

(http://www.nematodes.org/teaching/tutorials/phylogenetics/Bayesian_Workshop/Bayesian%20mini-conference.htm#_Toc145477467)

Model selection criteria

Review

Posada D & Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Syst. Biol.*, **53**, 793-808. [PDF](#)

Sanderson MJ & Kim J (2000) Parametric phylogenetics? *Syst. Biol.*, **49**, 817–829. [PDF](#)

Akaike information criterion

The Kullback-Liebler (K-L) distance (Kullback & Liebler, 1951) is a measure of the reduction in likelihood obtained by using an incorrect model in place of the ‘true’ model. The expected K-L distance can be estimated in phylogenetics by using the Akaike information criterion, AIC (Akaike 1974):

$$AIC = -2\log L + 2K$$

Where $\log L$ is the maximised log-likelihood and K is the parameter richness of the model. As parameter richness is increased, $\log L$ is expected to decrease while the ‘penalty term’ increases so the model with the lowest AIC will be a balance between parameter richness and the informativeness of additional parameters.

The AIC, as applied in Modeltest (Posada & Crandall 1998), can be

expressed more usefully in terms of the weight of support for each of a hierarchical series of models. The Akaike weight, w , for the i th model in a set of models R is:

$$w_i = \frac{e^{-(AIC_i - AIC_{min})}}{\sum_{j=1}^R e^{-(AIC_j - AIC_{min})}}$$

Cumulative Akaike weights for each of the R models (in order of highest w) can be used to determine the 95% set of models that minimise K-L. This provides a way to quantify model selection uncertainty.

Bayesian information criterion

The Bayesian information criterion, BIC, (Schwartz 1978) approximates marginal LogL for a candidate models:

$$\text{BIC} = -2\log L + K \log n$$

Where n is the sample size (in phylogenetics, $n \approx$ the number of characters in the alignment). Since the BIC approximates marginal $\log L$, the difference between two BIC estimates provides an approximation to Bayes factors (see **Bayes factors** below) with considerably less computational effort.

Likelihood ratio tests

Hierarchical likelihood ratio tests, hLRTs, (Fрати *et al.* 1997; Huelsebeck & Crandall 1997) require that the hypotheses under investigation be nested and can be tested against the χ^2 distribution (Goldman 1993). Non-hierarchical LRTs are also possible using the rule of thumb that two $\log L$ s constitutes a significant difference (Edwards 1972; Pagel 1999) to avoid computationally intensive Monte Carlo methods (Goldman 1993). The LRT test statistic for model 1 over null model 0 is:

$$\text{LRT} = 2(\log L_1 - \log L_0)$$

This is a hierarchical LRT if the null model is a special case of model 1. In phylogenetic model testing, the one-parameter JC69 model can be obtained from the two-parameter K80 model by assuming that transitions and transversions occur at the same rate so JC69 is nested within K80. Further nesting levels are shown in Figure 5 below. LRTs are perceived to be more accurate than the approximate AIC and BIC approaches as they allow comparison of actual maximum likelihoods from analyses under different models. LRTs are also convenient as they can be applied after analyses have been performed. There are, however, problems with the LRT approach (Sanderson & Kim 2000): (i) the non-hierarchical LRT method for non-nested models can only provide an approximate result; (ii) situations exist for which an optimal model cannot be selected by hLRTs; (iii) the starting point can affect the model selection procedure; and (iv) hLRTs involve performing multiple tests with the same data which may lead to false positives.

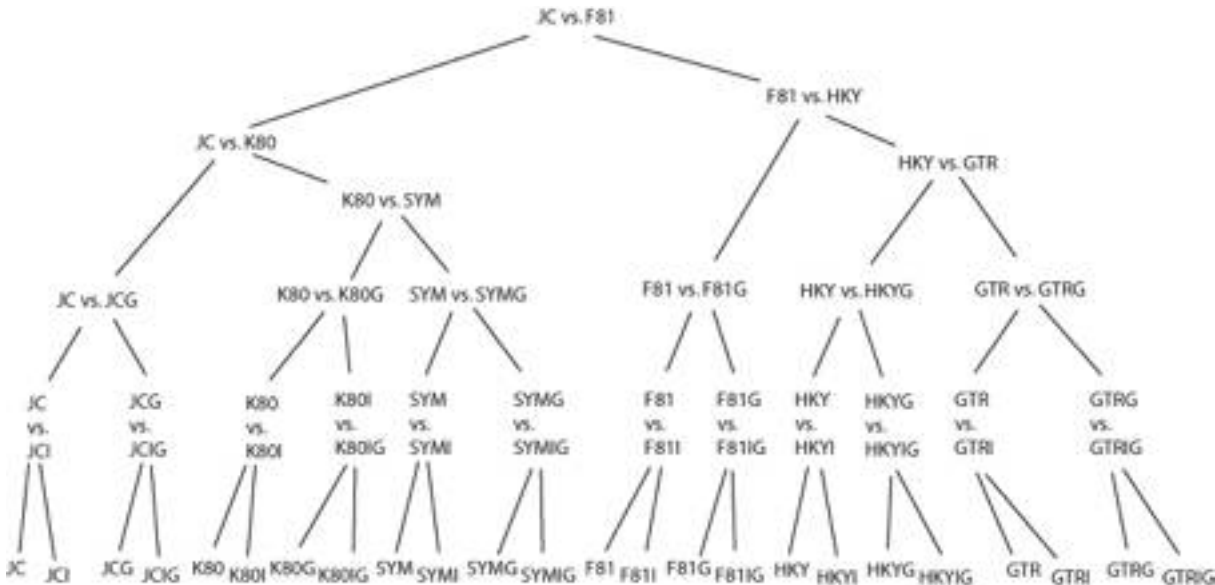


Figure 5. Modeltest hLRT hierarchy (after Posada and Crandall 2001)

Bayes factors

Bayes factors, B , were introduced by Kass & Raftery (1995) as a Bayesian equivalent to LRTs. Comparison between competing models M_i and M_j takes the form:

$$B_{ij} = \frac{P(D|M_i)}{P(D|M_j)}$$

Bayes factors compare marginal log-likelihoods and support for M_i over M_j can be quantified using the criteria of Kass and Raftery (1995): $B_{ij} < 1$, negative (support for M_j); $1 < B_{ij} < 3$, barely worth mentioning; $3 < B_{ij} < 12$, positive; $12 < B_{ij} < 150$, strong; and $B_{ij} > 150$, very strong. Bayes factors have the advantage that, unlike LRTs, the same test can theoretically be applied to both nested and non-nested data. The major drawback is that the ‘critical value’ for strong evidence of support for M_i over M_j , which is typically simplified to $B_{ij} > 10$ (e.g. Irestedt *et al.* 2004) should, according to Kass & Raftery (1995) be set at $B_{ij} > 20$ in phylogenetics to reduce the occurrence of false positives. Results of studies using Bayes factors should, therefore, be interpreted with caution.

Model averaging

Within a Bayesian framework, it is possible to extend the model testing approach from selection of a single best model to model averaging over the set of candidate models. In practise this is very straightforward as parameter estimates for each candidate models can be sampled from the stationary phase of a MCMC run in proportion to the probability of that model (see **Example** below). Probabilities for each model (or more sensibly the 95% set) can be obtained through AIC or BIC weights (as described above for AIC weights) or through Bayes factors by determining the posterior probability for each candidate model. For R candidate models, the posterior probability of the i th model is:

$$P(M_i | D) = \frac{P(D|M_i)P(M_i)}{\sum_{r=1}^R P(D|M_r)P(M_r)}$$

While the use of posterior probabilities is the ‘purest’ Bayesian approach to model averaging, uncertainty in prior specification can affect the posterior probability assigned to each model (see *Choosing and using Bayesian priors* below; this is also an issue in use of BIC) so AIC weights are often the best guide to model probabilities for use in Bayesian model averaging.

Example

References

Edwards AWF (1972) *Likelihood: an account of the statistical concept of likelihood and its application to scientific inference*. Johns Hopkins Univ. Press, London.

Frati F, Simon C, Sullivan J & Swofford DL (1997) Gene evolution and phylogeny of the mitochondrial cytochrome oxidase gene in Collembola. *J. Mol. Evol.*, **44**, 145–158. [PDF](#)

Goldman N (1993) Statistical tests of models of DNA substitution. *J. Mol. Evol.*, **36**, 182–198. [PDF](#)

Huelsenbeck JP & Crandall KA (1997) Phylogeny estimation and hypothesis testing using maximum likelihood. *Annu. Rev. Ecol. Syst.*, **28**, 437–466. [PDF](#)

Irestedt M, Fjeldså J, Nylander JAA & Ericson P (2004) Phylogenetic relationships of typical Antbirds (Thamnophilidae) and test of incongruence based on Bayes factors. *BMC Evol. Biol.*, **4**, 23. [PDF](#)

Kass RE & Raftery AE (1995) Bayes factors. *J. Am. Stat. Assoc.*, **90**, 773–795. [PDF](#)

Pagel M (1999) The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst.*

Biol., **48**, 612-622. [PDF](#)

Posada D & Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14**, 817-818. [PDF](#)

Posada D & Crandall KA (2001) Selecting the best-fit model of nucleotide substitution. *Syst. Biol.*, **50**, 580-601. [PDF](#)

Schwartz G (1978) Estimating the dimension of a model. *Ann. Stat.*, **6**, 461-464. [PDF](#)