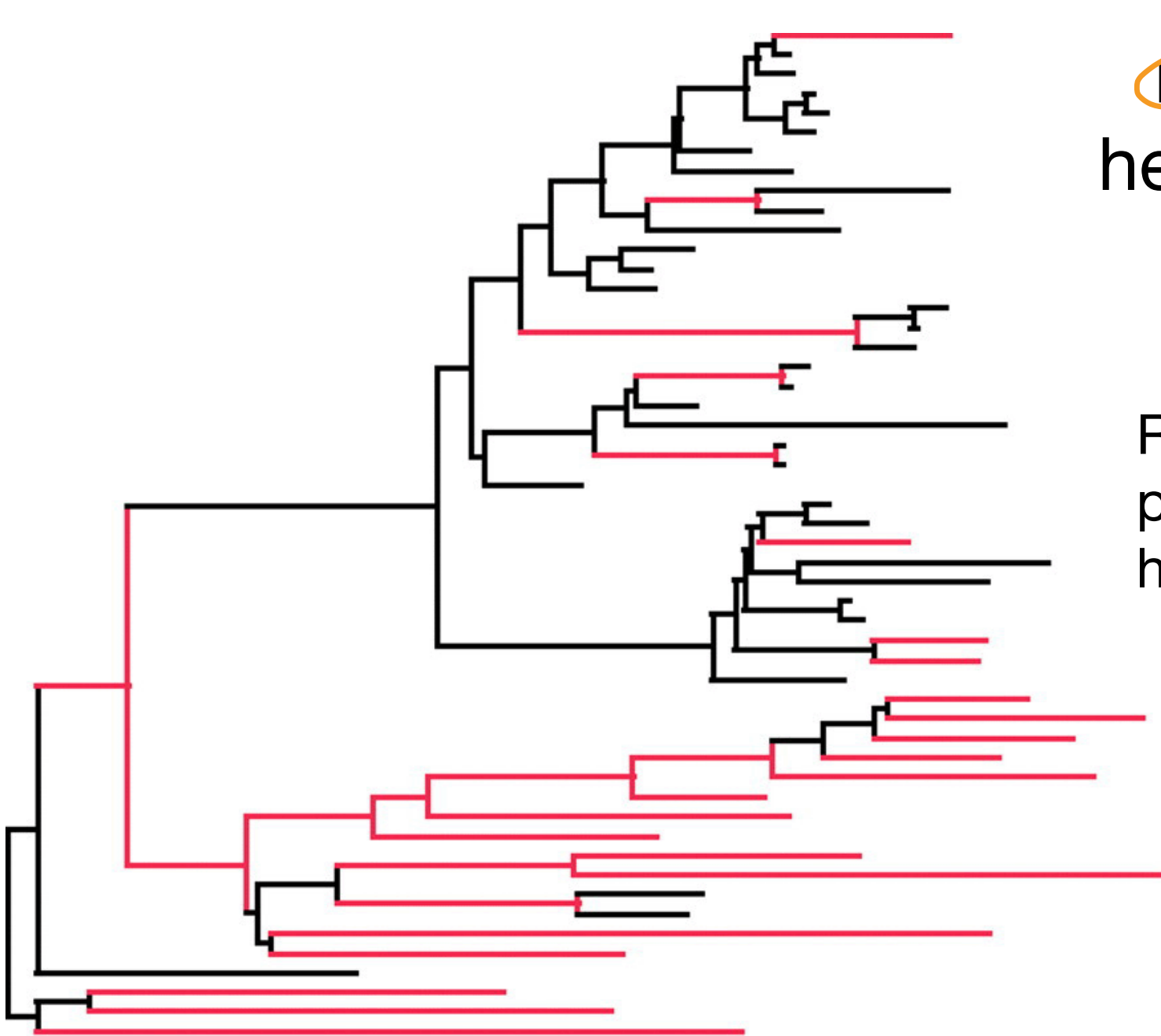
Heterotachy

A mixture of 6 sets of branch lengths yields the minimum AIC score

A mixture of 6 sets of branch lengths yields maximum discrimination between the widely-recognized true tree (MF) and the artifactual tree (MA)



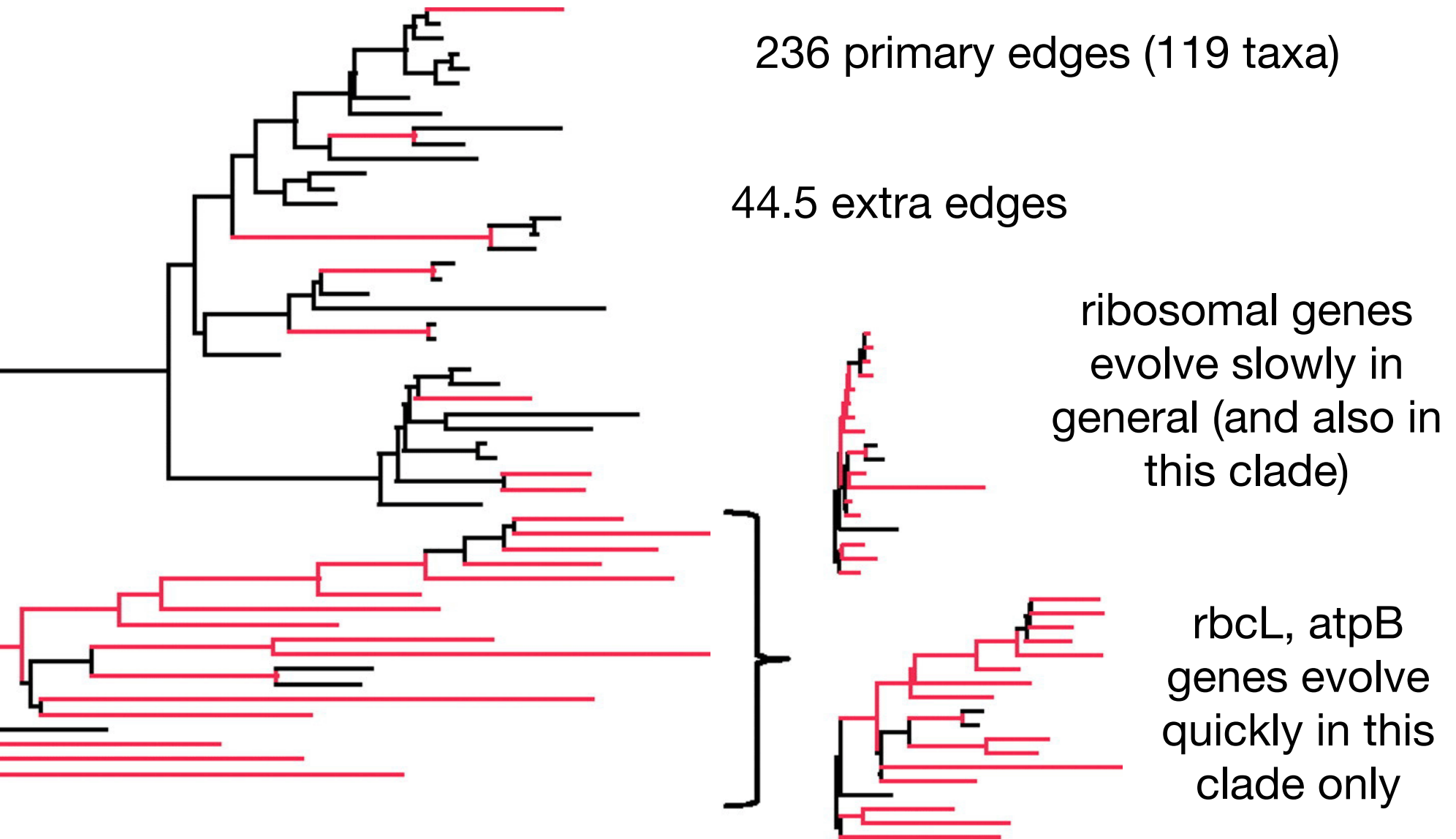
Kolaczkowski and Thornton (2008)



rjMCMC
heterotachy
model

Red edges have
prob. > 0.5 of
having 2 edges

rjMCMC heterotachy model



Software: <http://www.evolution.reading.ac.uk/BayesPhyHeterotachy.html>

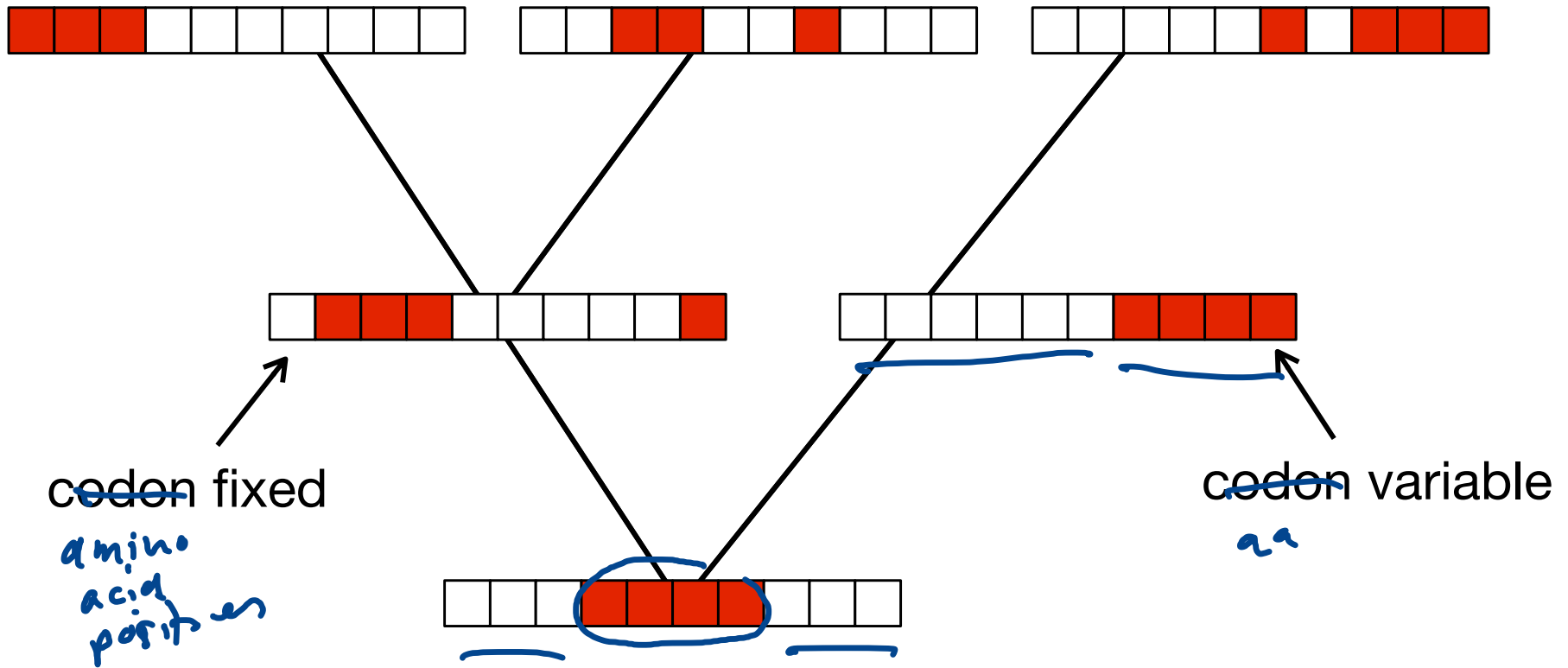
Covarion Models

Covariation Hypothesis

*"...at any one point in time only a very **restricted number of positions can fix mutations** but that as mutations are fixed the **positions capable of accepting mutations also change** so that examination of a wide range of species reveals a wide range of altered positions. We define this restricted group as the **concomitantly variable codons.**"*

Fitch and Markowitz (1970)

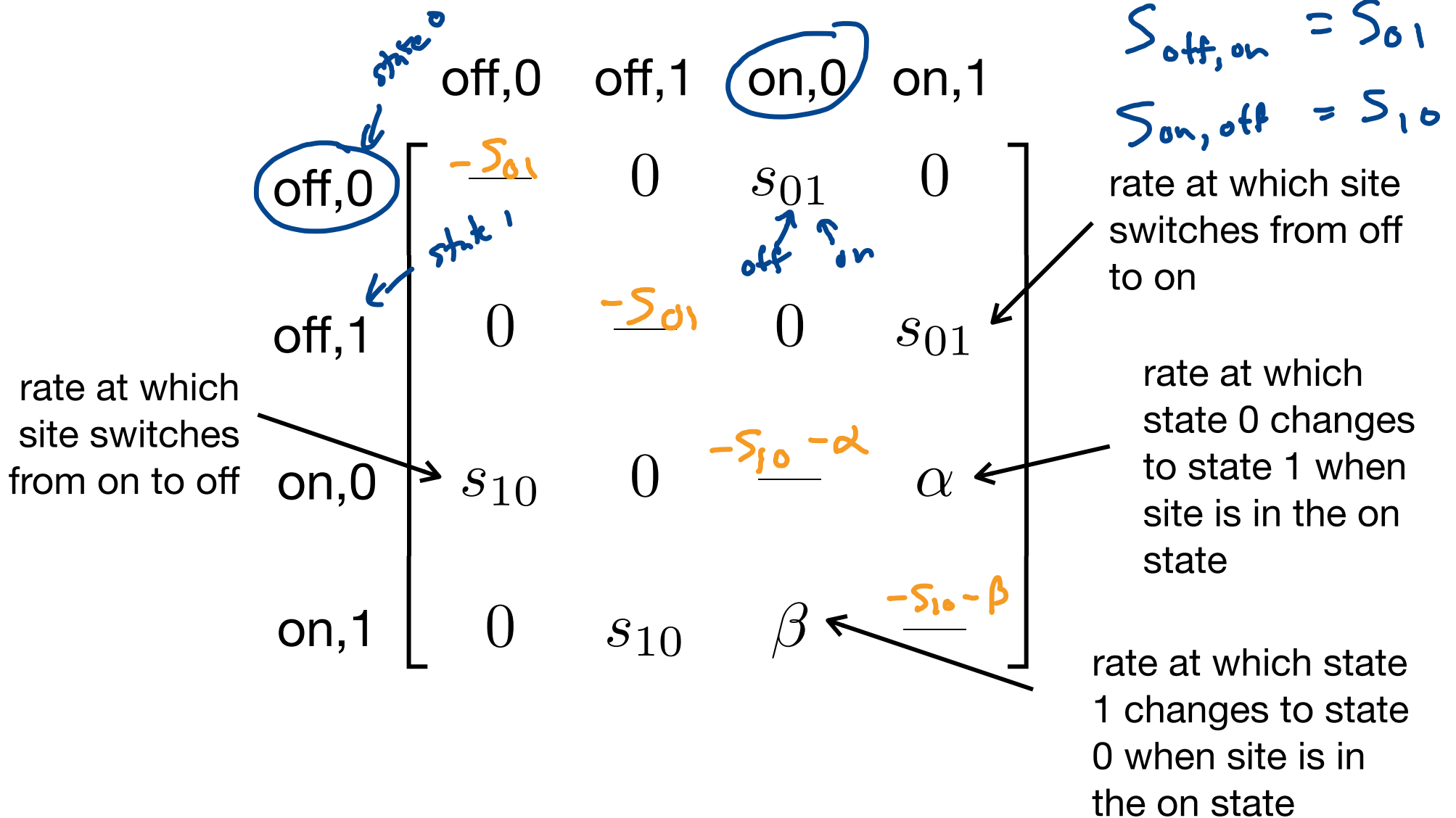
Covariation Hypothesis



Extreme form of heterotachy

Fitch and Markowitz (1970)

Covarion Model

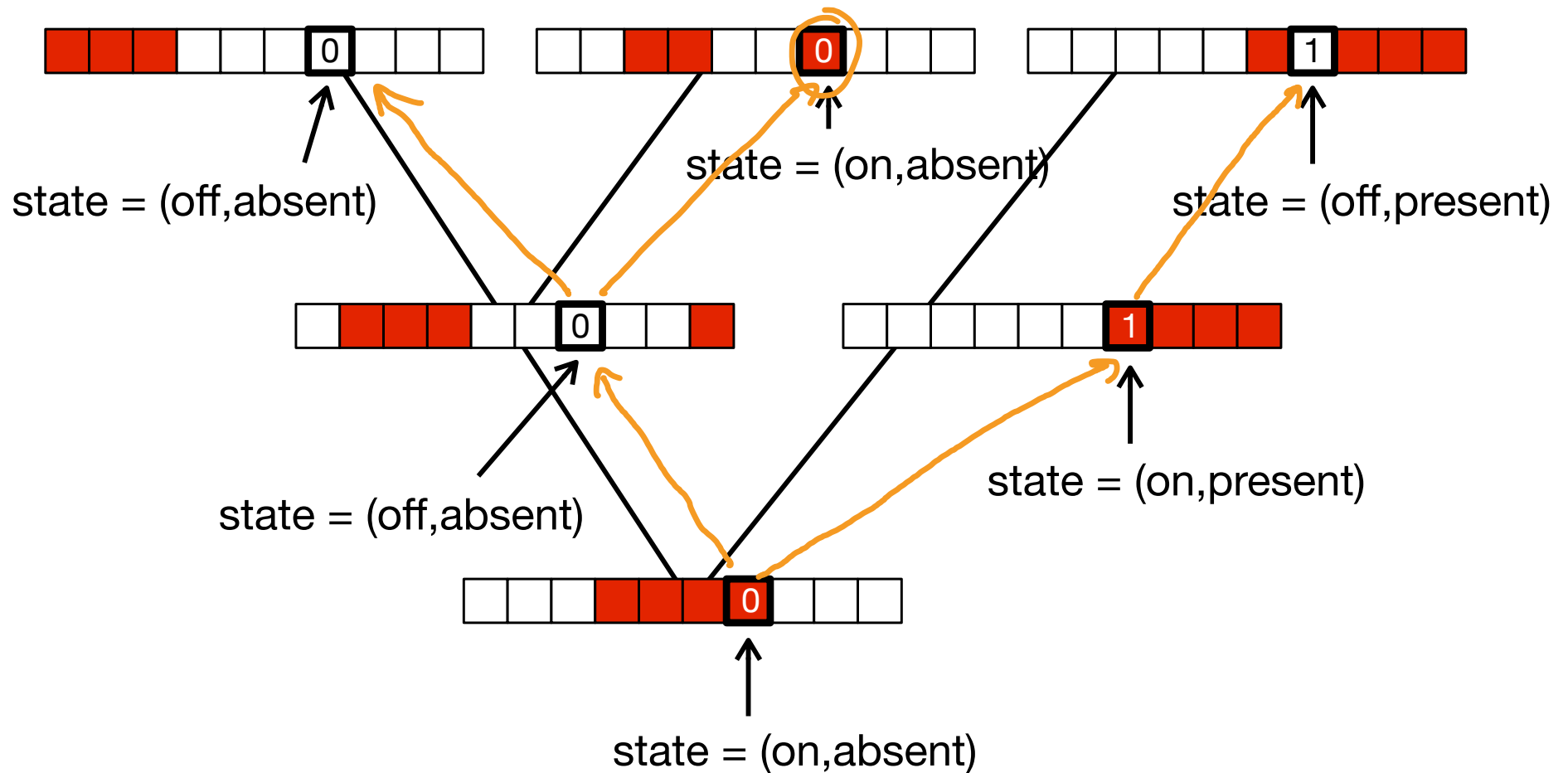


Tuffley and Steel (1998)

Covarion Model

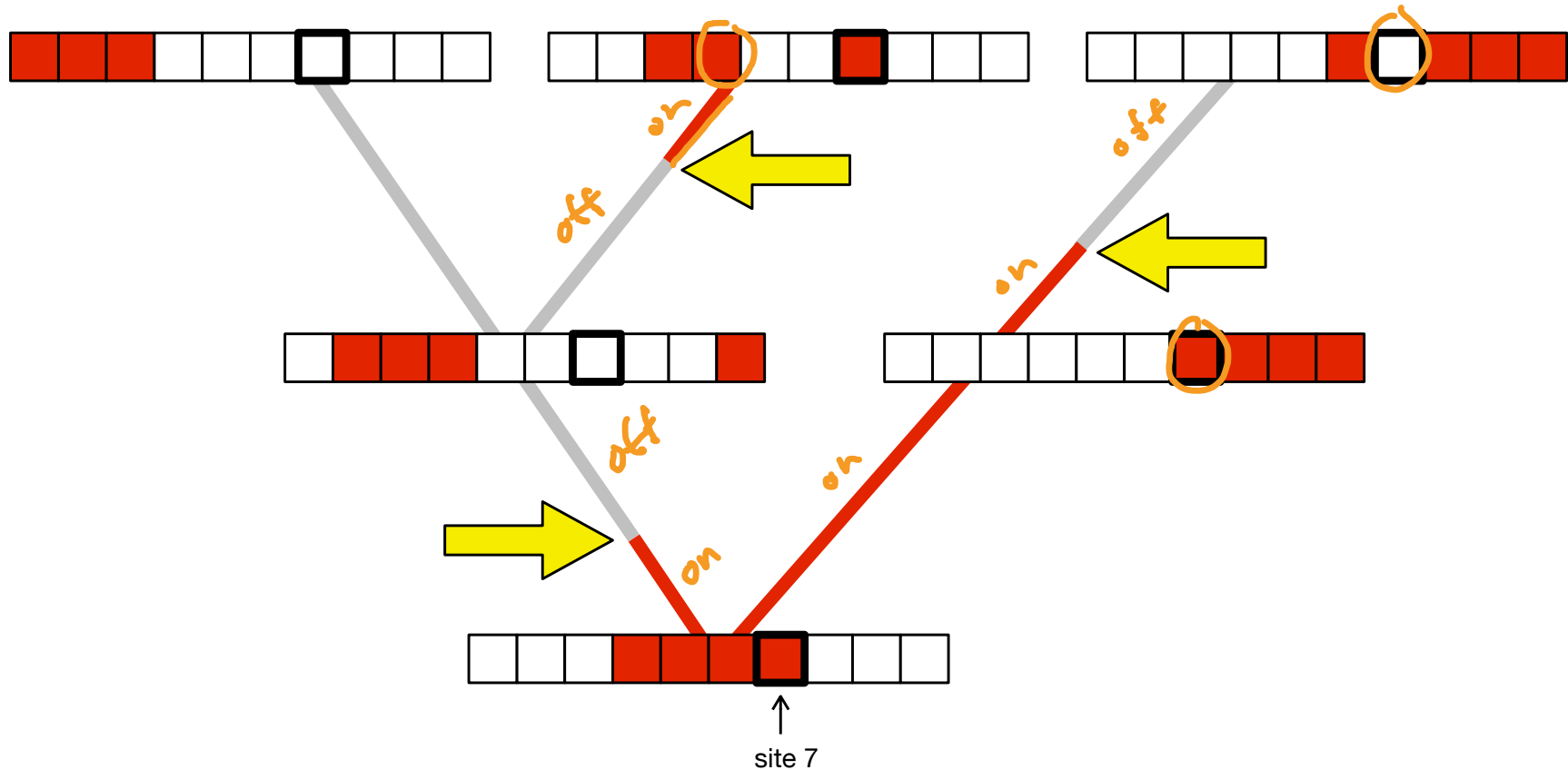
Pretend there are 10 presence/absence characters, each of which can be in an on or off state at any point in time..

Consider only site 7 for the moment (bold boxes)...



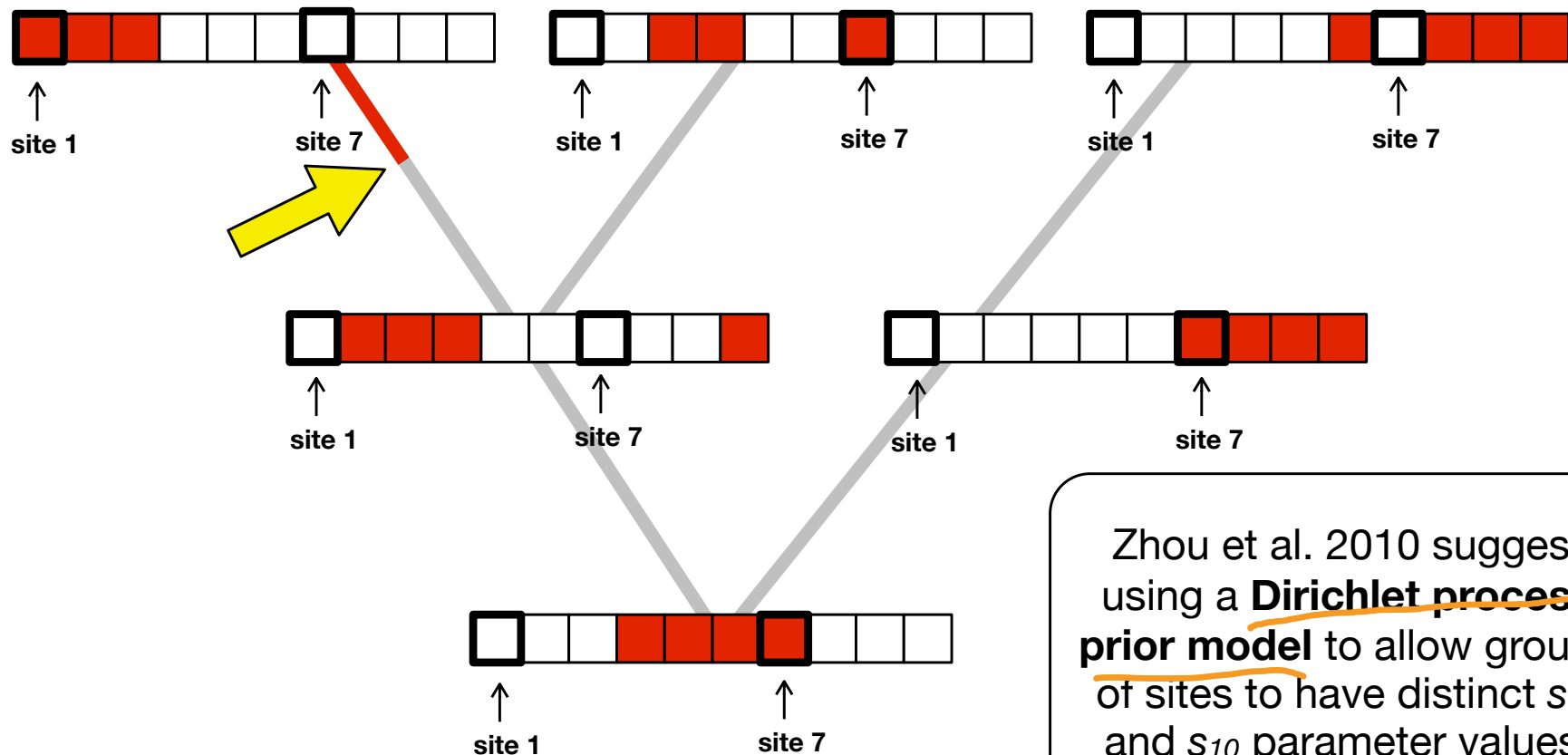
Covariation Mixture Model

Note that some sites switch from off to on (or vice versa) often (e.g. site 7)...



Covariation Mixture (CM) Model

Note that some sites switch from off to on (or vice versa) often (e.g. site 7), while other sites rarely switch (e.g. site 1).



Software:

<http://www.phylobayes.org/>

Zhou et al. (2010)