Farris S. J. (2001) Branch lengths do not indicate support - even in maximum likelihood. Cladistics 17: 298-299

“As a simple example, suppose that the data comprise four terminals A–D and 500 informative characters, of which half split the terminals AB/CD and the rest AC/BD. The (undirected) consensus of most parsimonious trees is unresolved, so that the Bremer support (Bremer, 1988, 1994; cf. Farris, 1996) for either split is zero, but the internal branch of either tree has length 250 according to parsimonious optimization (Farris, 1970).” [italics added]
Tests of two trees

- Test the null hypothesis that the differences between two trees (A and B) are no greater than expected from sampling error.

- The simplest test is the 'winning sites' test
  - sums the number of sites supporting tree A over tree B and vice versa (those having fewer steps on, and better fit to, one of the trees).

- If the null hypothesis is true: characters are equally likely to support tree A or tree B
  - a binomial distribution gives the probability of the observed difference in numbers of winning sites.

The Templeton test

- A non-parametric Wilcoxon signed ranks test of the differences in fits of characters to two trees.

- Like the 'winning sites' test but also takes into account the magnitudes of differences in the support of characters for the two trees.

Templeton’s test - an example

Recent studies of the relationships of turtles using morphological data have produced very different results with turtles grouping either within the parareptiles (H1) or within the diapsids (H2) the result depending on the morphologist. This suggests there may be:
- problems with the data
- special problems with turtles
- weak support for turtle relationships.

The Templeton test was used to evaluate these trees and showed that the slightly longer H1 tree found in the constrained analyses was not significantly worse than the unconstrained H2 tree. The morphological data do not allow choice between H1 and H2.

Comparing competing phylogenetic hypotheses - tests of two trees

1. Winning sites test (MP & ML)
2. Templeton test (MP & ML)
3. Kishino-Hasegawa test (MP & ML)
4. Shimodaira-Hasegawa test (ML only)
5. Bayesian Posterior Probabilities (BI only)

Kishino-Hasegawa test

- Uses differences in the support provided by individual sites for two trees to determine if the overall differences between the trees are significantly greater than expected from random sampling error.

- A parametric test - depends on assumptions:
  - that the characters are independent
  - and identically distributed (the same assumptions underlying the statistical interpretation of bootstrapping).

- It can be used with parsimony and maximum likelihood - implemented in PHYLIP and PAUP*.

Kishino-Hasegawa test

If the difference between trees (tree lengths or likelihoods) is attributable to sampling error, then characters will randomly support tree A or B and the total difference will be close to zero. The observed difference is significantly greater than zero if it is greater than 1.96 standard deviations. This allows us to reject the null hypothesis and declare the sub-optimal tree significantly worse than the optimal tree (p < 0.05).
**Kishino-Hasegawa test - an example**

<table>
<thead>
<tr>
<th>Tree</th>
<th>Length</th>
<th>Rank sums*</th>
<th>N</th>
<th>z</th>
<th>P**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Templeton</td>
<td>93</td>
<td>3</td>
<td>1.68184</td>
<td>1.7838</td>
<td>0.0831</td>
</tr>
</tbody>
</table>

Approximate probability of getting a more extreme test statistic under the null hypothesis of no difference between the two trees (Monte Carlo test). Estimated values in table (if any) indicate significant difference at P < 0.05. Consult a table for critical values of Wilcoxon rank sum when N is 25 or less.

**Questions**
- How reliable is this result?
- In particular, how well supported is the idea of multiple origins?
- How many origins can we confidently infer?

**Comparing competing phylogenetic hypotheses - tests of two trees**

1. Winning sites test (MP & ML)
2. Templeton test (MP & ML)
3. Kishino-Hasegawa test (MP & ML)
4. Shimodaira-Hasegawa test (ML only)
5. Bayesian Posterior Probabilities (BI only)
**Shimodaira-Hasegawa Test (ML only)**

- To be statistically valid, the Kishino-Hasegawa test should be of trees that are selected a priori.
- However, most people have used trees selected a posteriori on the basis of the phylogenetic analysis.
- When we test the 'best' tree against some other tree, the KH test will be biased towards rejection of the null hypothesis (so it is a conservative result when it doesn't reject).
- The SH test is a similar but more statistically correct technique in these circumstances and should be preferred, but is limited to comparing likelihood scores.

**Bayesian Hypothesis Testing**

Bayesian inference yields **Posterior Probabilities** for clades (branches).

Posterior Probabilities (BPP, or PP) are interpreted quite simply as:

**THE PROBABILITY THE HYPOTHESIS** (branch / tree) **IS CORRECT**

(Given the model, priors, & data)

These probabilities can be used to compare competing hypotheses instead of using the KH or SH tests.

---

**Example - DNA & Morphology** yield different trees

Morphology - preferred tree (best Bayes tree)

DNA [COII] - preferred tree (best Bayes tree)

* COII data reject COII tree at P<0.001

* Morphology data reject COII tree at P<0.005

Rare result - strongly supported incongruence

**Independent Dataset Congruence**

- Trees inferred from different data sets (different genes, morphology) should agree if they are accurate.
- Congruence between trees is best explained by their accuracy.
- Congruence can be investigated using consensus methods, bootstrapping, and KH / SH testing - use **dataset1 on the best tree of dataset2 (vice versa)** to see if the different datasets significantly reject the best trees of the other dataset.
- Incongruence requires further work to explain or resolve disagreements.

---

**Bayesian Hypothesis Testing - Simple & straightforward**

In MRBayes - check the output file name, `probpr` for the posterior probabilities of all likely bipartitions (if bipartition of interest isn't listed, its P is lower than the lowest listed, usually P<0.000).

This clade is a hypothesis that is rejected at P=0.000.

---

**SH Test - PAUP* output**

Tree

<table>
<thead>
<tr>
<th>Tree</th>
<th>-ln L</th>
<th>Diff -ln L</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16383</td>
<td>9935.26295</td>
<td>(best)</td>
</tr>
</tbody>
</table>

**Time used to compute likelihoods = 1.50 sec**

**Shimodaira-Hasegawa test:**

- SH test using RELL bootstrap (one-tailed test)
- Number of bootstrap replicates = 1000

<table>
<thead>
<tr>
<th>Tree</th>
<th>-ln L</th>
<th>Diff -ln L</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16383</td>
<td>10206.72132</td>
<td>271.45838</td>
</tr>
</tbody>
</table>

* P < 0.05

**This ML tree is significantly better than the alternative**
Systematics - Bio 615

Monophyly rejected
PP = 0.000067%

Character evolution:
Melanism - homology or homoplasy?

Confidence - Assessment of the Strength of the Phylogenetic Signal - part 3
1. Consistency Index
2. g1 statistic, PTP - test
3. Consensus trees
4. Decay index (Bremer Support)  [pp.190-193 - Text]
5. Bootstrapping / Jackknifing
6. Statistical hypothesis testing (frequentist) - Independent Dataset Congruence
7. Posterior probability - Bayesian Inference

Bayesian Phylogenetic Inference
Outline
1. Introduction, history
2. Advantages over ML
3. Bayes' Rule
4. The Priors
5. Marginal vs Joint estimation
6. Posteriors vs Bootstrap values
7. MCMC

Bayesian Statistics
Bayesian statistics is fundamentally different from the more commonly taught and used frequentist statistics
Based on the work of Reverend Thomas Bayes (1764)
Probability theory was Bayesian until ca. 1920s
Fisherian / Frequentist statistics developed 1929 & have since dominated
Logical problems with Bayesian Statistics were fixed in 1955 - re-emergence of Bayesian methods - not used for phylogenetics until 1996

The Classical Hypothesis Testing Setup
Step 1. Define the Research Hypothesis
A Research or Alternative Hypothesis is a statement derived from theory about what the researcher expects to find in the data.

Step 2. Define the Null Hypothesis
The Null Hypothesis is a statement of what you would not expect to find if your research or alternative hypothesis was consistent with reality.

Step 3. Conduct an analysis of the data to determine whether or not you can reject the null hypothesis with some pre-determined probability
If you can reject the null hypothesis with some probability, then the data are consistent with the model
If you cannot reject the null hypothesis with some probability, then the data are not consistent with the model

The Bayesian Approach
Bayesians, in contrast, try to do the following:
1) Make inferences based on all information at our disposal
2) See how new data effects our (old) inferences
3) Need to identify all hypotheses (or states of nature) that may be true
4) Need to know what each hypothesis (or state of nature) predicts that we will observe
5) Need to know how to compute the consequences. i.e. we need to know how to update our old inferences in light of our observations

In sum, their statistics matches how scientists actually think
Bayesian vs Frequentist

Coin - Flipping example - Flip a coin 50 times, get 28 heads

"Is this a fair coin?" - Frequentist approach: test -
- $H_0$: the probability of heads = 0.5
- $H_1$: the probability of heads ≠ 0.5

Bayesians would say "This approach is flawed. There is virtually no chance that the coin is perfectly fair ($P = 0.5000000000\ldots$) one only need collect enough data to demonstrate this. So this isn't a test of whether the coin is fair, it is a test of whether we have collected enough data to prove it isn't!"

(there is a difference between biological significance and statistical significance)

Bayesian vs Frequentist

Coin - Flipping example - Flip a coin 50 times, get 28 heads

"Is this a fair coin?" - Bayesian approach

Bayesian statistics allows us to answer a more useful question than is "P = 0.05?" namely...

“What is the probability that $P$ is between 0.4 and 0.6?”

Bayesians specify in advance the range they feel would represent a "fair coin" [ the hypothesis ]

Can calculate the probability of a hypothesis being true (give the data and some other stuff..)

Bayesian Phylogenetic Inference

Outline
1. Introduction, history
2. Advantages over ML
3. Bayes’ Rule
4. The Priors
5. Marginal vs Joint estimation
6. Posterior vs Bootstrap values
7. MCMC

Bayesian Phylogenetic Inference

Bayesian methods build on maximum likelihood
- Can use the same models of evolution
- The likelihood is incorporated into the equation

- Major advantages over ML:
  1. The likelihood is not maximized so considerable time is saved
  2. Branch support values analogous to bootstrap values are obtained in $\sim 10\times$ less time
  3. More realistic mixed-model analysis is possible
  4. Does not return a point-estimate of phylogeny but a quantification of the strength of the signal in the data

Bayesian Phylogenetic Inference

Maximum log-Likelihood score:

The probability of the data given the hypothesis (sub. model, tree, branch lengths)
- this is neither intuitive nor desirable

Bayesian Posterior Probability score:

The probability of the hypothesis (tree or branch) being correct given the data, priors, sub. model, and branch lengths
- this is intuitive and what we want!
Bayesian Phylogenetic Inference

Bayesian inference
- optimal topology is the one that maximizes the posterior probability (PP)
- PP is proportional to the prior probability and the likelihood
- We use Bayes' Rule to calculate the PP

Outline
1. Introduction, history
2. Advantages over ML
3. Bayes' Rule
4. The Priors
5. Marginal vs Joint estimation
6. Posteriors vs Bootstrap values
7. MCMC

Bayesian Phylogenetic Inference
Bayes' Rule - converts likelihood into Posterior Prob.
Pr = probability; D = data; tree = parameter (hypothesis)

\[ Pr(\text{tree} | D) = \frac{Pr(\text{tree}) \cdot Pr(D | \text{tree})}{Pr(D)} \]

posterior probability
marginal probability of data given model - a normalizing constant

Four Components of Bayes' Rule
1. Posterior Probability = probability of hypothesis given (...) after we have observed the data
2. Prior Probability = probability of hypothesis before the data are seen (industry of thought on this component!) - based on belief
3. Likelihood = probability of the data given the hypothesis
4. \( Pr(D) \) = marginal probability of the data
   - a constant scaling factor to get PP to range from 0.0 to 1.0

Four components to Bayes' Rule
Since \( Pr(D) \) is constant the equation can be written

\[ Pr(p | D) = Pr(p) \cdot Pr(D | p) \]
or in words
the posterior probability is proportional to the prior probability times the likelihood
Bayesian Phylogenetic Inference

Four components to Bayes’ Rule

And if we use a completely flat, uninformative prior, such that all values are equally likely \( a \ priori \):

\[
\Pr(p \mid D) = \Pr(D \mid p)
\]

or in words
the posterior probability is proportional to the likelihood!

But what are those priors anyway?

Subjectivity of the Prior is a common criticism of Bayesian Inference

Flat prior

Informative prior

1. Influence of the prior weakens as the signal in the data strengthens - few or very weak data can be dangerous
2. Priors are explicit and must be justified
3. Try different priors to see how they influence the posteriors

Bayesian Phylogenetic Inference

Outline

1. Introduction, history
2. Advantages over ML
3. Bayes’ Rule
4. The Priors
5. Marginal vs Joint estimation
6. Posteriors vs Bootstrap values
7. MCMC

Bayesian Phylogenetic Inference

Bayesian Inference differs subtly from Maximum Likelihood in how parameter values are obtained

- BI uses marginal estimation (volume of dist)
- ML uses joint estimation (maximum of dist)


Bayesian Phylogenetic Inference

BI & ML will prefer different trees when the data do not strongly prefer one tree (like comparing a bootstrap tree to a most parsimonious tree)
Terms - from lecture & readings

winning sites test
Templeton test
Kishino-Hasegawa test
Shimodaira-Hasegawa test
Bayesian Statistics
Reverend Thomas Bayes
Frequentist Statistics
posterior probability
prior probability
Bayes’ Rule
flat prior
marginal estimation
joint estimation

Study questions

Explain how a branch length (under parsimony) of 250 steps can have zero branch support?

The Templeton test, KH, and SH test all test what null hypothesis?

How must one select the trees to be compared for the KH to be statistically valid? What alternative test can one use if this requirement of the KH cannot be met?

What are four advantages of Bayesian Phylogenetic Inference over Maximum Likelihood?

Although BI & ML are expected to prefer the same tree if the same model of evolution is used for both - when might BI & ML prefer different trees & why?