

Likelihood ratio test for positive selection

Two nested models:

Model 0 no positive selection

(H0: ω is always ≤ 1)

Model 1 allows positive selection

(H1: $\omega > 1$ for some sites or in certain lineages)

LRT statistic: $2\Delta\ell = 2(\ell_1 - \ell_0) \sim \chi^2_{d.f.}$

$d.f.$ = difference in numbers of parameters

Modeling selection variability

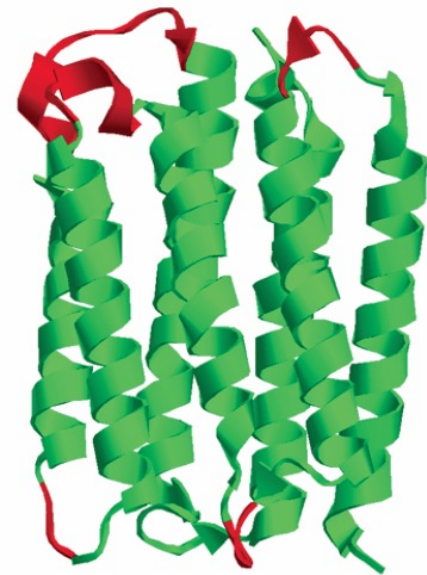
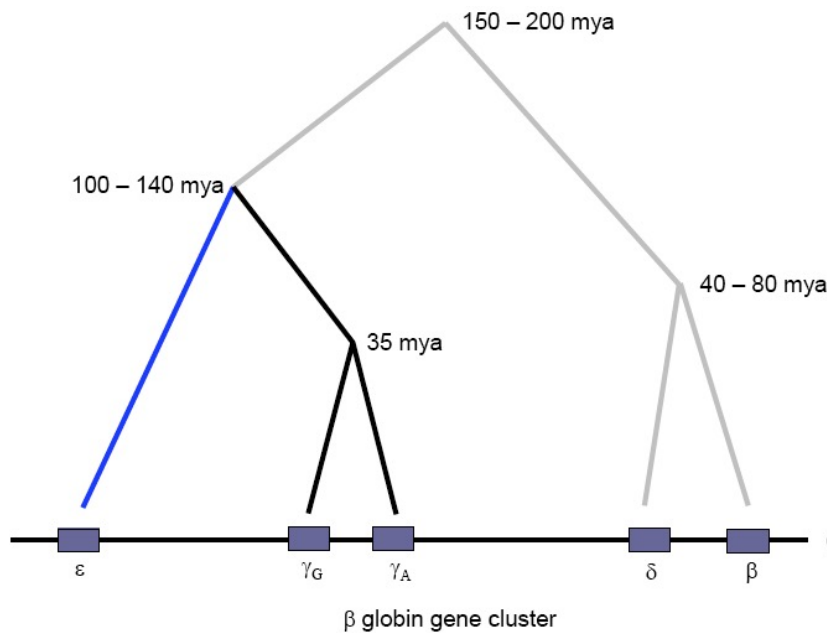
Assuming *constant selective pressure* across the whole sequence and over the whole phylogeny renders the *power of the test low*

e.g., Endo et al (1996) detected only 17 out of 3595 analyzed genes to be under selection

Positive selection usually affects:

only in a few lineages/branches

only few codon sites



Modeling selection variability

By modeling variable ω over time and across sites
we can study:

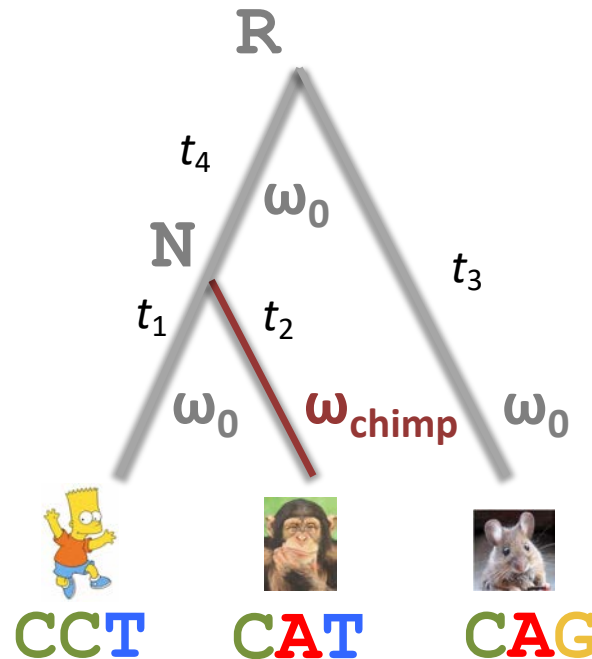
WHEN (in which lineages) did positive selection occur?

WHERE in the sequence did positive selection occur?

Modeling variability over time

Assign independent ω parameters to different branches on the tree:

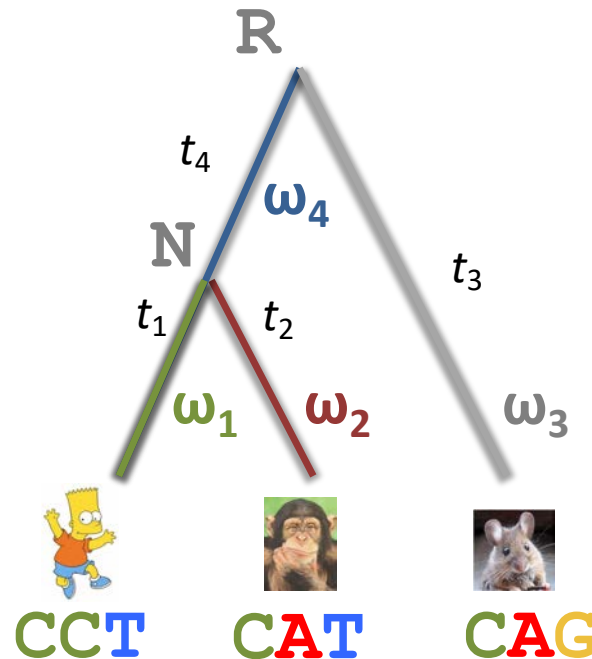
$$L_h = \sum_R \pi_R p_{R \rightarrow CAG}(t_3 | \omega_0) \sum_N p_{R \rightarrow N}(t_4 | \omega_0) p_{N \rightarrow CCT}(t_1 | \omega_0) p_{N \rightarrow CAT}(t_2 | \omega_{\text{chimp}})$$



Modeling variability over time

Assign independent ω parameters to different branches on the tree:

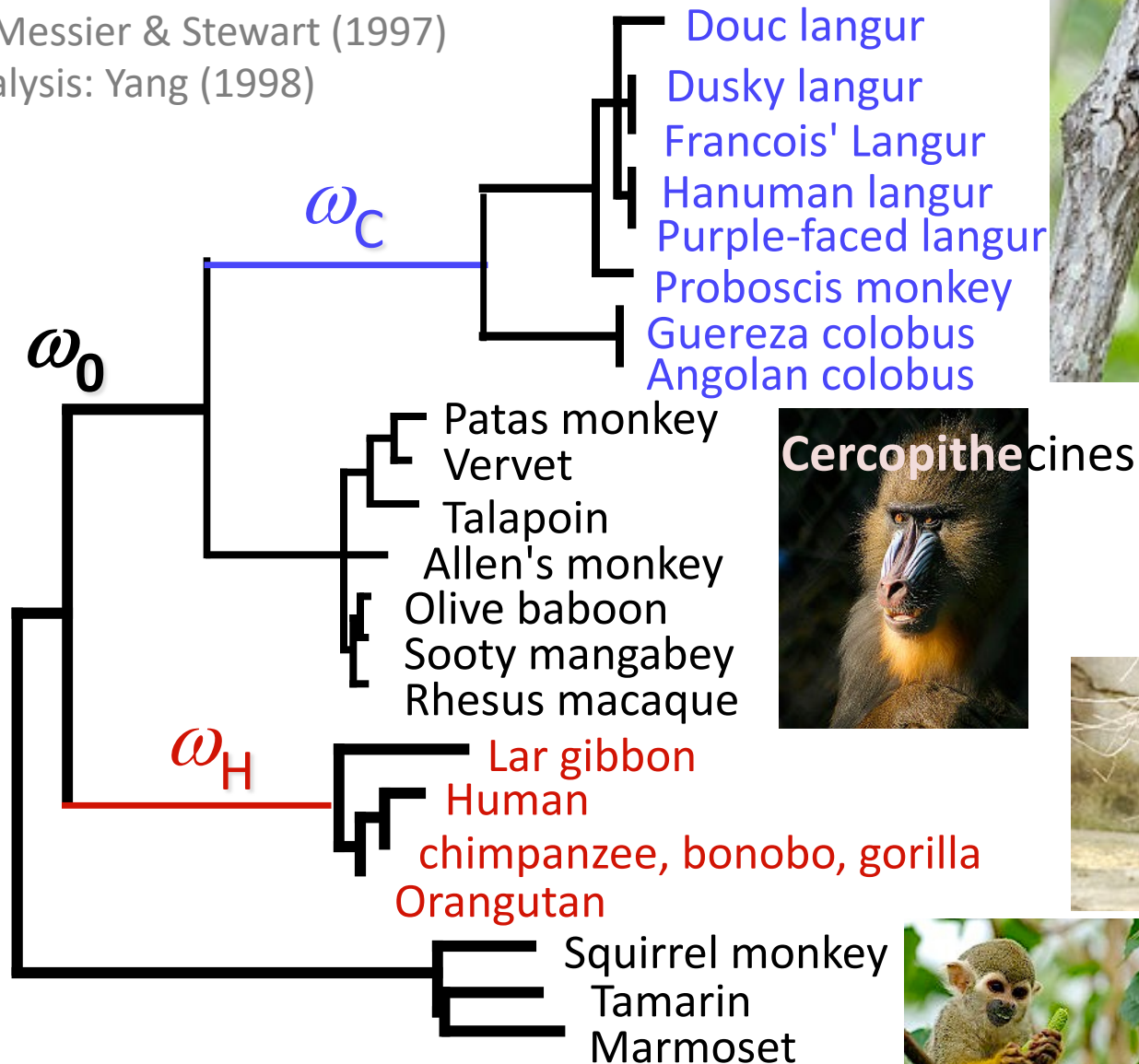
$$L_h = \sum_R \pi_R p_{R \rightarrow CAG}(t_3 | \omega_3) \sum_N p_{R \rightarrow N}(t_4 | \omega_4) p_{N \rightarrow CCT}(t_1 | \omega_1) p_{N \rightarrow CAT}(t_2 | \omega_2)$$



Adaptive evolution in primate lysozyme: ω variability over time

Data: Messier & Stewart (1997)

Re-analysis: Yang (1998)



Primate lysozyme: ML estimates

Model	p	ℓ	ω_0	ω_C
A. 1-ratio: $\omega_0 = \omega_C$	35	-1043.84	0.574	$= \omega_0$
B. 2-ratios: ω_0, ω_C	36	-1041.70	0.489	3.383
C. 2-ratios: $\omega_0, \omega_C = 1$	35	-1042.50	0.488	1 (fixed)

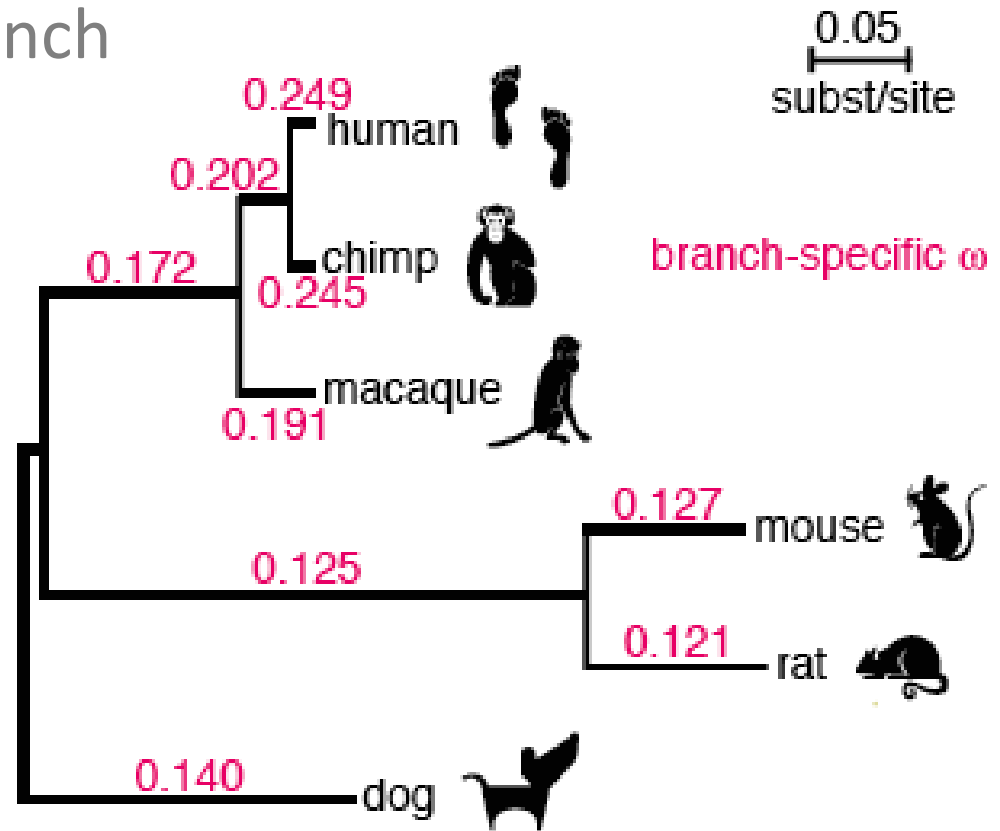
LRT

Null hypothesis	$2\Delta\ell$	d.f.
$\omega_C = \omega_0$	4.24*	1
$\omega_C = 1$	1.60	1

Free ω -ratio LRT with branch model

H_0 : one ω for all branches

H_1 : different ω for each branch



Free ω -ratio LRT with branch model

H_0 : one ω for all branches

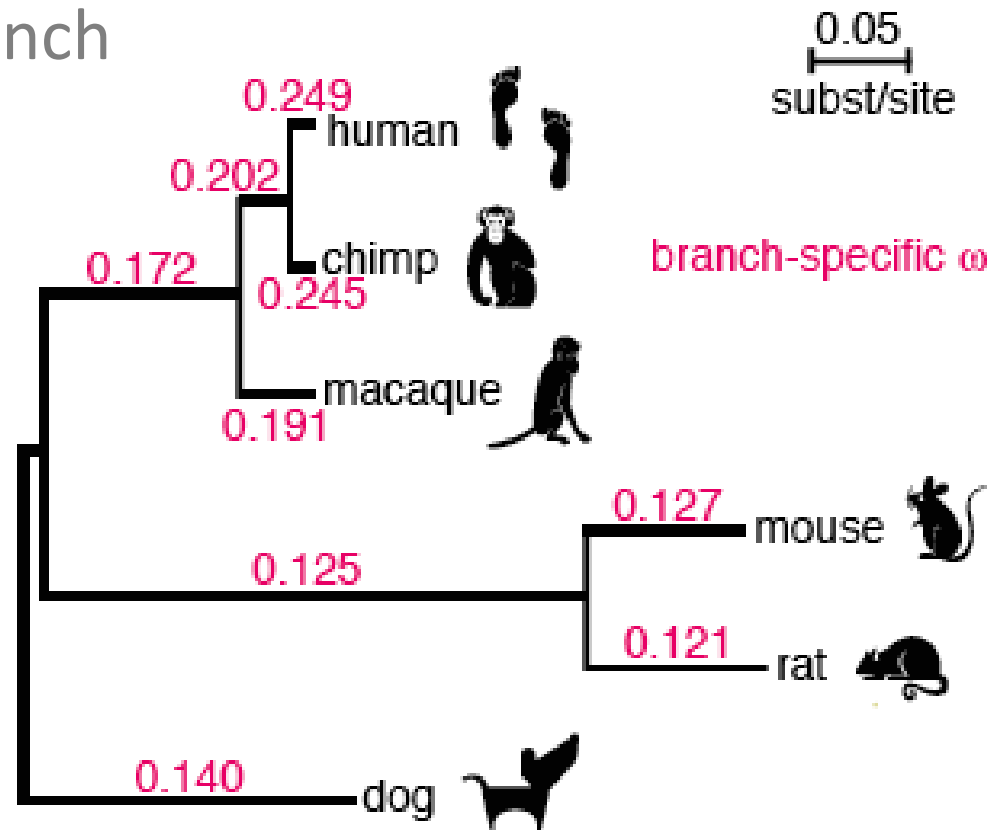
H_1 : different ω for each branch

#branches (for unrooted tree with T leaves):

$$2T-3$$

$$\text{d.f.} = (2T-3) - 1 = 2T - 4$$

Here: d.f. = 8



Exercises with codeml

Focus of exercise #2:

ML estimation with branch models