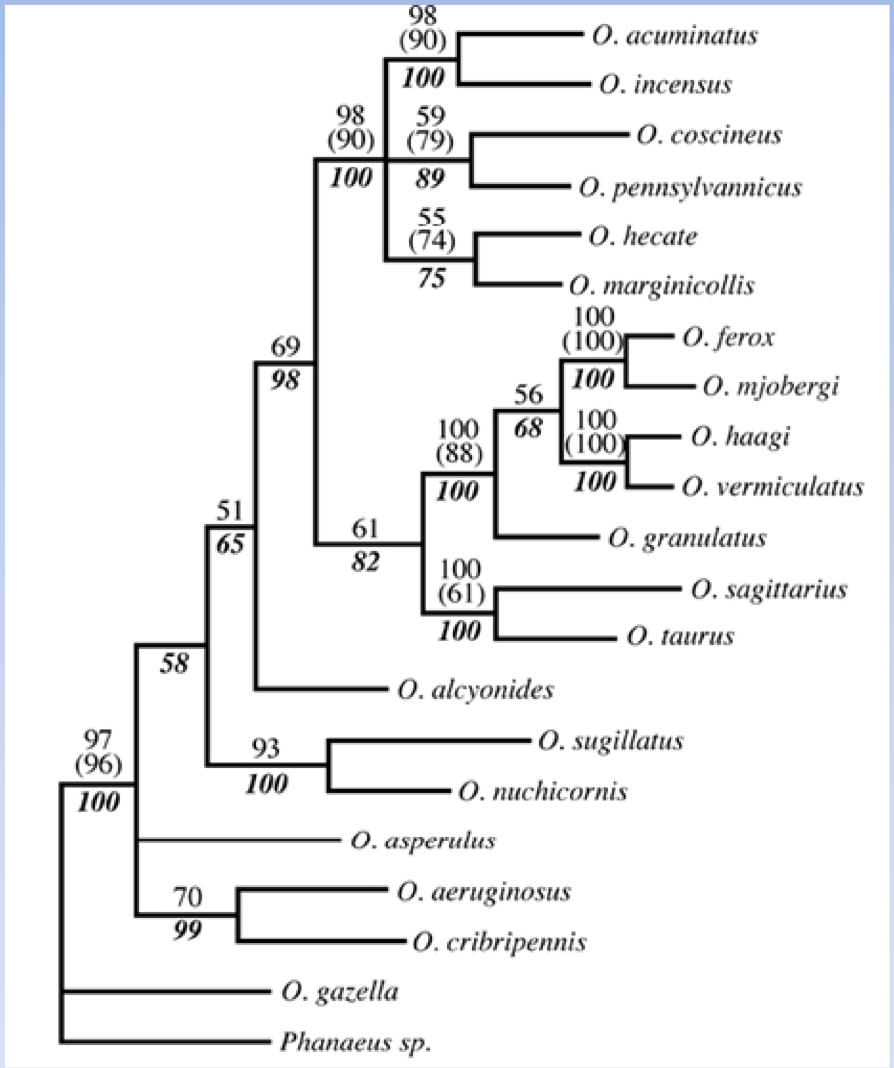
Is there enough signal in the data to favour one hypothesis (evolutionary model) over another?

Pete Lockhart
Thomas Buckley



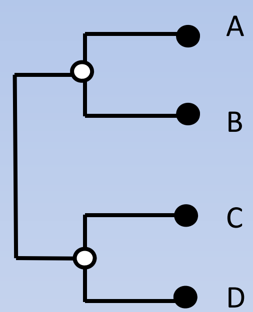
Non-parametric bootstrapping

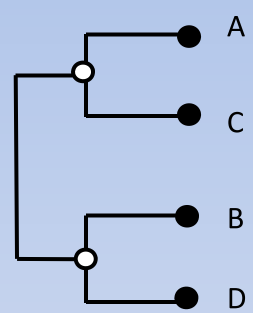
```
          123456
Species 1 CGCTAA
Species 2 CGATTA
Species 3 ATCGTC
Species 4 ATAGAC
```

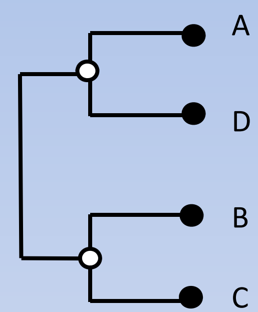
Species 1
Species 2
Species 3
Species 4

Species 1
Species 2
Species 3
Species 4

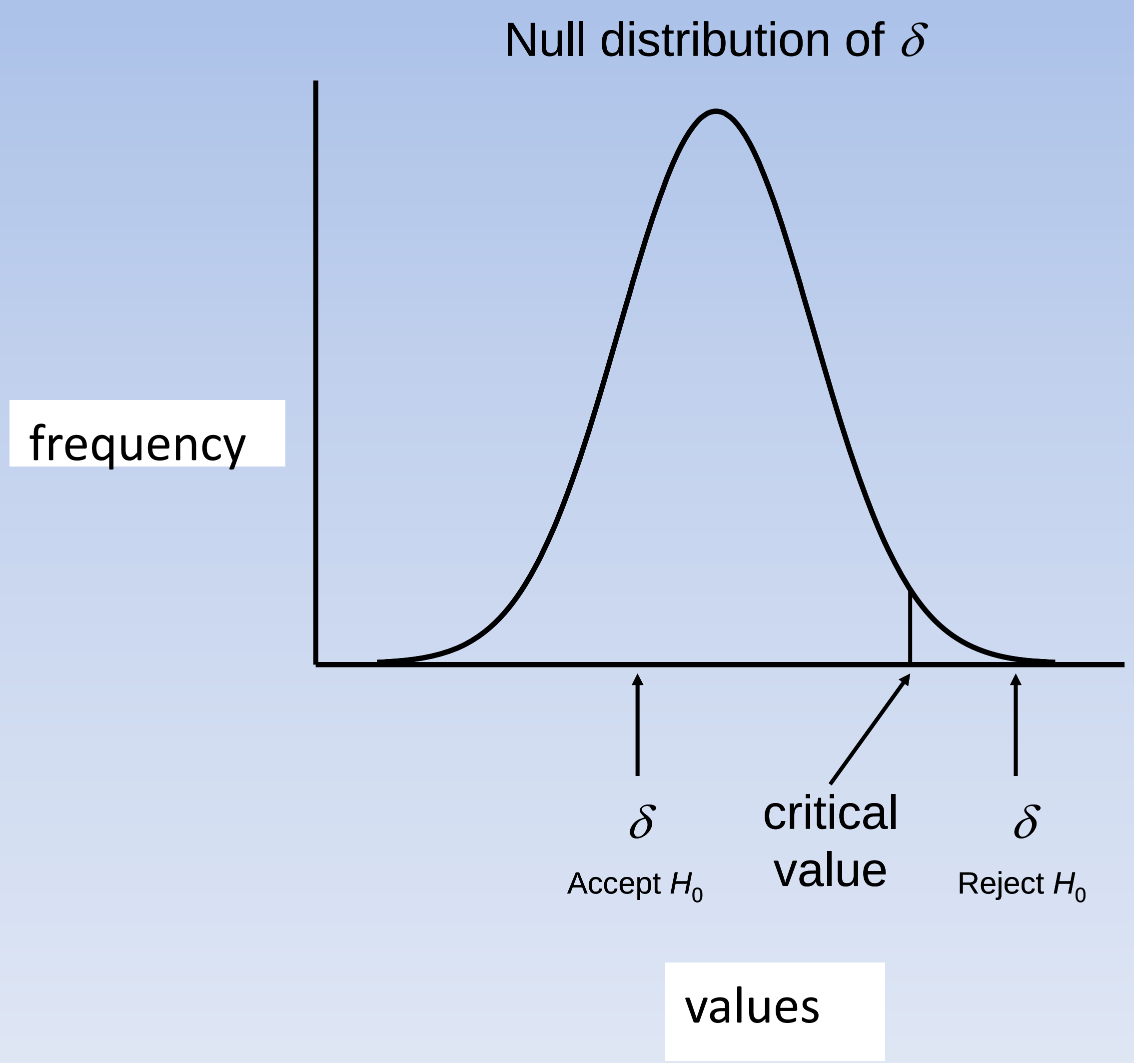
Species 1
Species 2
Species 3
Species 4


A
B
C
D


A
C
B
D


A
D
B
C

Hypothesis testing

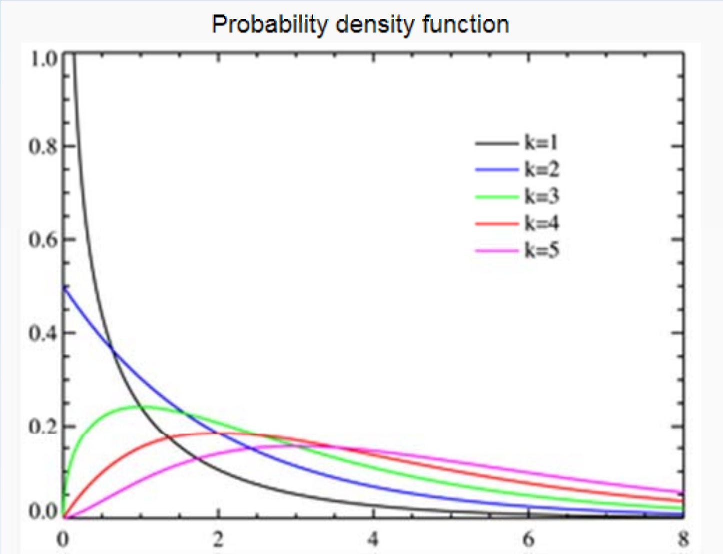


likelihood ratio test

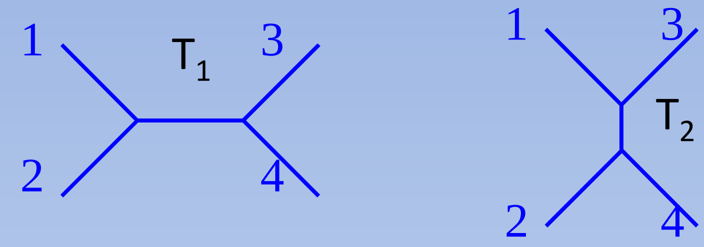
$$\delta = 2(\ln L_1 - \ln L_0)$$

ln L_o is likelihood of a simpler model

ln L_1 is likelihood of a more general model



Kishino-Hasegawa test



H_O: T_1 and T_2 explain the data equally well; i.e. $E[\delta]=0$

- Calculate the value of δ (compare T_1 and T_2) from original data
- Use the nonparametric bootstrap generate pseudoreplicates
- Optimise parameter values for T_1 and T_2 on each pseudoreplicate
- Calculate δ^i for each pseudoreplicate
- Centre the distribution by subtracting the mean of δ^i from each value of δ^i
- If δ lies outside of 2.5% - 97.5% of the ranked distribution of δ^l, then H_O is rejected

Shimodaira-Hasegawa test

The Shimodaira-Hasegaewa (SH) test can be used to simultaneously compare sets of topologies, and like the KH test is a nonparametric test. Unlike the SH test it includes a correction for the multiple comparison of topologies. Under the SH test the hypotheses are:

H_O = That all topologies are equally good explanations of the data

H_A = That some or all of the topologies are not equally good explanations of the data.

The SH test is performed as follows, where T_x is a topology belonging to the set of candidate topologies:

- Calculate the test statistic δ_x for each topology, where δ_x is L_{ML}-L_x.
- Bootstrap the data to generate *n* pseudoreplicates.
- Maximise the likelihood under each T_x for each pseudoreplicate.
- "Centre" the distribution by subtracting the mean of $L^{(i)}$ from each value of $L^{(i)}$.
- For each pseudoreplicate find the maximum likelihood, over all of the topologies in T_x, and subtract this value from likelihood of T_x for that pseudoreplicate.
- Determine whether the value of d_x for each T_x lies between 0% and 95% of the null distribution.; if any T_x falls outside of this distribution then H_0 is rejected.

Approximately unbiased test

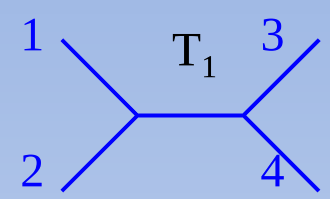
Generates pseudoreplicates that differ in length from the original dataset, in order to explore site-pattern space. This allows more accurate correction for comparing multiple topologies. In the AU test the null hypothesis is framed in terms of *m*, which is the expected value of the likelihood for a tree. The observed likelihood of tree *i* is a random variable drawn from μ.

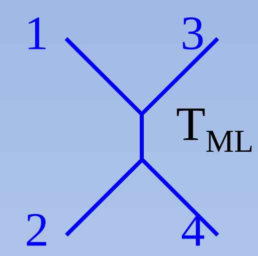
The null hypothesis states that tree *i* has a larger m_i than u_j from tree *j*:
$H_O : E[m_i] \geq [m_j]$
$H_A : E[m_i] < [m_j]$

The AU test is a complex procedure, which we have greatly simplified below:

- Specify a number of scaling constants, *r*, such that , where *n* is the original sequence length and is the modified sequence length.
- For each of the *M* sequence lengths generate a number of pseudoreplicates.
- Calculate *P*-values from the pseudoreplicates.
- Fit a curve through these *P*-values to obtain values for a correction formula.
- Use the correction formula to obtain the AU test *P*-values.

SOWH test





H_O = That T_1 is the true tree
H_A = That some other tree is the true tree

The SOWH test in its most general form is performed as follows:

- Take T_1 and T_{ML} and optimise the branch lengths and all substitution model parameters from the original data. The likelihood ratio, δ is the test statistic.
- Simulate n data sets using T_1 with associated branch lengths and substitution model parameters.
- For each data set, optimise the likelihood on T_1 to yield and maximise the likelihood over topologies, branch lengths and substitution model parameters, to yield .
- Calculate the difference between these two likelihoods, $\delta^{(i)}$, for each data set. If δ is greater than 95% of the ranked values of $d^{(i)}$ then the null hypothesis is rejected.

- There are a number of approximations that we can make that will ease the large computational burden of the SOWH test. The most useful of these is to optimise the parameter estimates for each pseudoreplicate on the null topology, and then fix these parameter values when optimising .

- The SOWH test can be used to compare the ML tree (T_{ML}) to any other specified tree (T_1). In this case, the null hypothesis implies that T_1 is the true tree:

example

In the tree below the optimal topology for our data set has chimpanzee and human as sister species. We have tested two hypotheses, the first that chimpanzee and gorilla are sister species, and the second that human and gorilla are sister species.

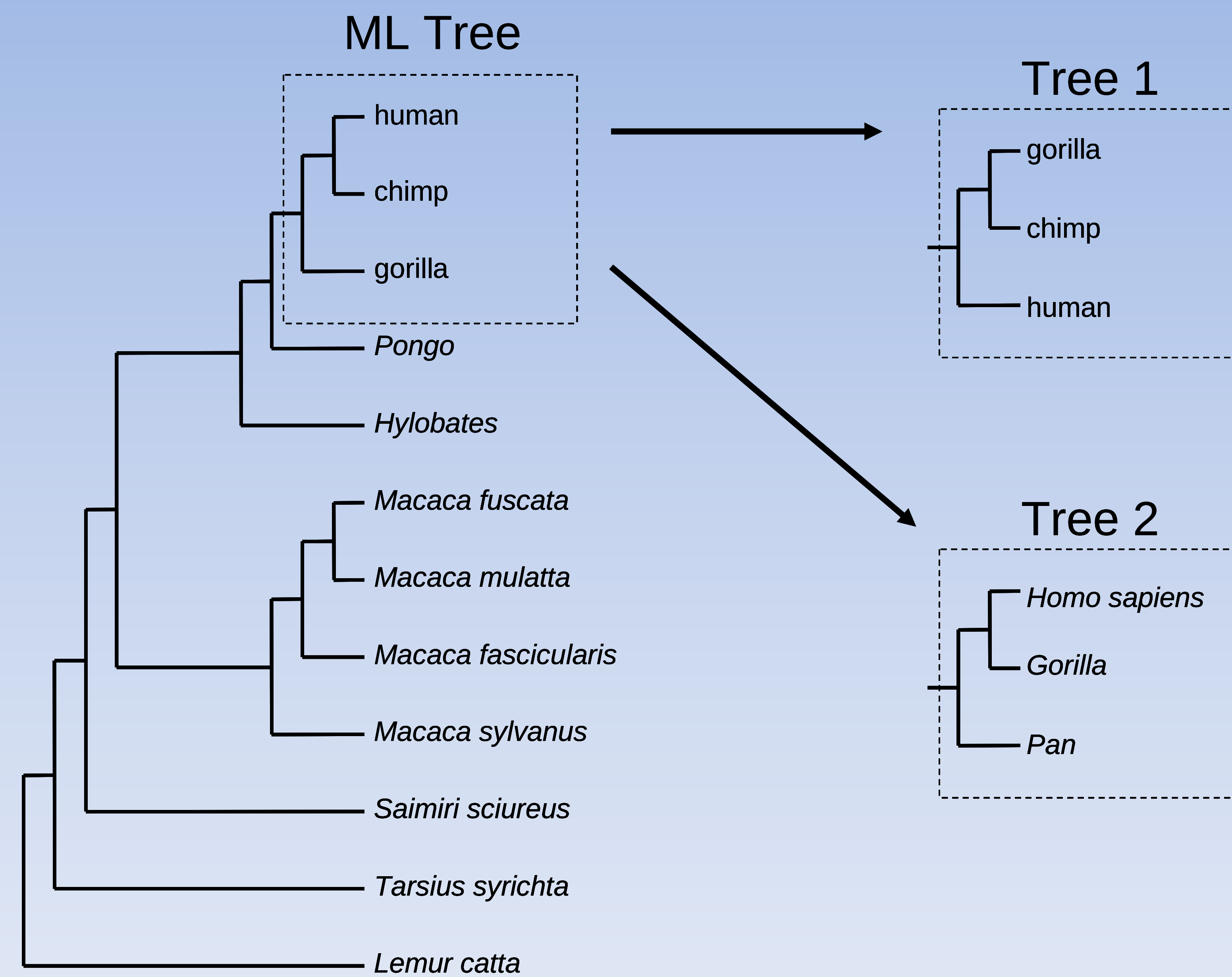
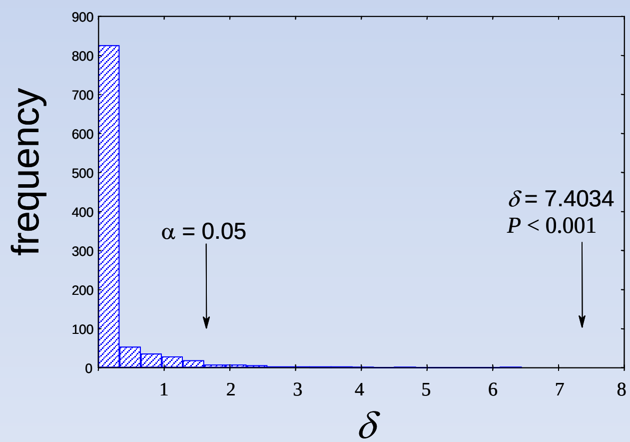

ML Tree
human
chimp
gorilla
Pongo
Hylobates
Macaca fuscata
Macaca mulatta
Macaca fascicularis
Macaca sylvanus
Saimiri sciureus
Tarsius syrichta
Lemur catta
Tree 1
gorilla
chimp
human
Tree 2
Homo sapiens
Gorilla
Pan

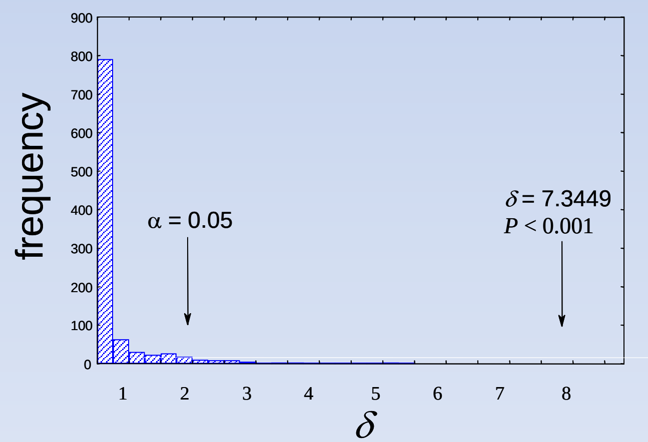
Table 7.2. Test statistic (δ) and P-values from the four tests of topology.

	δ	KH	SH	AU	SOWH
T_1	7.4034	0.096	0.169	0.090	<0.001
T_2	7.3449	0.093	0.171	0.076	<0.001

The test statistic and P-values for the four tests are shown in Table 7.2. The most striking observation from these results is how the SOWH test leads to rejection of the null hypothesis whereas the other tests do not. In figures below we show the null distributions from the two SOWH tests, which demonstrate just how extreme the observed test statistics are.



gorilla + chimp



human + gorilla

Which hypothesis test to use? – Matt Phillips

If just pairwise hypothesis testing and neither is a priori known to be the ML tree, then the KH test

If comparing many topologies simultaneously and the curvature of site-pattern space can be defined, then AU test, otherwise, SH test, which is very conservative and the power to reject H_O is dependent on the number of topologies tested)

SOWH tests the full phylogenetic model (topology, branch-lengths, substitution parameters), so can be difficult to interpret when the object is specifically to compare topologies. If the model is misspecified it will not be conservative enough.

Application in substitution model selection

<u>example:</u>

JC model = -6424.20245   JC+I model = -6277.78093

$\delta = 2(\ln L_1 - \ln L_0) = 2(-6277.78093+6424.20245)$

$\delta = 292.84304$

- since the proportion of null distribution values (mixed chi square distribution) as large as this is less than 0.0001 we reject the null hypothesis (the simpler model) in favour of the more general model