LETTERS

jModelTest: Phylogenetic Model Averaging

David Posada
Departamento de Genética, Bioquímica e Inmunología, Facultad de Biología, Universidad de Vigo, Vigo, Spain

jModelTest is a new program for the statistical selection of models of nucleotide substitution based on “Phyml” (Guindon and Gascuel 2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52:696–704. It implements 5 different selection strategies, including “hierarchical and dynamical likelihood ratio tests,” the “Akaike information criterion,” the “Bayesian information criterion,” and a “decision-theoretic performance-based” approach. This program also calculates the relative importance and model-averaged estimates of substitution parameters, including a model-averaged estimate of the phylogeny. jModelTest is written in Java and runs under Mac OSX, Windows, and Unix systems with a Java Runtime Environment installed. The program, including documentation, can be freely downloaded from the software section at http://darwin.uvigo.es.

Introduction

Models of nucleotide substitution allow for the calculation of probabilities of change between nucleotides along the branches of a phylogenetic tree. The use of a particular substitution model may change the outcome of the phylogenetic analysis (e.g., Buckley 2002; Buckley and Cunningham 2002; Lemmon and Moriarty 2004), and statistical model selection has become an essential step for the estimation of phylogenies from DNA sequence alignments. In-depth reviews about model selection in phylogenetics are available elsewhere (Johnson and Omland 2003; Posada and Buckley 2004; Sullivan and Joyce 2005). Indeed, the performance of different model selection strategies has been the subject of active research (Posada 2001; Posada and Crandall 2001; Pol 2004; Abdo et al. 2005; Alfaro and Huelsenbeck 2006).

Several programs already exist for the statistical selection of models of nucleotide substitution (e.g., Nylander 2004; Keane et al. 2006). Among these, Modeltest (Posada and Crandall 1998) has been one of the most popular. This note describes a new program called jModelTest that supersedes Modeltest in several aspects. jModelTest allows for the definition of restricted sets of candidate models (table 1), implements customizable “hierarchical likelihood ratio tests” (hLRTs) (Fratti et al. 1997; Huelsenbeck and Crandall 1997; Sullivan et al. 1997) and “dynamic likelihood ratio tests” (dLRTs) (Posada and Crandall 2001), provides a rank of models according to the “Akaike Information Criterion” (AIC) (Akaike 1973), to the “Bayesian Information Criterion” (BIC) (Schwarz 1978) or to a “decision-theoretic performance-based” approach (DT) (Minin et al. 2003) (table 2), calculates the relative importance of every parameter, and computes model-averaged estimates of these, including a model-averaged estimate of the tree topology (Posada and Buckley 2004).

Model Selection with jModelTest

jModelTest is essentially a front-end to a computational pipeline that takes advantage of existing programs for running different tasks. Basically, this pipeline (fig. 1) includes:

- “ReadSeq” (Gilbert 2007): for conversion among different DNA sequence alignment formats.
- “Phyml” (Guindon and Gascuel 2003): for the likelihood calculations, including estimates of model parameters and trees.
- “Consense” (from the PHYLIP package) (Felsenstein 2005): to calculate weighted and strict consensus trees representing model-averaged phylogenies.

Likelihood Calculations

Likelihood calculations, including model parameters and tree estimates, are carried out with Phyml (Guindon and Gascuel 2003). The tree topology used in these calculations can be the same across models (fixed) or optimized for each one. Fixed tree topologies can be estimated with the BIONJ algorithm (Gascuel 1997) upon JC distances (Jukes and Cantor 1969) or user-defined. Alternatively, a BIONJ or an ML tree can be estimated under each model. In all cases, branch lengths are estimated and counted as parameters.

Custom Set of Models

Currently, there are 11 different nucleotide substitution schemes implemented in jModelTest, which combined with equal or unequal base frequencies (+F), a proportion of invariable sites (+I), and rate variation among sites (+G), result in 88 distinct models (table 1). The program offers the possibility of defining to a reasonable extent which models are included in the candidate set.

Sequential Likelihood Ratio Tests

A series of likelihood ratio tests (LRTs) can be implemented under a particular hierarchy (hLRTs), in which the user can specify their order, and whether parameters are added (forward selection) or removed (backward selection). Alternatively, the order of the LRTs can be set dynamically (dLRTs) (Posada and Crandall 2001), 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. The hLRTs and dLRTs will be available only if the likelihood scores were calculated upon a fixed topology, due to the nesting requirement of the $\chi^2$ approximation.

### Information Criteria

The program implements 3 different information criteria: the AIC (Akaike 1973), the BIC (Schwarz 1978), and a performance-based approach based on decision theory (DT) (Minin et al. 2003). Under the AIC framework, there is also the possibility of using a corrected version for small samples (AICc) (Sugiura 1978; Hurvich and Tsai 1989), instead of the standard AIC. In this case, sample size has to be specified, which by default is approximated as the number of sites in the alignment (note that the sample size of an alignment is presently an unknown quantity).

#### Table 1

<table>
<thead>
<tr>
<th>Substitution Models Available in jModelTest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
</tr>
<tr>
<td>JC</td>
</tr>
<tr>
<td>F81</td>
</tr>
<tr>
<td>K80</td>
</tr>
<tr>
<td>HKY</td>
</tr>
<tr>
<td>TrNe</td>
</tr>
<tr>
<td>TPM1</td>
</tr>
<tr>
<td>TPM1u</td>
</tr>
<tr>
<td>TPM2</td>
</tr>
<tr>
<td>TPM2u</td>
</tr>
<tr>
<td>TPM3</td>
</tr>
<tr>
<td>TPM3u</td>
</tr>
<tr>
<td>TIM1</td>
</tr>
<tr>
<td>TIM1e</td>
</tr>
<tr>
<td>TIM2</td>
</tr>
<tr>
<td>TIM2e</td>
</tr>
<tr>
<td>TIM3</td>
</tr>
<tr>
<td>TIM3e</td>
</tr>
<tr>
<td>TVM</td>
</tr>
<tr>
<td>TVMe</td>
</tr>
<tr>
<td>SYM</td>
</tr>
<tr>
<td>GTR</td>
</tr>
</tbody>
</table>

**Note.**—The same number of branch lengths (k) needs to be estimated for every model.


- Any of these can include invariable sites (+I), rate variation among sites (+G), or both (+I+G).

- Equal AC = 0; Unequal AC, AG, AT, CG, CT, GT = 1, 2, 3, 4, 5.

#### Table 2

<table>
<thead>
<tr>
<th>Model Selection Strategies Implemented in jModelTest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hierarchical Likelihood Ratio Tests</strong></td>
</tr>
<tr>
<td>Abbreviation</td>
</tr>
<tr>
<td>Base tree</td>
</tr>
<tr>
<td>Nesting requirement</td>
</tr>
<tr>
<td>Simultaneous comparison</td>
</tr>
<tr>
<td>Selection uncertainty</td>
</tr>
<tr>
<td>Parameter importance</td>
</tr>
<tr>
<td>Model averaging</td>
</tr>
</tbody>
</table>

* DT weights are simply the rescaled reciprocal DT scores. This is a gross implementation very likely to change.
model-averaged estimates of these, using all the models in the candidate set, or a fraction included in a particular CI (see Posada and Buckley 2004).

Model-Averaged Phylogenies

jModelTest is able to compute an average estimate of the tree topology by building a consensus of the maximum likelihood (ML) trees for every model in the candidate set, weighting them with their model weights (AIC, BIC, or DT) (fig. 2). Indeed, this option is only available when the tree topology has been optimized for every model. The consensus tree is constructed using the Consense program from the PHYLIP package (Felsenstein 2005).

Software Platform and Availability

jModelTest is written in Java and can be started in any operating system with a Java Runtime Environment (see http://www.java.com). However, jModelTest uses other programs for different tasks, and these have been compiled for Mac OSX, Windows XP, and Linux. The package, including installation instructions, documentation, executables, and example data, is distributed free of charge for academic use from the software section at http://darwin.uvigo.es.

Conclusions

Model selection is an important issue in statistical phylogenetics, around which some questions still remain open (Kelchner and Thomas 2007). jModelTest addresses some of these, providing an increased flexibility for the user to explore the data and the role of the substitution model on the estimation of phylogenetic trees.

Fig. 1.—jModelTest pipeline. Alignments are loaded using the ReadSeq library (Gilbert 2007). Likelihood calculations, including estimates of model parameters and trees, are carried out with Phyml (Guindon and Gascuel 2003). A custom program called Ted (D. Posada) is used to compute Euclidean distances between trees for performance-based model selection (DT), whereas Consense (Felsenstein 2005) is used to calculate weighted and strict consensus trees representing model-averaged phylogenies.

Fig. 2.—Model-averaged tree of HIV-1 pol sequences. The topology shown is the consensus of 88 ML tree topologies, one for every model, weighted according to the AIC weights. The numbers on the branches represent uncertainty due to model selection. In this case, clades (AJ), (AJC), and (HG) are supported by the best and fourth best AIC models (GTR + G, AIC weight = 0.83; TIM3 + G, AIC weight = 0.01; respectively) and others, but not by the second or third best AIC models (GTR + I + G, AIC weight = 0.15; GTR + I, AIC weight = 0.01; respectively).
Acknowledgments

I want to thank a number of users of Modeltest that had made numerous comments and suggestions through the years. Special thanks to Stephane Guindon for his generous help with Phyml and to John Huelsenbeck for suggesting the stochastic calculation of CIs. I want to acknowledge Sudhir Kumar for inviting me to present the latest advances in Modeltest at the 2006 SMBE annual meeting, which finally prompted the completion of jModelTest.

Literature Cited


Sudhir Kumar, Associate Editor

Accepted April 2, 2008