jModelTest 0.1.1 (April 2008)

© David Posada 2008 onwards

dposada@uvigo.es
http://darwin.uvigo.es/

See the jModeltest FORUM and FAQs at http://darwin.uvigo.es/

1. Disclaimer

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version. This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details. You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 59 Temple Place - Suite 330, Boston, MA 02111-1307, USA. The jModelTest distribution includes Phyml and Consense (Phylip package) executables. These programs are protected by their own license and conditions, and using jModelTest implies agreeing with those conditions as well.

2. Purpose

jModelTest is a tool to carry out statistical selection of best-fit models of nucleotide substitution. It implements five different model selection strategies: hierarchical and dynamical likelihood ratio tests (hLRT and dLRT), Akaike and Bayesian information criteria (AIC and BIC), and a decision theory method (DT). It also provides estimates of model selection uncertainty, parameter importances and model-averaged parameter estimates, including model-averaged phylogenies. The theoretical background is described elsewhere (Posada and Buckley 2004b; Sullivan and Joyce 2005).

3. Citation

When using jModelTest you should cite all these:

Posada D. In press. jModelTest: Phylogenetic Model Averaging. Molecular Biology and Evolution.

Guindon S and Gascuel O (2003). A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood". *Systematic Biology* 52: 696-704.

And if you use jModelTest to build a model-averaged tree, you should also cite this:

Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle (USA). http://evolution.genetics.washington.edu/phylip.html

4. History

Version 0.0 (March 2005):	started the program.
Version 0.1 (February 2008):	the first version of jModelTest is released.
Version 0.1.1 (April 2008):	minor changes.

5. Usage

jModelTest is a Java program developed in Xcode under MacOS X 10.5. It provides a GUI where the user can select the input file (a DNA alignment) and specify the different options for the model selection analysis. To accomplish most of its tasks, jModelTest builds up a pipeline with several freely available programs:

- ReadSeq (Gilbert 2007): DNA alignment reading.
- Phyml (Guindon and Gascuel 2003): likelihood calculations. In Mac OS X and Linux the program uses Phyml beta version 3.0, which is faster but can be unstable.
- Consense (Felsenstein 2005): consensus trees.
- Ted (D. Posada): euclidean distances between trees.

5.1 Operative Systems

Although jModelTest it is optimized for MacOS X, executables are provided to run the program under Windows XP and Linux. In order to avoid potential problems during execution, the program folder and folders therein should be located under a path without spaces.

5.2 Starting the program

jModelTest can be started in the OS systems described above if a recent Java environment is properly installed (see http://www.java.com/). After double-clicking on the jModelTest.jar file, the program console, with several menus and a text panel, should open (Figure 1) (if this does not work, open a console prompt, move to the jModelTest folder and type "java –jar jModeltest.1.0.jar"). The text in this console can be edited ("Edit > …") and saved to a file, ("Edit > Save console") or printed ("Edit > Print console") at any time.

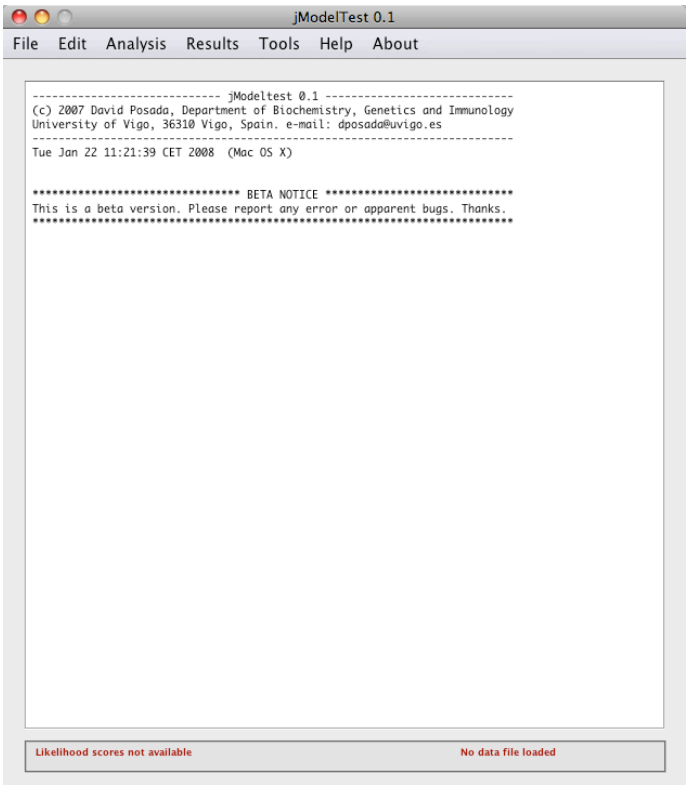


Figure 1. jModelTest console.

5.3 Input datafiles

The input file for jModelTest is a DNA sequence alignment in any of the formats allowed by ReadSeq (Gilbert 2007) (see http://iubio.bio.indiana.edu/soft/molbio/readseq/java/), including fasta, phylip, nexus, and other standards (**note that ReadSeq does not read in long sequence names in several formats**). An input file can be specified by clicking on the menu "File > Load DNA alignment".

5.4 Likelihood settings

Likelihood calculations are carried out with Phyml (Guindon and Gascuel 2003). There are 88 models currently implemented in jModelTest, including 11 substitution schemes, equal or unequal base frecuencies (+F), a proportion of invariable sites (+I) and rate variation among sites with a number of rate categories (+G) (Table 1). The panel for the likelihood calculations is available from the menu Analysis > "Compute likelihood scores" (Figure 2). Here it is possible to a large extent to specify which models will be compared (a minimum of 3, and a maximum of 88). For example, the user can select a diferent number of substitution schemes, whose exact names will depend on the +F, +I and +G options.

3 schemes: JC, HKY and GTR.
5 schemes: JC, HKY, TN, TPM1, and GTR.
7 schemes: JC, HKY, TN, TPM1, TIM1, TVM and GTR.
11 schemes: JC, HKY, TN, TPM1, TPM2, TPM3, TIM1, TIM2, TIM3, TVM and GTR.

For each model, there is the option of fixing the topology or to optimize it. In all cases branch lengths will be estimated and therefore counted as model paramenters. A fixed tree can be estimated using the BIONJ algorithm (Gascuel 1997) with the JC model, or it can be specified by the user from a file (in Newick format). Alternativel, potentially different BIONJ or a ML tree can be estimated for each model, which will require more computation time, specially for the ML optimization. Note that the LRTs methods will only be available when the likelihoods scores are calculated upon a fixed topology.

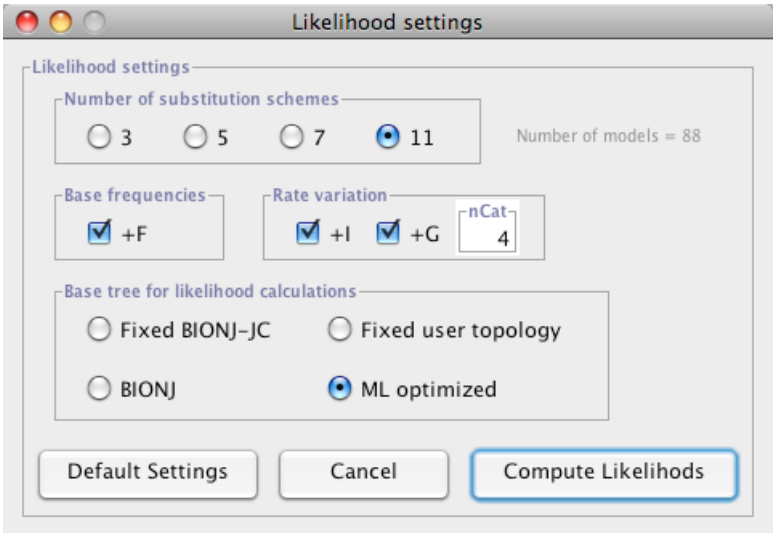


Figure 2. Settings for the likelihood calculations.

The likelihood computations can take a very variable amount of time depending on the data, number of candidate models and tree optimization. The console will print out the parameter estimates and likelihood scores (Figure 3). In addition, a progress bar will show how these calculations proceed (Figure 4).

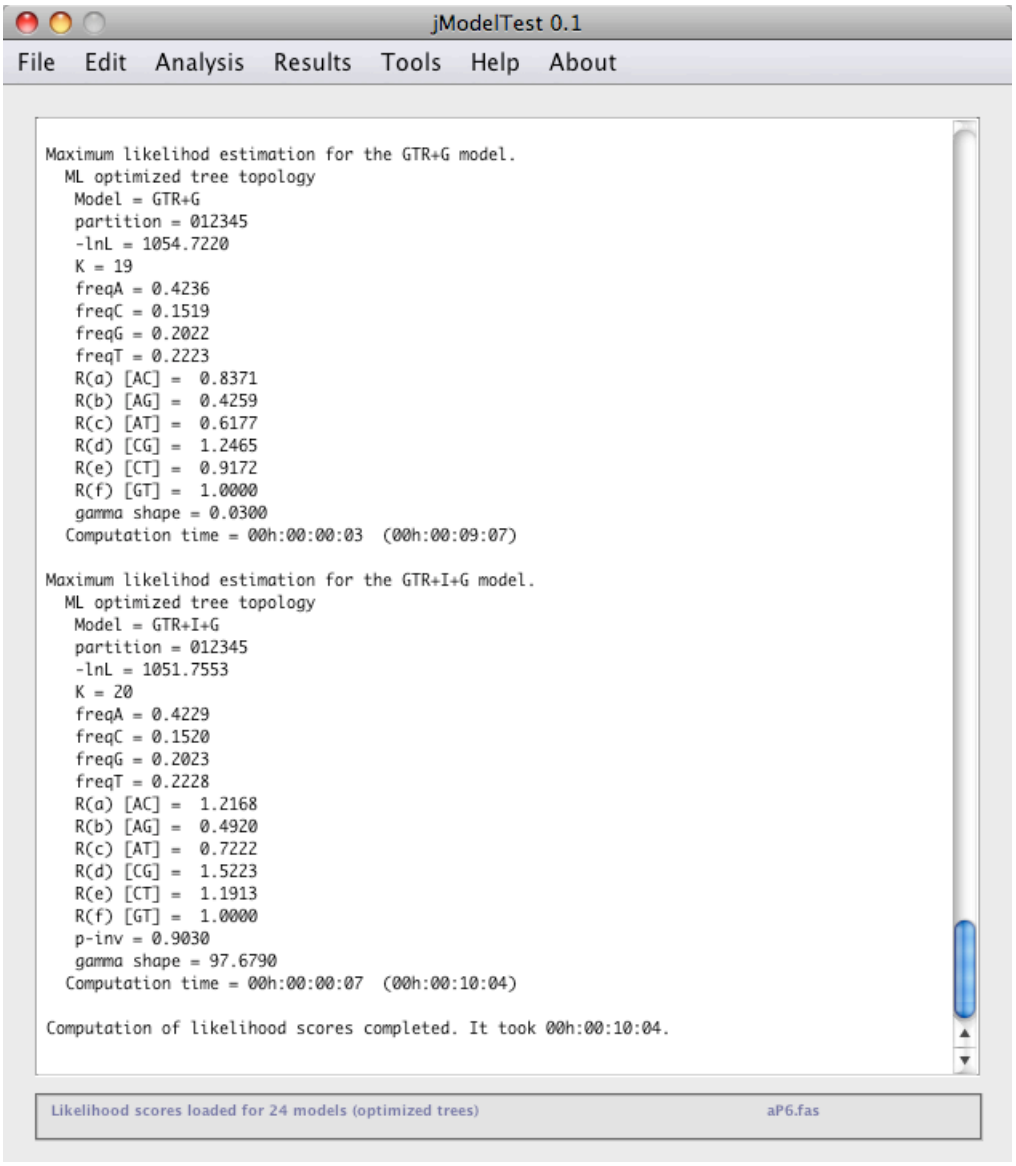


Figure 3. Likelihood calculations.

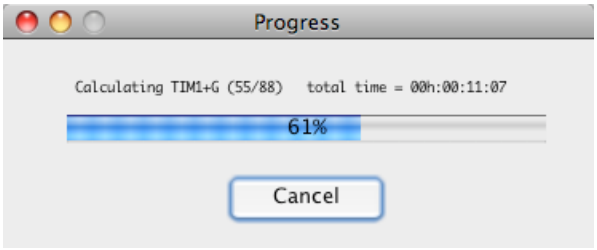


Figure 4. Progress bar for the likelihood calculations.

Table 1. Substitution models available in jModelTest. Any of these models can include invariable sites (+I), rate variation among sites (+G), or both (+I+G).

Model	Reference	Free parameters	Base frequencies	Substitution rates	Substitution code
JC	(Jukes and Cantor 1969)	0	equal	AC=AG=AT=CG=CT=GT	000000
F81	(Felsenstein 1981)	3	unequal	AC=AG=AT=CG=CT=GT	000000
K80	(Kimura 1980)	1	equal	AC=AT=CG=GT; AG=CT	010010
HKY	(Hasegawa, Kishino, and Yano 1985)	4	unequal	AC=AT=CG=GT; AG=CT	010010
TNef	(Tamura and Nei 1993)	2	equal	AC=AT=CG=GT; AG; CT	010020
TN	(Tamura and Nei 1993)	5	unequal	AC=AT=CG=GT; AG; CT	010020
TPM1	= K81 (Kimura 1981)	2	equal	AC=GT; AT=CG; AG=CT	012210
TPM1uf	(Kimura 1981)	5	unequal	AC=GT; AT=CG; AG=CT	012210
TPM2		2	equal	AC=AT; CG=GT; AG=CT	010212
TPM2uf		5	unequal	AC=AT; CG=GT; AG=CT	010212
TPM3		2	equal	AC=CG; AT=GT; AG=CT	012012
TPM3uf		5	unequal	AC=CG; AT=GT; AG=CT	012012
TIM1ef	(Posada 2003)	3	equal	AC=GT; AT=CG; AG, CT	012230
TIM1	(Posada 2003)	6	unequal	AC=GT; AT=CG; AG, CT	012230
TIM2ef		3	equal	AC=AT; CG=GT; AG; CT	010232
TIM2		6	unequal	AC=AT; CG=GT; AG; CT	010232
TIM3ef		3	equal	AC=CG; AT=GT; AG; CT	012032
TIM3		6	unequal	AC=CG; AT=GT; AG; CT	012032
TVMef	(Posada 2003)	4	equal	AC; AT, CG; GT; AG=CT	012314
TVM	(Posada 2003)	7	unequal	AC; AT; CG; GT; AG=CT	012314
SYM	(Zharkikh 1994)	5	equal	AC; AG; AT; CG; CT; GT	012345
GTR	= REV (Tavaré 1986)	8	unequal	AC; AG; AT; CG; CT; GT	012345

5.5 Model selection and averaging

Once all likelihood scores are in place, models can be selected according to different criteria from the menu ("Analysis > …").

5.5.1 Akaike information criterion (AIC)

Under the **AIC** framework, the user can select whether to use the **AICc** (corrected for small samples) instead of the standard AIC, which is the default (Figure 5). If the AICc is specified, the user needs to specify the sample size, which by default is the number of sites in the alignment. A **confidence interval** (CI) of models including a specified fraction of the models (by default 100%) will also be built according to the cumulative weight. When model does not fit completely within the CI (the previous model in the sorted table has a cumulative weight below the CI and this model has a cumulative weight above the CI and) it will be included in the CI with a probability equal to the portion of its cumulative weight that is inside the CI. For example, in the table displayed in the Figure 5, the model F81+G has a probability of (0.9500 – 0.9412) / 0.0220 = 0.4 of being include within the 95% CI.

Parameter importances are rescaled by the total weight of the models included in the confidence interval. In order to obtain and **model averaged estimates**, weights are rescaled by the parameter importance. If the user wish to do so, a block of PAUP* commands specifying the likelihood settings of the AIC model can be written to the console.

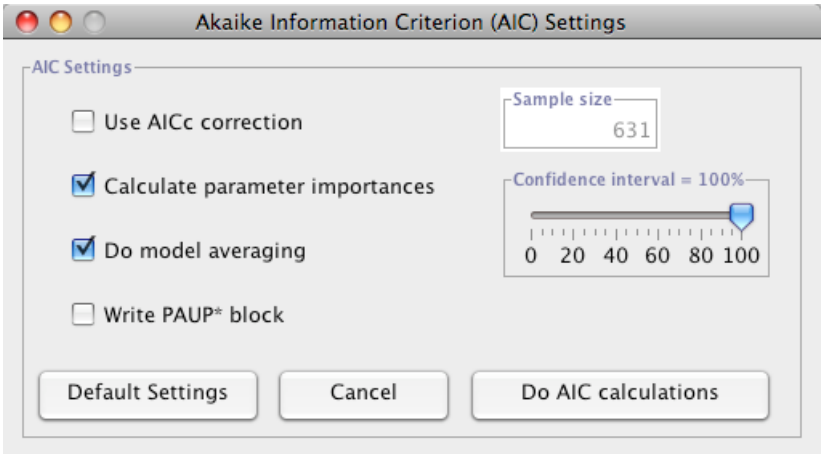


Figure 5. Options for the AIC selection.

Once the calculations have been carried out, the program reports the model selected, which is the one with the smallest AIC. Model selection uncertainty is displayed in a table in which models are sorted in increasing order according to their AIC score. This table also includes the AIC differences and the relative and cumulative AIC weights (Figure 6). These results are also available at the "Model > Show model table" menu, in which the selected model is displayed in red. After this, the confidence interval, parameter importances and model-averaged estimates are displayed (Figure 7).

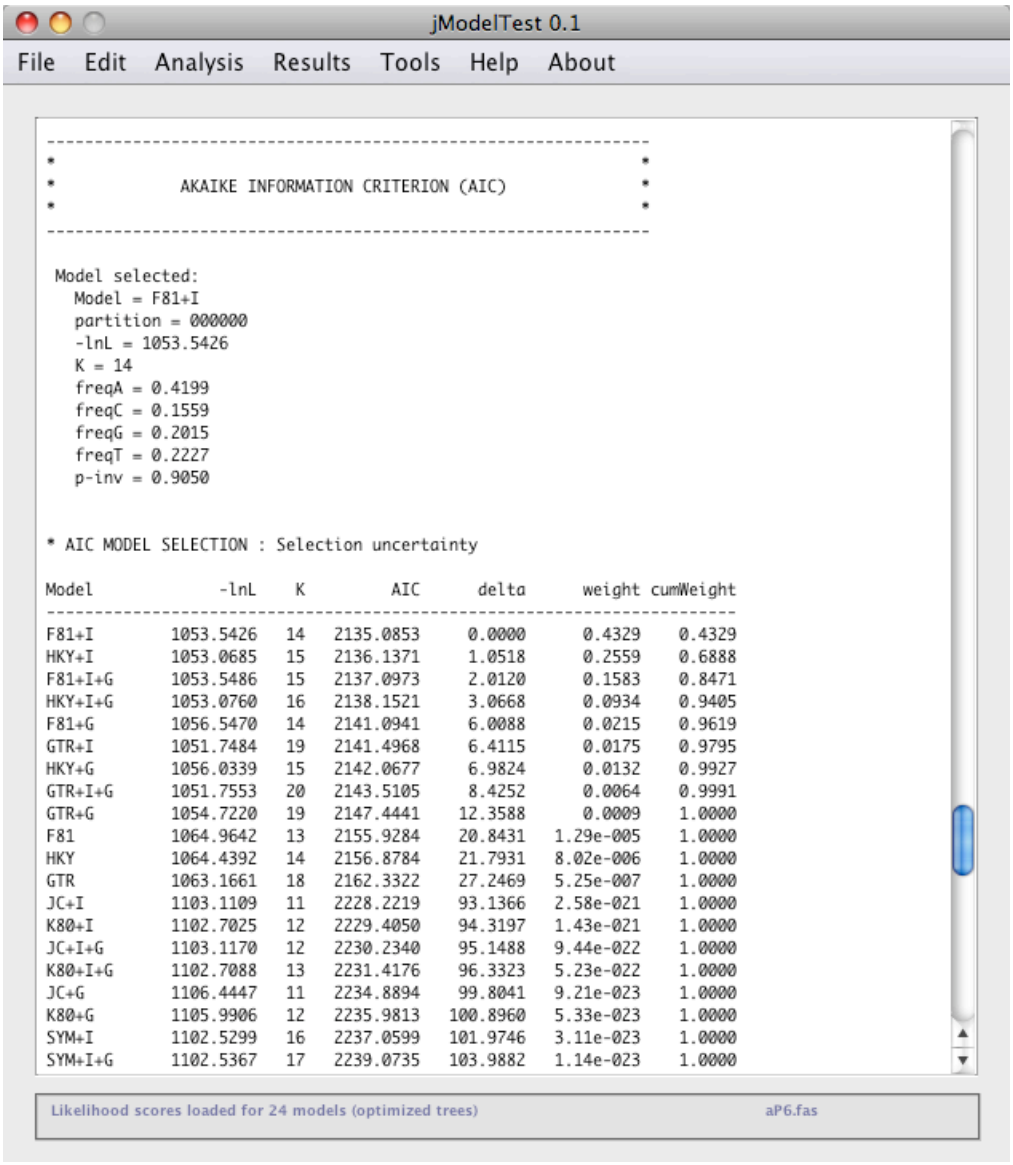


Figure 6. Results of the AIC selection. -lnL: negative log likelihood; K: number of estimated parameters; AIC: Akaike Information Criterion; delta: AIC difference; weight: AIC weight; cumWeight: cumulative AIC weight.

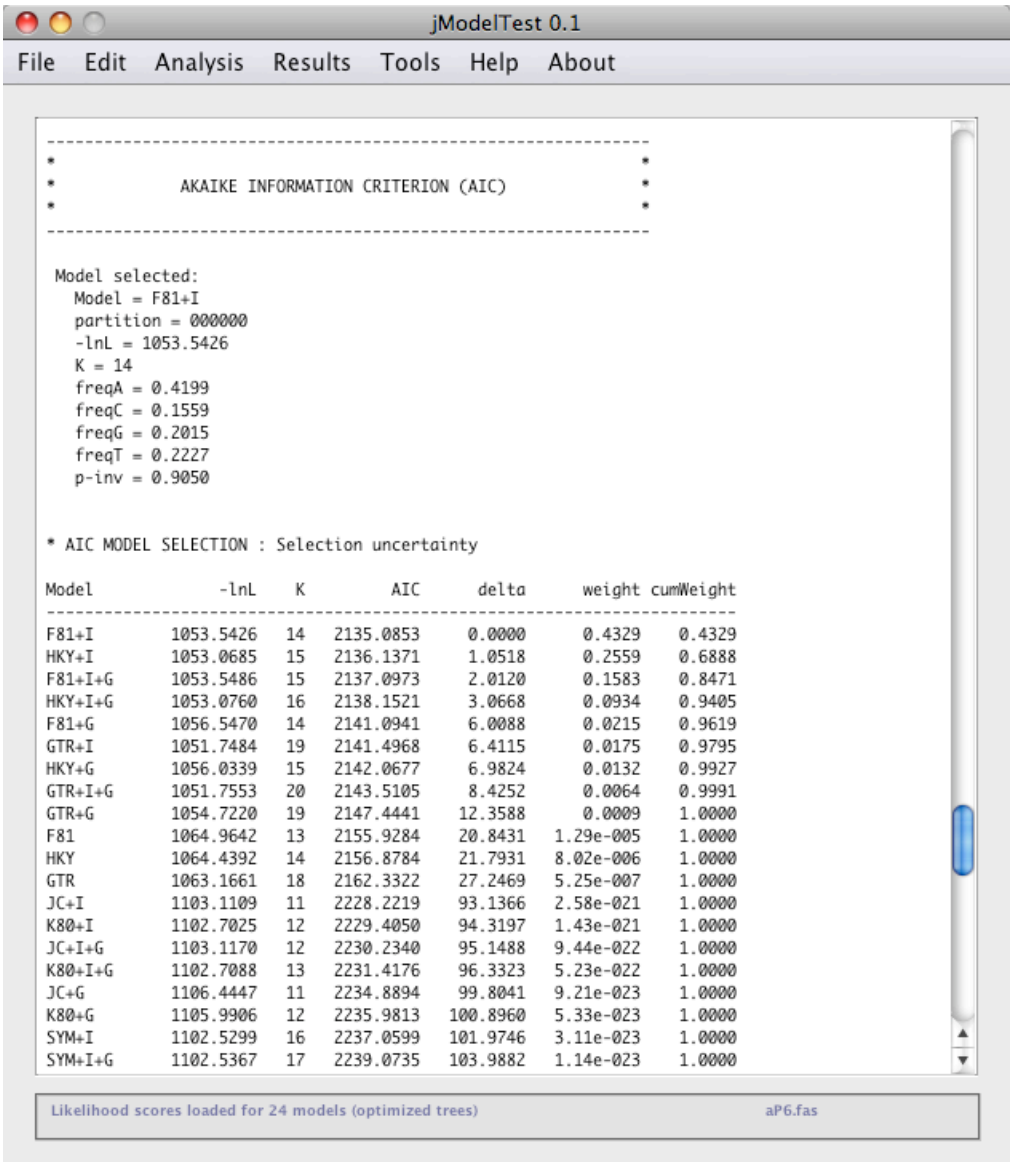


Figure 7. AIC confidence interval, parameter importances and model-averaged estimates.

5.5.2 Bayesian information criterion (BIC)

The options (Figure 8) and results for the BIC selection are analogous to those described above for the AIC, expect for the lack of a correction for small samples.

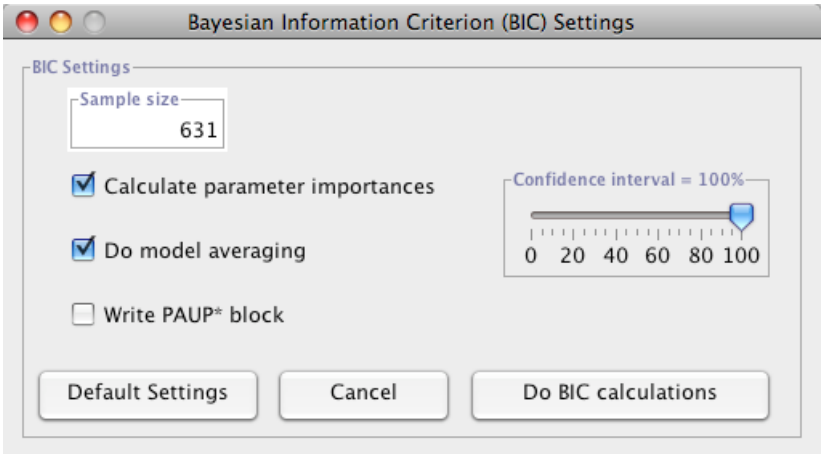


Figure 8. Options for the BIC selection.

5.5.3 Decision theory performance-based selection (DT)

The options for the DT selection (Figure 9) are analogous to those described above for the BIC. However, the calculation of weights here is different. This because DT statistic is of a different nature, and the standard theory does not apply anymore. Right now, the DT weights are simply the rescaled reciprocal DT scores ($(1/DT_i)$/sum). The weights reported here are very gross and should be used with caution. Remember also that parameter importances and model averaged estimates use these weights.

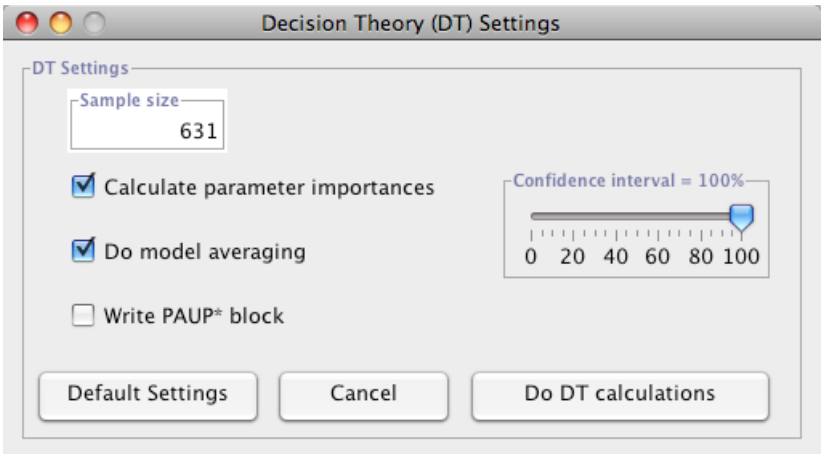


Figure 9. Options for the DT performance-based selection.

5.5.4 Sequential likelihood ratio tests (sLRTs)

Sequential likelihood ratio tests for model selection can be implemented under a particular hierarchy of likelihood ratio tests (**sLRTs**), in which the user can specify the order of the LRTs and whether parameters are added (*forward* selection) or removed (*backward* selection) (Figure 10). Alternatively, the order of the LRTs can be set automatically or dynamically (**dLRTs**) by comparing the current model with the one that is one hypothesis away and provides the largest increase (under forward selection) or smallest decrease (under backward selection) in likelihood (Figure 11). The sLRTs will be available only if the likelihoods scores were calculated upon a fixed topology, due to the nesting requirement of the chi-square approximation.

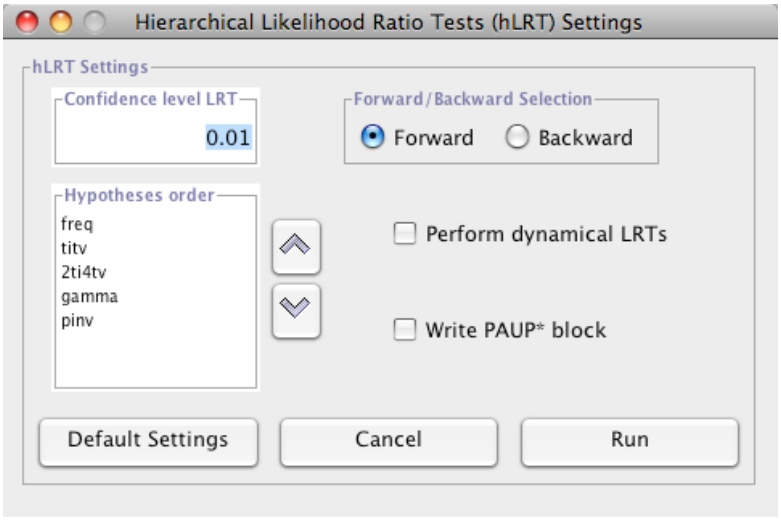


Figure 10. Options for the sequential LRTs.

The number and type of hypotheses tested (i.e., of LRTs performed) will depend on the particular models included in the candidate set. The possible tests are:

- Base frequencies
  - freq = unequal base frequencies (option +F, Figure 2).

- Substitution constraints (*nss* = number of substitution schemes in Figures 2 and 11)
  - titv = transition/transversion ratio (*nss* = 3 , 5, 7, 11).
  - 2ti4tv = 2 different transition rates and 4 different transversion rates (*nss* = 3).
  - 2ti = 2 different transition rates (*nss* = 5, 7, 11).
  - 2tv = 2 different transversion rates (AC=GT and AT=CG when *nss* = 3 and 5; all options for *nss* = 11).
  - 4tv = 4 different transversion rates (*nss* = 5, 7, 11).

- Rate variation among sites
  - gamma = gamma-distributed rate heterogeneity (option +G, Figure 2).
  - pinv = proportion of invariable sites (option +I, Figure 2).

In the case of the TIM and TPM family for *nss* = 11, the model with highest likelihood (TIM1 or TIM2 or TIM3; TPM1 or TPM2 or TPM3) will be used in the LRT. The level of significance for each individual LRT can be specified. By default this is value 0.01. The standard chi-square approximation is used in all tests, except for those involving gamma distributed rate variation among sites or a proportion of invariable sites, where a mixed chi-square is used instead. The default hierarchy for 24 models (nss=3, +F, +I, +G), in which the order of tests is titv-2titv-pinv-gamma is displayed in Figure 13.

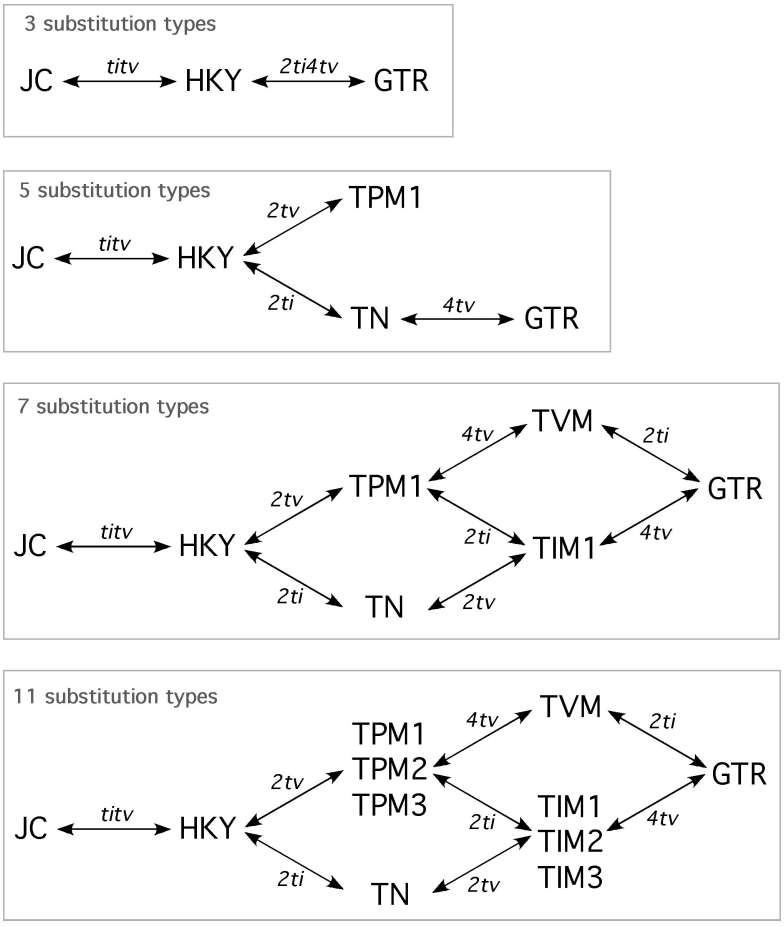


Figure 11. Possible LRTs for different substitution types according to the number of substitution schemes specified (Figure 2). The exact names of the models compared will change according to the +F, +I and +G options and the outcome of their LRTs.

5.6 Model averaged phylogeny

Like any other model parameter, the program can compute a model averaged estimate of the tree topology (Figure 12). This estimate is obtained by calculating a weighted (using will be AIC, BIC or DT weights) consensus (majority rule) from all the trees corresponding to the models in the candidate set (or within a given confidence interval) (Figure 13). This option is only available when the tree topology has been optimized for every model. A strict consensus can also be computed, although in this case the weights have no meaning.

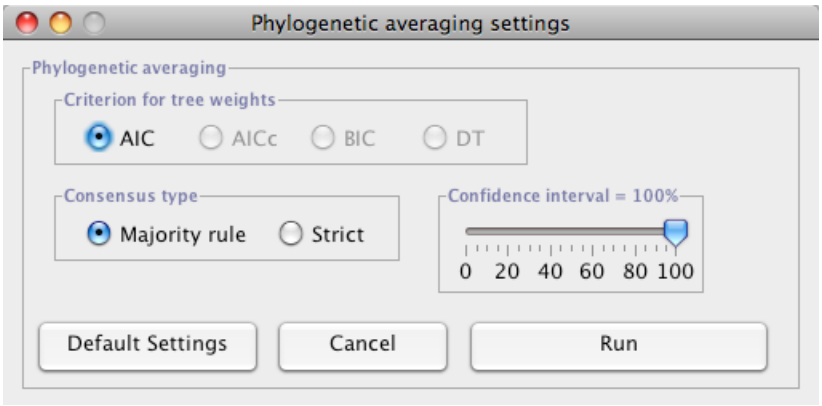


Figure 12. Options for the sequential LRTs.

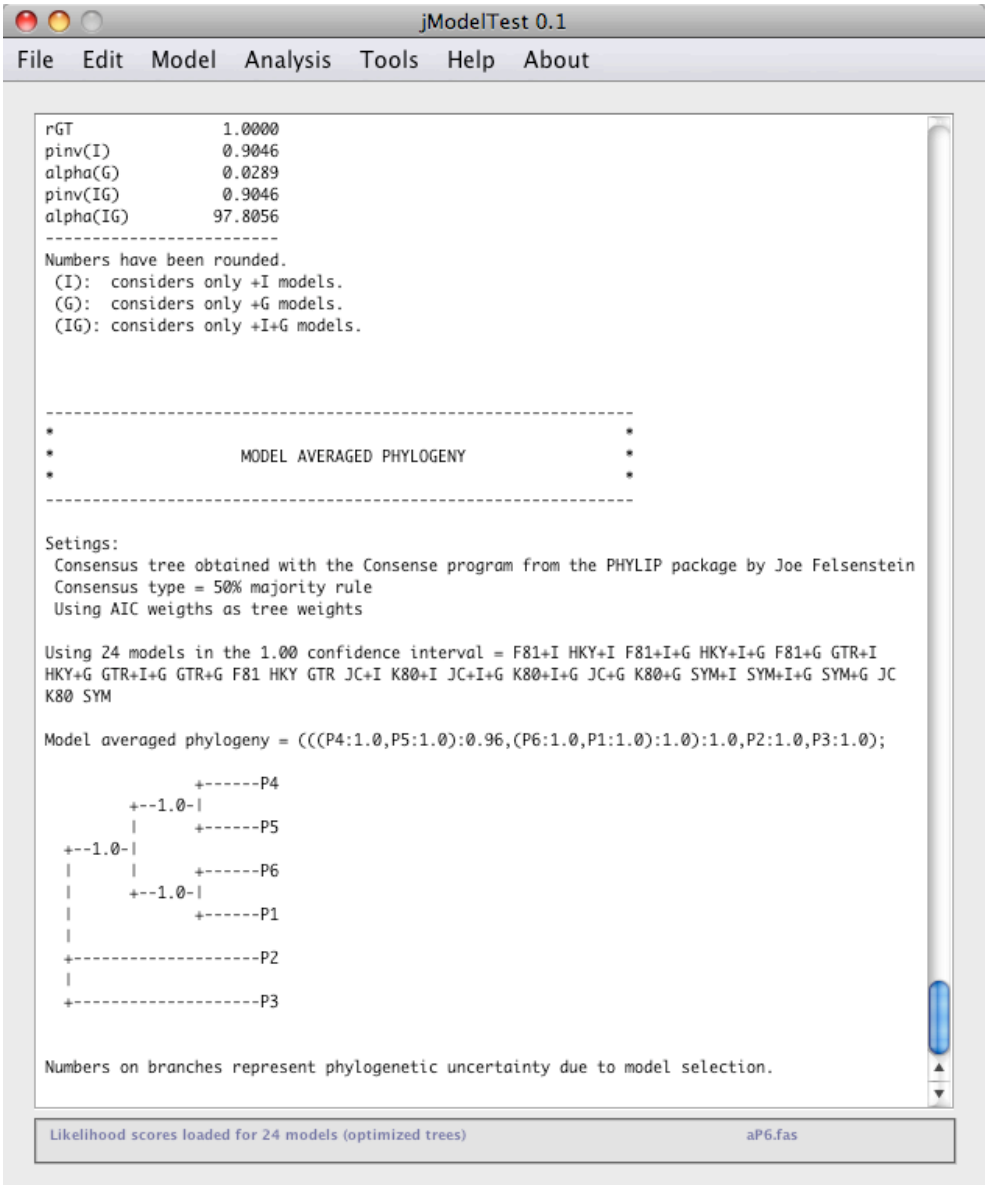


Figure 13. Console output showing a model-averaged phylogeny for 24 models with the AIC.

6. Miscellaneous options

6.1 LRT calculator

The program includes a very simple calculator to perform likelihood ratios tests using the standard or a mixed chi-square approximation (Figure 14). The models tested should be nested (the null hypothesis is a special case of the alternative hypothesis).

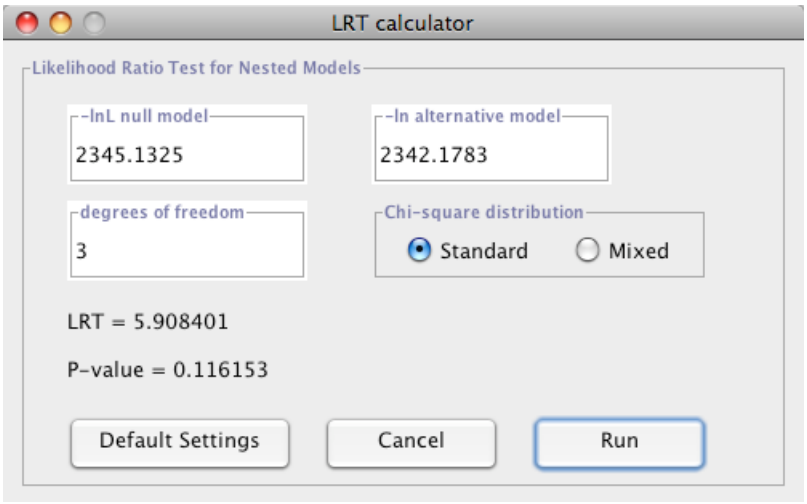
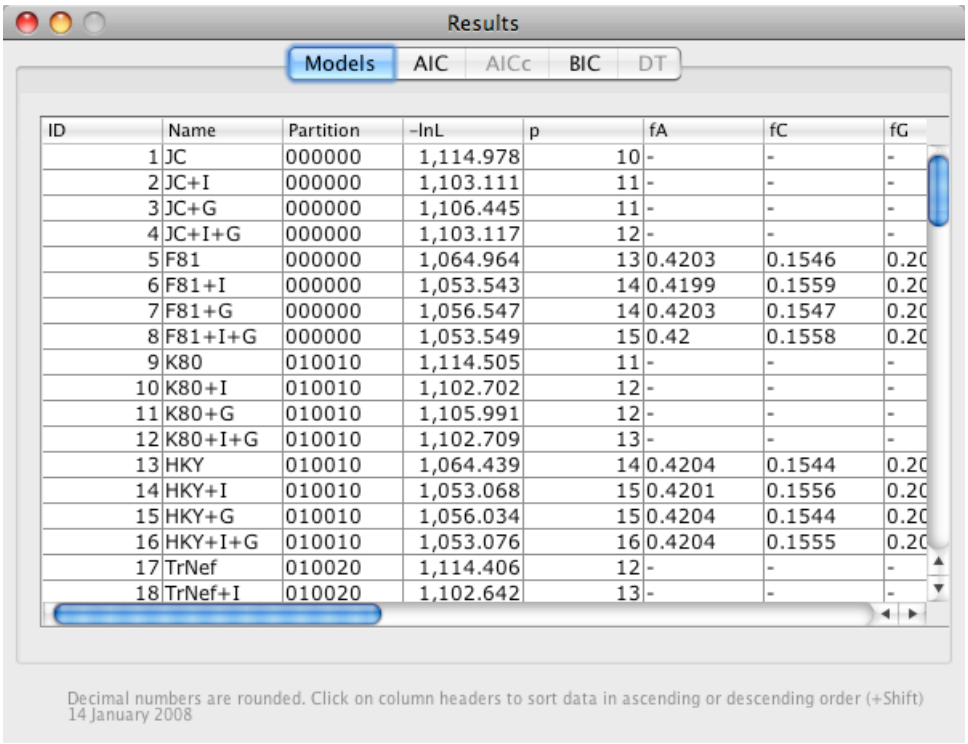


Figure 14. LTR calculator.

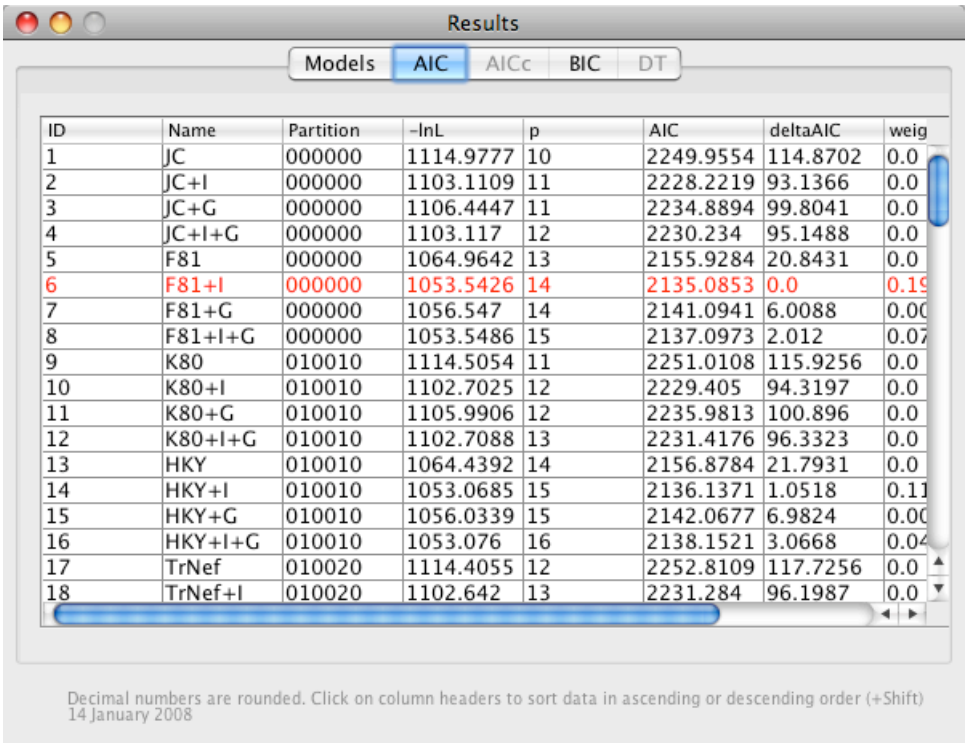
6.2 Results table

The likelihood scores (Figure 15) and the results from the different analyses (Figure 16) are stored in a table that can be displayed at any time from the menu “Results > Show results table”.



ID	Name	Partition	-lnL	p	fA	fC	fG
1	JC	000000	1,114.978	10	-	-	-
2	JC+I	000000	1,103.111	11	-	-	-
3	JC+G	000000	1,106.445	11	-	-	-
4	JC+I+G	000000	1,103.117	12	-	-	-
5	F81	000000	1,064.964	13	0.4203	0.1546	0.20
6	F81+I	000000	1,053.543	14	0.4199	0.1559	0.20
7	F81+G	000000	1,056.547	14	0.4203	0.1547	0.20
8	F81+I+G	000000	1,053.549	15	0.42	0.1558	0.20
9	K80	010010	1,114.505	11	-	-	-
10	K80+I	010010	1,102.702	12	-	-	-
11	K80+G	010010	1,105.991	12	-	-	-
12	K80+I+G	010010	1,102.709	13	-	-	-
13	HKY	010010	1,064.439	14	0.4204	0.1544	0.20
14	HKY+I	010010	1,053.068	15	0.4201	0.1556	0.20
15	HKY+G	010010	1,056.034	15	0.4204	0.1544	0.20
16	HKY+I+G	010010	1,053.076	16	0.4204	0.1555	0.20
17	TrNef	010020	1,114.406	12	-	-	-
18	TrNef+I	010020	1,102.642	13	-	-	-

Figure 15. Model table showing the likelihood scores and parameter estimates for each value.



ID	Name	Partition	-lnL	p	AIC	deltaAIC	weig
1	JC	000000	1114.9777	10	2249.9554	114.8702	0.0
2	JC+I	000000	1103.1109	11	2228.2219	93.1366	0.0
3	JC+G	000000	1106.4447	11	2234.8894	99.8041	0.0
4	JC+I+G	000000	1103.117	12	2230.234	95.1488	0.0
5	F81	000000	1064.9642	13	2155.9284	20.8431	0.0
6	F81+I	000000	1053.5426	14	2135.0853	0.0	0.19
7	F81+G	000000	1056.547	14	2141.0941	6.0088	0.00
8	F81+I+G	000000	1053.5486	15	2137.0973	2.012	0.07
9	K80	010010	1114.5054	11	2251.0108	115.9256	0.0
10	K80+I	010010	1102.7025	12	2229.405	94.3197	0.0
11	K80+G	010010	1105.9906	12	2235.9813	100.896	0.0
12	K80+I+G	010010	1102.7088	13	2231.4176	96.3323	0.0
13	HKY	010010	1064.4392	14	2156.8784	21.7931	0.0
14	HKY+I	010010	1053.0685	15	2136.1371	1.0518	0.11
15	HKY+G	010010	1056.0339	15	2142.0677	6.9824	0.00
16	HKY+I+G	010010	1053.076	16	2138.1521	3.0668	0.04
17	TrNef	010020	1114.4055	12	2252.8109	117.7256	0.0
18	TrNef+I	010020	1102.642	13	2231.284	96.1987	0.0

Figure 16. Model table showing the AIC scores and related measures. The AIC model is indicated in red.

7. The package

The jModelTest package includes several files in different subdirectories. These files should not be moved around. It is best to put the jModelTest folder in a path without spaces.

jModelTest
- + -------- doc
- + -------- examples
- + -------- exe
  - + ---- phyml
  - + ---- consense
  - + ---- ted
- + jModelTest.x.x.jar
- + -------- license
- + README.html

README.html: quick instructions and comments for users.
/doc/jModelTest.x.x.pdf: Documentation in PDF format
/examples/example.nex: an example data file in NEXUS format

*/exe/phyml/**: phyml executables for mac OS X, windows and linux.
*/exe/consense/**: consense executables for mac OS X, windows and linux.
*/exe/ted/**: ted executables for mac OS X, windows and linux.

/license/gpl.html: GNU general public license in HTML format

7.1 Example file

The example file (*example.nex*) included is an alignment of 10 DNA sequences 1000 bp long. This alignment was simulated on a tree obtained from the coalescent process and under the HKY+I model, with these parameter values:

Effective population size = 10000
Mutation rate per nucleotide per site = 5e-5
Base frequencies (A, C, G, T) = 0.4, 0.2, 0.1, 0.3
Transition/transversion rate = 4
Alpha parameter of the gamma distribution = 0.4

8. Theoretical background

All phylogenetic methods make assumptions, whether explicit or implicit, about the process of DNA substitution (Felsenstein 1988). Consequently, all the methods of phylogenetic inference depend on their underlying substitution models. To have confidence in inferences it is necessary to have confidence in the models (Goldman 1993b). Because of this, it makes sense to justify the use of a particular model. Statistical model selection is one way of doing this. For a review of model selection in phylogenetics see Sullivan and Joyce (2005) and Johnson and Omland (2003). The strategies includes in jModelTest include sequential likelihood ratio tests (LRTs), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and performance-based decision theory (DT).

8.1 Sequential Likelihood Ratio Tests (sLRT)

In traditional statistical theory, a widely accepted statistic for testing the goodness of fit of models is the likelihood ratio test statistic (*LRT*):

$$LRT = 2\ (\ell_1 - \ell_0)$$

where ℓ_1 is the maximum likelihood under the more parameter-rich, complex model (alternative hypothesis) and ℓ_0 is the maximum likelihood under the less parameter-rich simple model (null hypothesis). When the models compared are nested (the null hypothesis is a special case of the alternative hypothesis) and the null hypothesis is correct, the *LRT* statistic is asymptotically distributed as a χ^2 with q degrees of freedom, where q is the difference in number of free parameters between the two models (Kendall and Stuart 1979; Goldman 1993b). Note that, to preserve the nesting of the models, the likelihood scores need to be estimated upon the same tree. When some parameter is fixed at its boundary (*p-inv*, α), a mixed χ^2 is used instead (Ohta 1992; Goldman and Whelan 2000). The behavior of the χ^2 approximation for the LRT has been investigated with quite a bit of detail (Goldman 1993a; Goldman 1993b; Yang, Goldman, and Friday 1995; Whelan and Goldman 1999; Goldman and Whelan 2000).

8.1.1 hLRT

Likelihood ratio tests can be carried out sequentially by adding parameters (*forward selection*) to a simple model (JC), or by removing parameters (*backward selection*) from a complex model (GTR+I+G) in a specific order or hierarchy (hLRT; see Figure 17). The performance of hierarchical LRTs for phylogenetic model selection has been discussed by Posada and Buckley (2004a) .

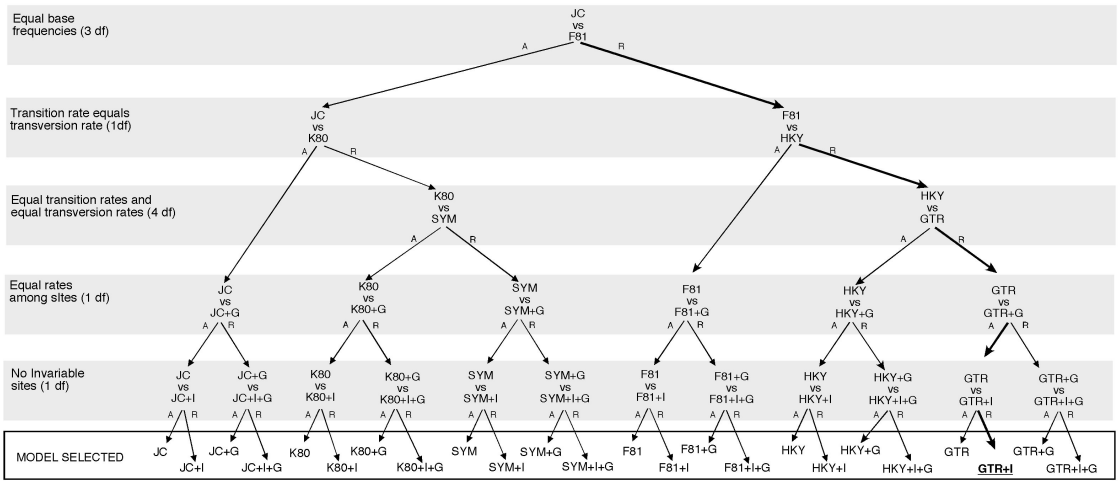


Figure 17. Example of a particular forward hierarchy of likelihood ratio tests for 24 models. At any level the null hypothesis (model on top) is either accepted (A) or rejected (R). In this example the model selected is GTR+I.

8.1.2 dLRT

Alternatively, the order in which parameters are added or removed can be selected automatically (Figure 18). One option to accomplish this is to add the parameter that maximizes a significant gain in likelihood during forward selection, or to add the parameter that minimizes a non-significant loss in likelihood during backward selection (Posada and Crandall 2001a). In this case, the order of the tests is not specified *a priori*, but it will depend on the particular data.

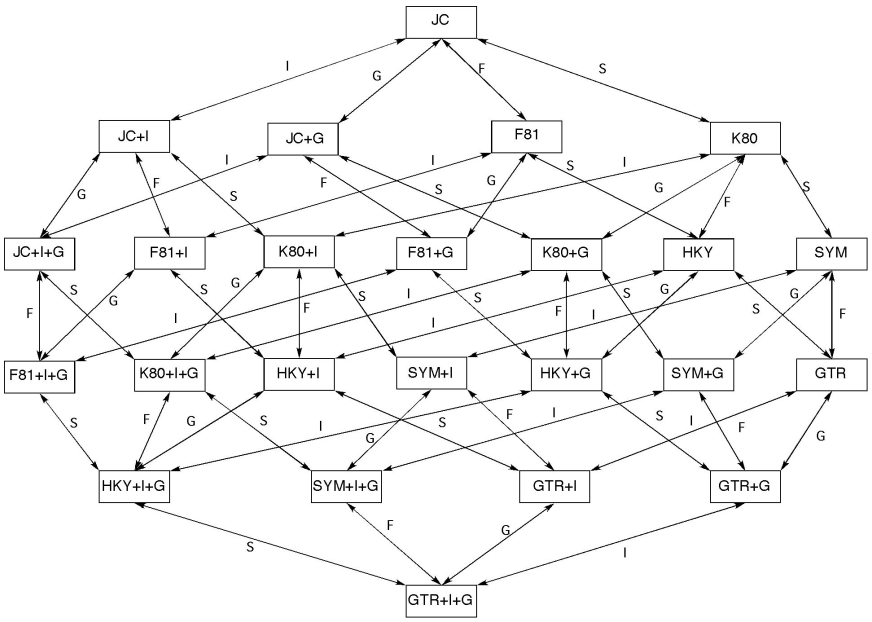


Figure 18. Dynamical likelihood ratio tests for 24 models. At any level a hypothesis is either accepted (A) or rejected (R). In this example the model selected is GTR+I. Hypotheses tested are: F = base frequencies; S = substitution type; I = proportion of invariable sites; G = gamma rates.

8.2 Akaike Information Criterion

The Akaike information criterion (AIC, (Akaike 1974) is an asymptotically unbiased estimator of the Kullback-Leibler information quantity (Kullback and Leibler 1951). We can think of the AIC as the amount of information lost when we use a specific model to approximate the real process of molecular evolution. Therefore, the model with the smallest AIC is preferred. The AIC is computed as:

$$AIC = -2\ell + 2K \ ,$$

where ℓ is the maximum log-likelihood value of the data under this model and K_i is the number of free parameters in the model, including branch lengths if they were estimated *de novo*. When sample size (n) is small compared to the number of parameters (say, $n/K < 40$) the use of a second-order AIC, AIC_c (Sugiura 1978; Hurvich and Tsai 1989), is recommended:

$$AIC_c = AIC + \frac{2K(K+1)}{n-K-1},$$

The AIC compares several candidate models simultaneously, it can be used to compare both nested and non-nested models, and model-selection uncertainty can be easily quantified using the AIC differences and Akaike weights (see Model uncertainty below). Burnham and Anderson (2003) provide an excellent introduction to the AIC and model selection in general.

8.3 Bayesian Information Criterion

An alternative to the use of the AIC is the Bayesian Information Criterion (BIC) (Schwarz 1978):

$$BIC = -2\ell + K \log n$$

Given equal priors for all competing models, choosing the model with the smallest BIC is equivalent to selecting the model with the maximum posterior probability. Alternatively, Bayes factors for models of molecular evolution can be calculated using reversible jump MCMC (Huelsenbeck, Larget, and Alfaro 2004). We can easily use the BIC instead of the AIC to calculate BIC differences or BIC weights.

8.4 Performance-based selection

Minin et al. (2003) developed a novel approach that selects models on the basis of their phylogenetic performance, measured as the expected error on branch lengths estimates weighted by their BIC. Under this **decision theoretic** framework (DT) the best model is the one with that minimizes the risk function:

$$C_i \approx \sum_{j=1}^{R} \left\| \hat{\mathbf{B}}_i - \hat{\mathbf{B}}_j \right\| \frac{e^{-BIC_i/2}}{\sum_{j=1}^{R} e^{-BIC_i/2}},$$

where

$$\left\| \hat{\mathbf{B}}_i - \hat{\mathbf{B}}_j \right\|^2 = \sum_{l=1}^{2t-3} (\hat{B}_{il} - \hat{B}_{jl})^2$$

and where t is the number of taxa. Indeed, simulations suggested that models selected with this criterion result in slightly more accurate branch length estimates than those obtained under models selected by the hLRTs (Minin et al. 2003; Abdo et al. 2005).

8.5 Model Uncertainty

The AIC, Bayesian and DT methods can rank the models, allowing us to assess how confident we are in the model selected. For these measures we could present their *differences* (Δ). For example, for the *i*th model, the AIC (BIC, DT) difference is:

$$\Delta_i = AIC_i - min(AIC),$$

where *min AIC* is the smallest AIC value among all candidate models. The AIC differences are easy to interpret and allow a quick comparison and ranking of candidate models. As a rough rule of thumb, models having Δ_i within 1-2 of the best model have substantial support and should receive consideration. Models having Δ_i within 3-7 of the best model have considerably less support, while models with $\Delta_i > 10$ have essentially no support. Very conveniently, we can use these differences to obtain the relative AIC (BIC) **weight** (w_i) of each model:

$$w_i = \frac{\exp(-1/2\Delta_i)}{\sum_{r=1}^{R}\exp(-1/2\Delta_r)}$$

which can be interpreted, from a Bayesian perspective, as the probability that a model is the best approximation to the truth given the data. The weights for every model add to 1, so we can establish an approximate 95% **confidence set of models** for the best models by summing the weights from largest to smallest from largest to smallest until the sum is 0.95 (Burnham and Anderson 1998, pp. 169-171; Burnham and Anderson 2003). This interval can also be set up stochastically (see above "Model selection and averaging"). **Note that this equation will not work for the DT (see the DT explanation on "Model selection and averaging").**

8.6 Model Averaging

Often there is some uncertainty in selecting the best candidate model. In such cases, or just one when does not want to rely on a single model, inferences can be drawn from all models (or an optimal subset) simultaneously. This is known as model averaging or multimodel inference. See Posada and Buckley (2004a) and references therein for an explanation of application of these techniques in the context of phylogenetics.

Within the AIC or Bayesian frameworks, it is straightforward to obtain a **model-averaged estimate** of any parameter (Madigan and Raftery 1994; Raftery 1996; Hoeting, Madigan, and Raftery 1999; Wasserman 2000; Burnham and Anderson 2003; Posada 2003). For example, a model-averaged estimate of the substitution rate between adenine and cytosine (φA-C) using the Akaike weights (w_i) for *R* candidate models would be:

$$\hat{\bar{\varphi}}_{A-C} = \frac{\sum_{i=1}^{R} w_i \; I_{\varphi_{A-C}}(M_i) \; \varphi_{A-Ci}}{w_+(\varphi_{A-C})},$$

where

$$w_+(\varphi_{A-C}) = \sum_{i=1}^{R} w_i I_{\varphi_{A-C}}(M_i),$$

and

$$I_{\varphi_{A-C}}(M_i) = \begin{cases} 1 \text{ if } \varphi_{A-C} \text{ is in model } M_i \\ 0 \text{ otherwise} \end{cases},$$

Note that need to be careful when interpreting the relative importance of parameters. When the number of candidate models is less than the number of possible combinations of parameters, the presence-absence of some pairs of parameters can be correlated, and so their relative importances.

8.6.1 Model averaged phylogeny

Indeed, the averaged parameter could be the topology itself, so we could construct a **model–averaged estimate of phylogeny**. For example, one could estimate a ML tree for all models (or a best subset) and with those one could build a weighted consensus tree using the corresponding Akaike weights. See Posada and Buckley (2004a) for a practical example.

8.7 Parameter importance

It is possible to estimate the **relative importance** of any parameter by summing the weights across all models that include the parameters we are interested in. For example, the relative importance of the substitution rate between adenine and cytosine across all candidate models is simply the denominator above, $w_+(\varphi_{A-C})$.

9. Acknowledgements

Thanks to Stephane Guindon for his generous help with Phyml, and to John Huelsenbeck for suggesting the stochastic calculation of confidence intervals.

10. Credits

- Phyml by Guindon and Gascuel
- ReadSeq by Don Gilbert (http://iubio.bio.indiana.edu/soft/molbio/readseq/java/).
- Consense by Joe Felnsenstein
- Table utilities by Philip Milne.;
- BrowserLauncher by Eric Albert.

11. References

Abdo, Z., V. N. Minin, P. Joyce, and J. Sullivan. 2005. Accounting for uncertainty in the tree topology has little effect on the decision-theoretic approach to model selection in phylogeny estimation. Molecular Biology and Evolution **22**:691-703.

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control **19**:716-723.

Alfaro, M. E., and J. P. Huelsenbeck. 2006. Comparative performance of Bayesian and AIC-based measures of phylogenetic model uncertainty. Systematic Biology **55**:89-96.

Burnham, K. P., and D. R. Anderson. 2003. Model selection and multimodel inference: a practical information-theoretic approach. Springer-Verlag, New York, NY.

Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag, New York, NY.

Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and reliability. Annual Review of Genetics **22**:521-565.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. Journal of Molecular Evolution **17**:368-376.

Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package). Department of Genome Sciences, University of Washington, Seattle.

Gascuel, O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Molecular Biology and Evolution **14**:685-695.
Gilbert, D. 2007. ReadSeq. Indiana University, Bloomington.
Goldman, N. 1993a. Simple diagnostic statistical test of models of DNA substitution. Journal of Molecular Evolution **37**:650-661.
Goldman, N. 1993b. Statistical tests of models of DNA substitution. Journal of Molecular Evolution **36**:182-198.
Goldman, N., and S. Whelan. 2000. Statistical tests of gamma-distributed rate heterogeneity in models of sequence evolution in phylogenetics. Molecular Biology and Evolution **17**:975-978.
Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology **52**:696-704.
Hasegawa, M., K. Kishino, and T. Yano. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution **22**:160-174.
Hoeting, J. A., D. Madigan, and A. E. Raftery. 1999. Bayesian model averaging: A tutorial. Statistical Science **14**:382-417.
Huelsenbeck, J. P., B. Larget, and M. E. Alfaro. 2004. Bayesian Phylogenetic Model Selection Using Reversible Jump Markov Chain Monte Carlo. Molecular Biology and Evolution **21**:1123-1133.
Hurvich, C. M., and C.-L. Tsai. 1989. Regression and time series model selection in small samples. Biometrika **76**:297-307.
Johnson, J. B., and K. S. Omland. 2003. Model selection in ecology and evolution. Trends in Ecology & Evolution **19**:101-108.
Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pp. 21-132 *in* H. M. Munro, ed. Mammalian Protein Metabolism. Academic Press, New York, NY.
Kendall, M., and A. Stuart. 1979. The Advanced Theory of Statistics. Charles Griffin, London.
Kimura, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proceedings of the National Academy of Sciences, U.S.A. **78**:454-458.
Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution **16**:111-120.
Kullback, S., and R. A. Leibler. 1951. On information and sufficiency. Annals of Mathematical Statistics **22**:79-86.
Madigan, D. M., and A. E. Raftery. 1994. Model selection and accounting for model uncertainty in graphical models using Occam's Window. Journal of the American Statistical Association **89**:1335-1346.
Minin, V., Z. Abdo, P. Joyce, and J. Sullivan. 2003. Performance-based selection of likelihood models for phylogeny estimation. Systematic Biology **52**:674-683.
Ohta, T. 1992. Theoretical study of near neutrality. II. Effect of subdivided population structure with local extinction and recolonization. Genetics:917-923.
Pol, D. 2004. Empirical problems of the hierarchical likelihood ratio test for model selection. Systematic Biology **53**:949-962.
Posada, D. 2001. The effect of branch length variation on the selection of models of molecular evolution. Journal of Molecular Evolution **52**:434-444.
Posada, D. 2003. Using Modeltest and PAUP* to select a model of nucleotide substitution. Pp. 6.5.1-6.5.14 *in* A. D. Baxevanis, D. B. Davison, R. D. M. Page, G. A. Petsko, L. D. Stein, and G. D. Stormo, eds. Current Protocols in Bioinformatics. John Wiley & Sons, Inc.
Posada, D., and T. R. Buckley. 2004a. Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. Systematic Biology **53**:793-808.
Posada, D., and T. R. Buckley. 2004b. Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology **53**:793-808.
Posada, D., and K. A. Crandall. 2001a. Selecting the best-fit model of nucleotide substitution. Systematic Biology **50**:580-601.
Posada, D., and K. A. Crandall. 2001b. Selecting models of nucleotide substitution: an application to human immunodeficiency virus 1 (HIV-1). Molecular Biology and Evolution **18**:897-906.
Raftery, A. E. 1996. Hypothesis testing and model selection. Pp. 163-187 *in* W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, eds. Markov chain Monte Carlo in practice. Chapman & Hall, London ; New York.
Schwarz, G. 1978. Estimating the dimension of a model. The Annals of Statistics **6**:461-464.
Sugiura, N. 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. Communications in Statistics–Theory and Methods **A7**:13-26.
Sullivan, J., and P. Joyce. 2005. Model selection in phylogenetics. Annual Review of Ecology, Evolution and Systematics **36**:445-466.

Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution **10**:512-526.

Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Pp. 57-86 *in* R. M. Miura, ed. Some mathematical questions in biology - DNA sequence analysis. Amer. Math. Soc., Providence, RI.

Wasserman, L. 2000. Bayesian Model Selection and Model Averaging. Journal of Mathematical Psychology **44**:92-107.

Whelan, S., and N. Goldman. 1999. Distributions of statistics used for the comparison of models of sequence evolution in phylogenetics. Molecular Biology and Evolution **16**:1292-1299.

Yang, Z., N. Goldman, and A. Friday. 1995. Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. Systematic Biology **44**:384-399.

Zharkikh, A. 1994. Estimation of evolutionary distances between nucleotide sequences. Journal of Molecular Evolution **39**:315-329.