February 6, 2009**. Maximum Likelihood Principles, and Applications to Discrete Characters**

I. Introduction

Maximum likelihood (ML) principles provide a fundamental alternative to parsimony in the reconstruction of phylogenies and estimation of ancestral states. ML also represents an important shift in thinking from standard probabilistic statistics. In standard statistics we focus on the probability of a given observation under a null hypothesis. If the actual observations are considered very unlikely, then we reject the null hypothesis. However, we don't actually accept a particular alternative hypothesis. For example, consider a t-test of the following observations:

Treatment A: 5, 8, 10 (mean = 7.7, sd = 2.5)
Treatment B: 8, 12, 15 (mean = 11.7, sd = 3.5)

Null hypothesis:
• Assume that observations represent a finite sample from a normal distribution (in other words, the process in the natural world that generates these data would generate a normal distribution if you collected an infinite sample)
• assume that the means and variances of those distributions are equal in the two samples
• if these assumptions are true, the probability of drawing two samples that differ by as much or more as the two above is $Pr(d|H_0) = 0.18$.
• note that we can describe the full distribution of possible outcomes, in terms of the differences between the two groups, which will be a normal distribution that sums to 1.

Maximum likelihood reverses the entire process. Let's assume that our data are real and true, and they reflect the outcome of some unknown process or model. Can we calculate the likelihood of the model, given these data, and compare that to the likelihood of alternative models. We are searching for the maximum likelihood model – the model of the world that best fits the data. The problem is that there are an infinite number of possible models, so unlike probability we can't describe the entire likelihood space as a distribution that sums to 1. So how can we calculate their relative likelihoods? The fundamental insight (Edwards 1972) that makes ML statistics possible is that:

$$L(m \mid d) \propto Pr(d \mid m)$$

Thus, we can obtain relative likelihoods of alternative models and compare them. One of those alternatives could be the traditional null hypothesis, leading to the same significance test, but in general the ability to specify a range of alternatives enhances our ability to explore specific hypotheses with the data. One of the main drawbacks is that it can be quite difficult in some cases to find the best model under ML, if there is no analytic solution.

Likelihood estimation of ancestral states and rates of character evolution on a phylogeny:

We'll start with one of the simplest problem in phylogenetics: estimating the rates of character evolution for a binary trait. First consider the problem of evolution along a single branch:

α = the instantaneous forward transition rate from 0 -> 1
β = the instantaneous reverse transition rate from 1 -> 0

With a little calculus one can show that the probabilities of change along a branch of length t are:

	From:	
To:	0	1
0	$P_{00} = 1 - P_{01}$	$P_{10} = \dfrac{\beta}{\alpha + \beta}\left(1 - \exp\left[-(\alpha + \beta)t\right]\right)$
1	$P_{01} = \dfrac{\alpha}{\alpha + \beta}\left(1 - \exp\left[-(\alpha + \beta)t\right]\right)$	$P_{11} = 1 - P_{10}$

If one assumes that the backward and forward transition rates are the same, P01 and P10 also are the same and simplify considerably.

For example:
$\alpha = 0.5, \beta = 0.5, t = 1$

To: From:	0	1
0	0.684	0.316
1	0.316	0.684

For example:
$\alpha = 0.8, \beta = 0.2, t = 1$

To: From:	0	1
0	0.494	0.126
1	0.506	0.874

For example:
$\alpha = 0.8, \beta = 0.0, t = 1$

|	0	1
0	0.449	0.0
1	0.551	1.0

The first REALLY important thing about maximum likelihood view of evolutionary change is that branch lengths matter (unlike parsimony). Given instantaneous rates of change, α and β, a branch will eventually converge on a probability α that it ends in state 1 and probability β that it ends in state 0, regardless of the initial state.

E.g. $\alpha = 0.8, \beta = 0.2$
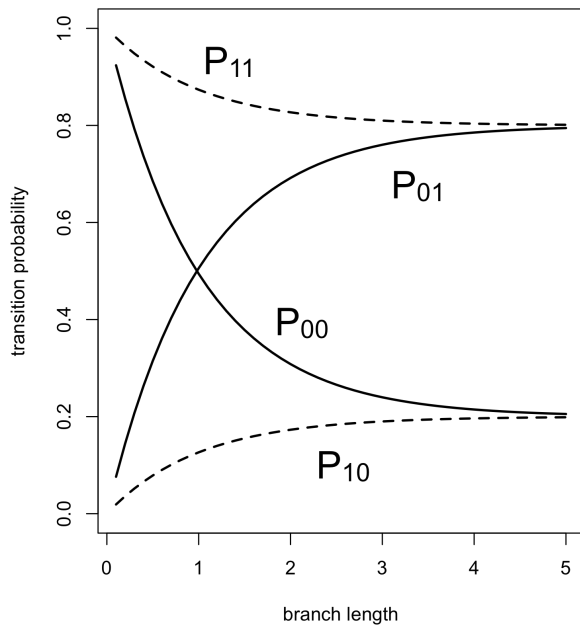
If both rates are lower, but similar ratio to each other, the branches will converge to the same point, but it will take longer. For $\alpha = 0.4, \beta = 0.1$
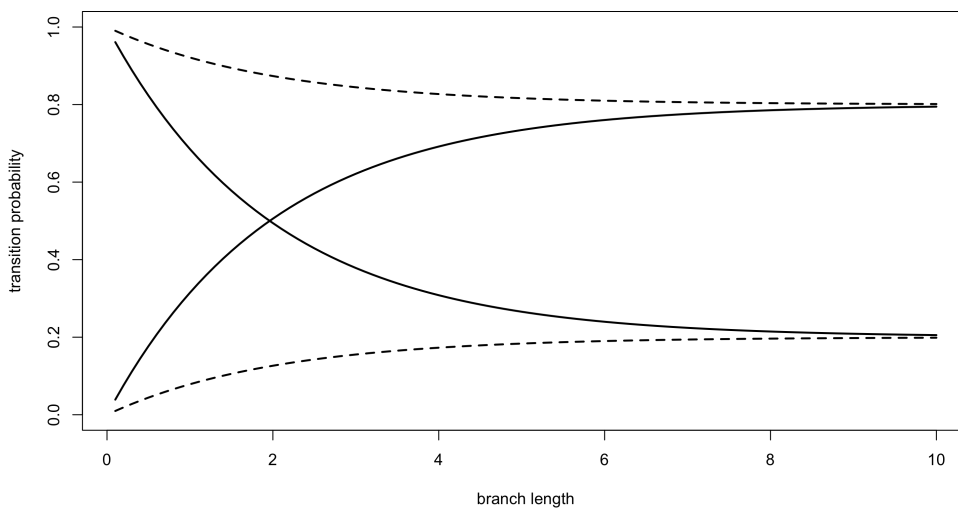
ML estimation of transition rates and ancestral states

1. Calculate likelihood for all possible rates over all possible combinations of ancestral states

2. Sum probabilities over each ancestral set; max value = ML estimate of transition rate

3. Now, using that rate, evaluate likelihood of each ancestral set

4. Sum likelihoods for individual states at each node overall all possible values at other nodes; normalize to sum of 1 to get likelihood of state 0 and 1 at each internal node

phylogeny and species trait values
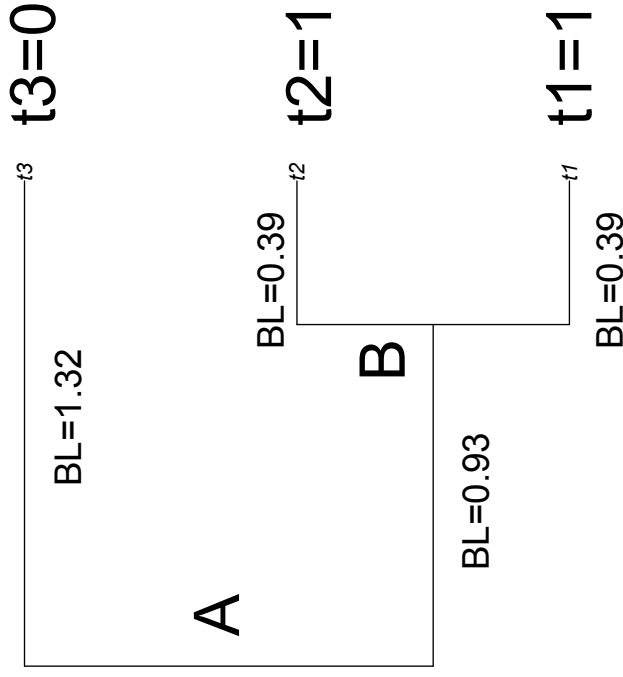
What is the likelihood of any one set of ancestral states given a hypothesized transition rate (assume equal forwards and backwards rates):

Let A = 0, B = 1, alpha = beta = 0.3

overall likelihood of this combination is the product of the individual likelihoods on each branch:

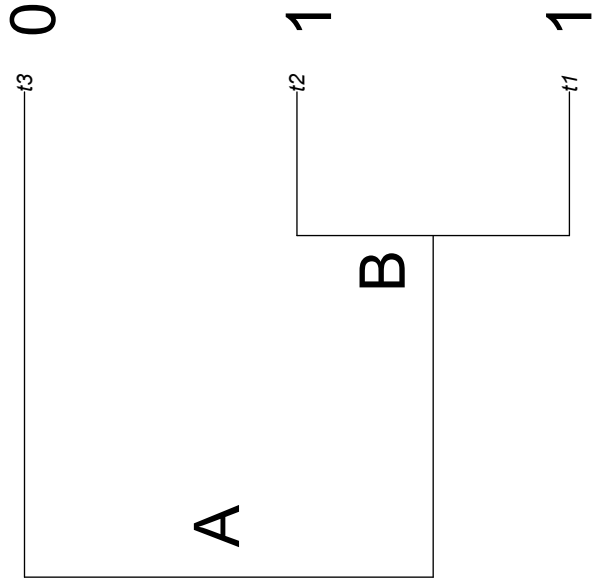$A \rightarrow B = P_{01, 0.93} = 0.213$
$B \rightarrow t1 = P_{11, 0.39} = 0.896$
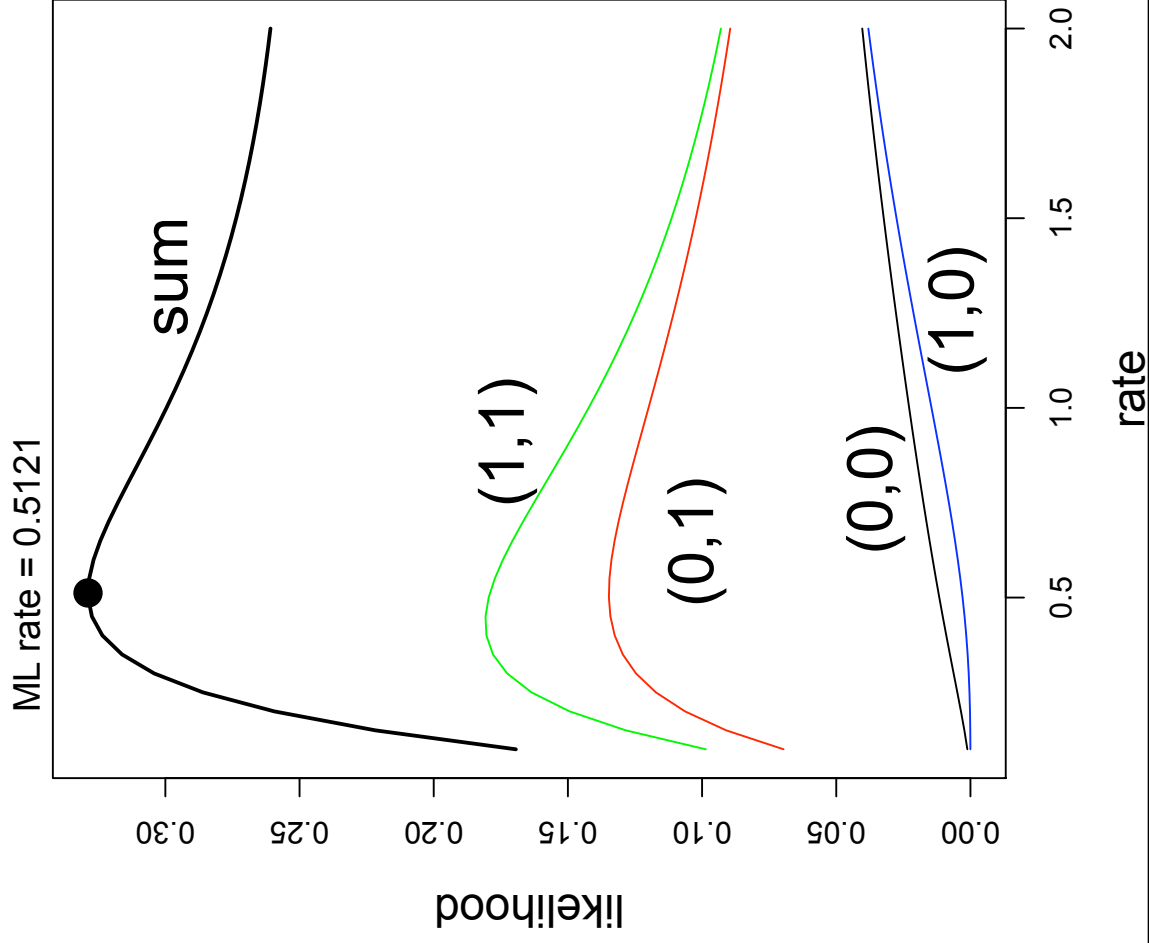$B \rightarrow t2 = P_{11, 0.39} = 0.896$
$A \rightarrow t3 = P_{00, 1.32} = 0.726$

prod = 0.1246

t3=0

BL=1.32

t2=1

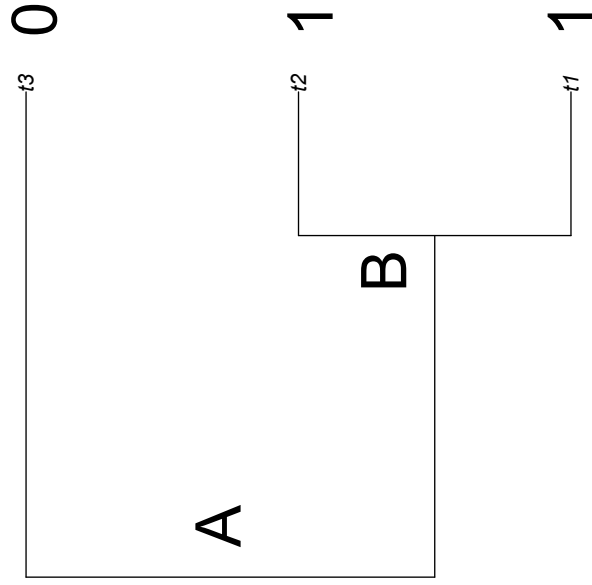BL=0.39

B

t1=1

BL=0.39

BL=0.93

A

phylogeny and species trait values

likelihood of transition rates depending on hypothesized ancestral node values (A,B)

phylogeny and
species trait values

given ML rate = 0.5121, what are
probabilities of each combination of
ancestral states (A,B)

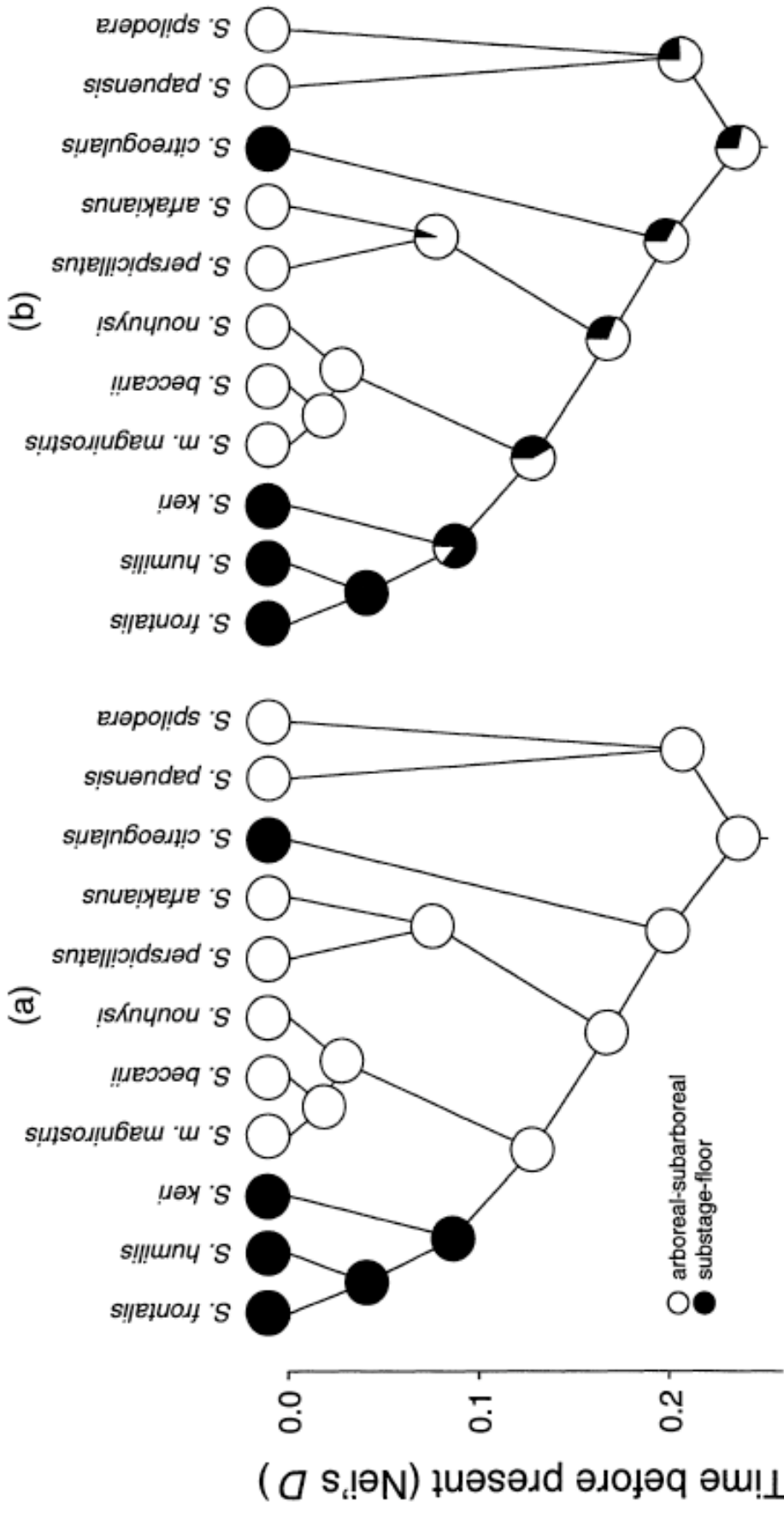| A \ B	0	1	P(A)	L(A) normalized
0	0.0057	0.1218	0.127	0.429
1	0.0005	0.1693	0.169	0.571
P(B)	0.0062	0.291		
L(B) normalized	0.021	0.979		

A

B

t3 0

t2 1

t1 1

FIG. 3. Vertical foraging stratum in 11 scrubwrens (*Sericornis*). The tree is from a UPGMA analysis of Nei's (1978) genetic distances calculated on allozyme frequencies (Christidis et al. 1988). Tree structure was corroborated by a reanalysis of mtDNA restriction sites in five of these species using RESTML (Felsenstein 1993; data from Joseph and Moritz 1993). (a) Ancestor states estimated using maximum parsimony (Maddison and Maddison 1992). (b) Maximum-likelihood ancestor states assuming that all transition rates q are equal; both q̂ = 5.93. Area of pies at nodes in (b) indicate relative support for each ancestral state.

Schluter et al. 1997 Evolution

Likelihood ratio tests

a measure of support for alternative hypotheses

$LR = -2*\ln(L_1/L_2)$

For two hypotheses with the same number of parameters, there is no exact significance value attached to the LR. Values greater than 2 are considered 'strong support'

For nested hypotheses with different numbers of parameters, LR is distributed as a chi-square with df = the difference in number of parameters

For example:
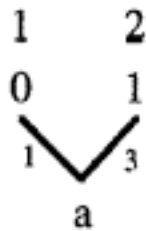
If we find the maximum likelihood with alpha and beta fit independently:

al = 0.59
be = 0.31
L(m) ∝ 0.256.

If we allow only one transition rate, such that al=be, then:

al = be = inf
L ∝ 0.25

LR = -2*ln(.25/.256) = 0.05
chisq(0.05, df=1) = 0.82

So these data are insufficient to reject a single rates model.

Pagel's (1994) discrete test of correlated evolution:

Same idea as above, but test for parameters of dependence in trait change.

$$
I + Q_{\mathrm{D}} =
$$

X, Y	$0, 0$	$0, 1$	$1, 0$	$1, 1$
$0, 0$	$1 - (q_{12} + q_{13})$	q_{12}	q_{13}	0
$0, 1$	q_{21}	$1 - (q_{21} + q_{24})$	0	q_{24}
$1, 0$	q_{31}	0	$1 - (q_{31} + q_{34})$	q_{34}
$1, 1$	0	q_{42}	q_{43}	$1 - (q_{42} + q_{43})$

,

For example:
q12 is the probability that trait 2 changes from 0 -> 1, when trait 1 = 0
q34 is the probability that trait 2 changes from 0 -> 1, when trait 1 = 1

For an instantaneous model of change, assume they don't change simultaneously

Model with full dependence has 8 parameters

If traits evolve independently, there are only 4 parameters, because:
q12 = q34; q13 = q24; q31 = q42; q21 = q43

Table 1. *Categories of hypothesis test using* $L(I)$ *and* $L(D)$

(The tests shown in the table do not exhaust the range of possible tests but represents ones likely to be of interest. Tests with more than 1 d.f. are also possible. The contingent change tests, temporal order/relative rate tests, and branch length transformation tests will have asymptotic χ^2 distributions but may also be tested by Monte Carlo simulation. The omnibus test, and tests of alternative models will not in general be distributed as χ^2. The test of branch length transformation can be applied to any likelihood, not just to the full model. Subscripts to $L(D)$ denote the number of parameters.)

test	hypothesis	description	statistic
omnibus χ^2 (4 d.f.)	$L(I) < L(D)$	correlated evolution	$L(I)$ versus $L(D)$
contingent change (1 d.f.)	$q_{12} \neq q_{34}$	change in Y depends upon state of X	$L(D_7)$ versus $L(D_8)$
	$q_{21} \neq q_{43}$	change in Y depends upon state of X	$L(D_7)$ versus $L(D_8)$
	$q_{13} \neq q_{24}$	change in X depends upon state of Y	$L(D_7)$ versus $L(D_8)$
	$q_{31} \neq q_{42}$	change in X depends upon state of Y	$L(D_7)$ versus $L(D_8)$
temporal order or relative rate (1 d.f.)	$q_{12} \neq q_{13}$	order of acquisition of X versus Y	$L(D_7)$ versus $L(D_8)$
	$q_{42} \neq q_{43}$	order of loss of X versus Y	$L(D_7)$ versus $L(D_8)$
alternative models	$q_{ij} = 0$	q_{ij} transitions excluded	$L(D_7)$ versus $L(D_8)$
branch transformation	$\kappa \neq 1$	transformation of branch lengths	$L(D_n)$ versus $L(D_{n+1})$
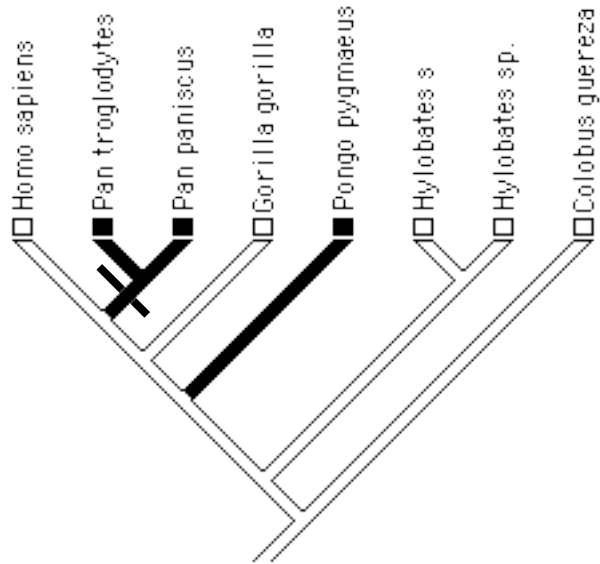
ln(L(independent)) = -11.91
ln(L(dependent)) = -8.43
-2*ln(LI/LD) = 6.96
chisq(6.96,4) = 0.14

q12 = 0.29   gain of oestrous swellings in single-male breeding systems
q34 = 3.45   gain of OS in multi-male breeding systems
q13 = 1.87   gain of multi-male BS in absence of OS

CITATIONS:

Pagel M. (1994) Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc. R. Soc. Lond. B, 255, 37-45

Pagel M. (1999) The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Syst. Biol., 48, 612-622

Schluter D., Price T., Mooers A. & Ludwig D. (1997) Likelihood of ancestor states in adaptive radiation. Evolution, 51, 1699-1711