Modeltest

Version 3.06

© David Posada

Department of Zoology, Brigham Young University, Provo UT 84602, USA

dp47@email.byu.edu

Disclaimer

Version 1-1.06 (June 98): several aesthetic variations

Version 2.0 (June 99): 40 models are included

Version 2.1 (October 99): A bug in the selection of the minimum AIC, which implied that the model GTR+I+G could not be selected is solved. Also, the number of free parameter is redefined. Now JC is considered to have 0 free parameters and GTR+I+G 10. This would affect only a few AIC calculations.

Version 3.0 beta 1 (December 99): 16 new models are added for a total now of 56 models. These models are variations of two main substitution schemes that I called TIM (transitional model) and TIV (transversional model). They are described in figure 1 below.

Version 3.0 beta 2 (December 99): Because in the new release of PAUP beta3, the likelihood score file includes now base frequencies estimates, the program was modified accordingly. The likelihood ratio tests are now explained with more detail, and the output of Modeltest is more consistent with the likelihood settings option in PAUP*.

Version 3.0 (February 99): several cosmetic variations. The output of Modeltest is designed now to specify the model selected accordingly to PAUP* likelihood settings.

Version 3.04 (July 00): The program outputs now a block of commands to implement the likelihood settings in PAUP* for the best-fit model selected. The mixed chi-square distribution is added as default for the I and G tests.

Version 3.05 (Feb 01): In the windows version, the AIC[55] gave an AIC of 0 to the GTRIG. Now dimension is AIC[56] (Juan Suarez). TIM+G reported invariable sites instead of gamma shape (Cymon Cox)

Version 3.06 (Apr 01): Print likelihood scores by default. In the windows version there was a bug by which the file scores.txt was always the standard input (Andy Vierstraete). Using GNU licencese (I should have done this a long time ago) (thanks to Naoki Takebayashi)

<u>**Note than only the last versions of Modeltest (3.x) are compatible with the new PAUP* version (4.0)**</u>

<u>This tool is provided without warranty of any kind. In no event shall the author be held responsible for any damages resulting from the use of this software.</u>

I really appreciate the input from several users and would like to show my appreciation. Thank you very much !!.

Introduction

All phylogenetic methods make assumptions, whether explicit or implicit, about the process of DNA substitution (Felsenstein 1988). For example, an assumption common to many phylogenetic methods is a bifurcating tree to describe the phylogeny of species (Huelsenbeck and Crandall 1997). Consequently, all the methods of phylogenetic inference depend on their underlying models. To have confidence in inferences it is necessary to have confidence in the models (Goldman 1993b). Because of this, all the methods based on explicit models of evolution should explore which is the model that fits the data best, justifying then its use. In traditional statistical theory, a widely accepted statistic for testing the goodness of fit of models is the likelihood ratio test statistic $\delta = -2 \log \Lambda$, being

$$\Lambda = \frac{\max\left[L_0\left(NullModel \mid Data\right)\right]}{\max\left[L_1\left(AlternativeModel \mid Data\right)\right]}$$

where L_0 is the likelihood under the null hypothesis (simple model) and L_1 is the likelihood under the alternative hypothesis (more complex, parameter rich, model). The value of this statistic is always equal to or greater than zero, as the likelihood under the more complex model will always be equal or higher than the likelihood under the simpler model. When the models compared are nested (the null hypothesis is a special case of the alternative hypothesis) and the null hypothesis is correct, the δ statistic is asymptotically distributed as χ^2 with q degrees of freedom, where q is the difference in number of free parameters between the two models; equivalently, q is the number of restrictions on the parameters of the alternative hypothesis required to derive the particular case of the null hypothesis (Goldman 1993b; Kendall and Stuart 1979). To preserve the nesting of the models, the likelihood scores are estimated upon the same tree, and then, once the models have been compared, a final tree is estimated using the chosen model of evolution.

When the models are not nested, an alternative means of generating the null distribution of the δ statistic is through Monte Carlo simulation (parametric bootstrapping) (Goldman 1993a). Another way of comparing different models without the nested requirement is the Akaike information criterion (minimum theoretical information criterion, AIC) (Akaike 1974). The AIC is a useful measure that rewards models for good fit, but imposes a penalty for unnecessary parameters (Hasegawa 1990a; Hasegawa 1990b; Hasegawa et al. 1991; Muse 1999). If L is the maximum value of the likelihood function for a specific model using n independently adjusted parameters within the model, then AIC= $-2\ln L + 2n$. Smaller values of AIC indicate better models.

The χ^2 approximation used to represent the underlying distribution of the LRT has been problematic. Goldman (1993b) first pointed out the difficulty in counting the number of degrees of freedom and the problem of the sparseness of the DNA data. Later, Yang et al. (1995) suggested that the LRT was well represented by a χ^2 distribution. However, the χ^2 distribution may not be reliable when the null model is equivalent to fixing some parameters at the boundary of the parameter space of the alternative model, e.g., rate homogeneity test, where the null

hypothesis is a special case of the gamma-distribution model with shape parameter (α) equal to infinity (Yang 1996). Whelan and Goldman (1999) have also shown that for comparisons of rate variation across sites and nucleotide frequencies estimated as the observed base frequencies, the χ^2 distribution was significantly different from the true distribution, and the tests were conservative. To account for the boundary problem Ota et al. (1992)and Goldman and Whelan (2000) suggested the use of a mixed χ^2 (or $\overline{\chi}^2$) distribution consisting of 50% χ_0^2 and 50% χ_1^2 to construct LRT tests for the invariable sites. *This distribution for model fitting is implemented in the Modeltest 3.04 and future versions as the default distribution for the invariable sites and rate heterogeneity among sites LRTs.*

The Program
MODELTEST is a simple calculator written in ANSI C and compiled for the Power Macintosh and Windows 95/NT using Metrowerks CodeWarrior and for Sun machines using gcc. It is designed to compare different nested models of DNA substitution in a hierarchical hypothesis-testing framework (Figure 1). MODELTEST calculates the likelihood ratio test statistic $\delta = -2$ log Λ and its associated *P*-value using a χ^2 distribution with *q* degrees of freedom in order to reject or fail to reject different null hypothesis about the process of DNA substitution. It also calculates the AIC estimate associated with each likelihood score.

Usage
The user communicates with the program using a standard console interface (Figure 2), in Macintosh machines where the input and output files can be specified, By clicking with the mouse in the left **File** button, the user can select an INPUT FILE. By clicking in the right **File** button, the user can specify an OUTPUT FILE (the default output is to the **Console** in the screen). In the **Argument** line the user can interact with the program. In Windows (DOS window) (Figure 3) and Unix machines the users interacts with the program through a command or argument line (Figure 4). These are the options:

    -d : Debug level (e.g. -d2)
    -a : alpha level (e.g. -a0.01)
    -c : Likelihood Ratio calculator mode
    -m : Turn off use of mixed chi square for I and G LRTs
    -i : AIC calculator mode
    -l : Prints likelihoods for all models
    -f : Input from a file for obtaining AIC values
    -? : Help

Default Mode
By default, the program will accept two classes of input files: a file containing ordered raw log likelihood scores corresponding to the tested models (JC, JC+I, JC+G, JC+I+G, K80, K80+I, K80+G, K80+I+G, TrNef, TrNef+I, TrNef+G, TrNef+I+G, K81, K81+I, K81+G, K81+I+G, TVMef, TVMef+I, TVMef+G, TVMef+I+G, TIMef, TIMef+I, TIMef+G, TIMef+I+G, SYM, SYM+I, SYM+G, SYM+I+G, F81, F81+I, F81+G, F81+I+G, HKY, HKY+I, HKY+G, HKY+IΓ, TrN, TrN+I, TrN+G, TrN+I+G, K81uf, K81uf+I, K81uf+G, K81uf+I+G, TVM,

TVM+I, TVM+G, TVM+I+G, TIM, TIM+I, TIM+G, TIM+I+G GTR, GTR+I, GTR+G, GTR+I+G; I: invariable sites; G: gamma distribution; see Figure 1 for abbreviations) and a PAUP* (Swofford 1998) file containing a matrix of the same log likelihood scores resulting from the execution of a block of PAUP* commands. This block of PAUP commands is included in the package (but see Appendix 1).

Alpha level (-a)
The user can set the alpha level of significance (by default 0.01) in the command line, inputting –a followed by the desired value. The program will use this level of significance in all its calculations.

Likelihood Ratio Calculator Mode (-c)
This is a useful mode when the user wants to calculate likelihood ratio tests and their associate probability for different hypotheses, or only some of the hypothesis tested by default by Modeltest. As the P-value is calculated using a chi-square distribution, **the models tested should be nested** (the null hypothesis is a special case of the alternative hypothesis). The user is guided by prompts for inputting a pair of likelihood scores and the number of degrees of freedom. Modeltest performs the likelihood ratio test and calculates its associated P-value using a chi-square distribution. In this case the user is responsible for calculating the appropriate number of degrees of freedom and log likelihood scores for testing the intended hypotheses.

AIC Calculator Mode (-i)
This mode is aimed to provide the user with a fast calculation of the AIC values corresponding to the likelihood scores. In this case, the models do not need to be nested. The user is guided by prompts for inputting the number of log likelihood scores, the positive log likelihood scores, and the corresponding number of free parameters.

Input file for obtaining AIC estimates (-f)
If the user is interested in inputting a large number of scores for obtaining the AIC estimates, one can input a file with two columns of data: the positive likelihood scores in one column with the corresponding number of free parameters in the same row constituting the second column.

Prints likelihood scores (-l)
If the user is interested in looking at the likelihood scores, by specifying this option, a list of all the models and their corresponding likelihood scores is printed.

Alpha level (-a)
The user can set the alpha level of significance (by default 0.01) in the command line, inputting "–a" followed by the desired value (for example, -a0.05). The program will use this level of significance in all its calculations.

Chi-square distribution (-m)
By default, the program uses now the mixed chi-square distribution (see above) as the default distribution to obtain P-values for the LRT of the invariable sites (I) and rate heterogeneity

among sites (G) hypotheses. By typing "-m" in the command line, the standard chi-square distribution is used also for the I and G hypotheses.

<u>Help (-?)</u>
By typing "-?" in the command line, the user can have access to some help. Nevertheless, the lecture of the manual is encouraged…

<u>Debug level (-d)</u>
The debug level is a programming feature that does not affect the calculations. If you don't know what is this, don't worry; simply don't use it…

Output
The output of MODELTEST consists of a description of the likelihood ratio tests performed, and their associated *P*-values. The null hypotheses tested are described in Figure 1. The program interprets the resulting *P*-values and chooses the model that fits the data best among those tested following the likelihood ratio test and/or AIC criteria, using a default individual alpha value of 0.01 (for maintaining an overall alpha value of 0.05, the standard Bonferroni correction - alpha/number of tests- results in a individual alpha value of 0.01), or another value specified by the user. The program also calculates the AIC values, indicating the smallest. Note that when the equal ti/tv (transition/transversion) rates hypothesis is not rejected, the equal ti and equal tv rates hypotheses are automatically rejected, and then will not be tested.
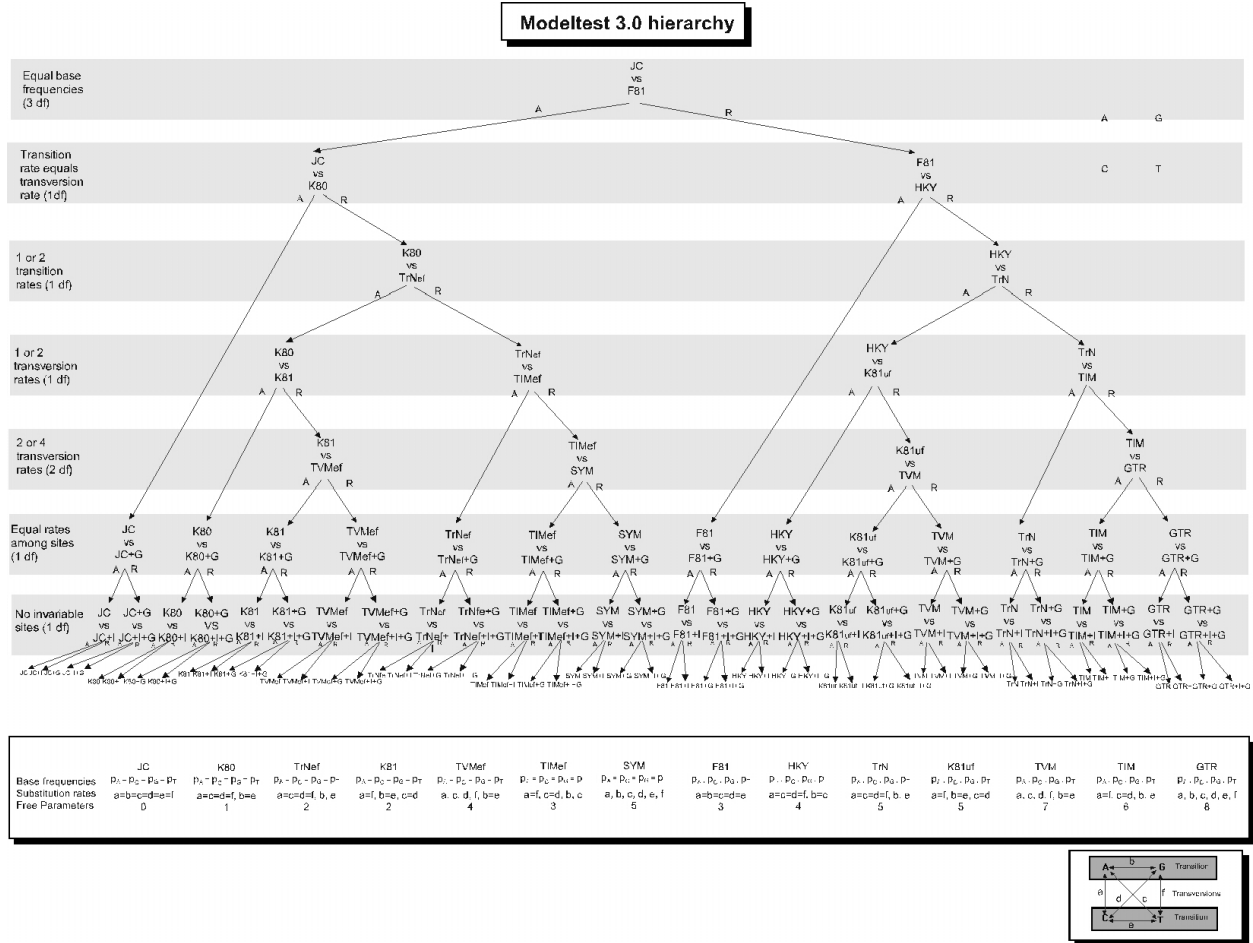
Modeltest hierarchy

Figure 1. Hierarchical hypothesis testing in MODELTEST. At each level the null hypothesis (upper model) is either accepted (A) or rejected (R). The models of DNA substitution are: JC (Jukes and Cantor 1969), K80 (Kimura 1980), TrNef (TrN equal base frequencies; see below), K81 (Kimura 1981), TIMef (TIM with equal base frequencies), TIV (TIV with equal base frequencies), SYM (Zharkikh 1994), F81 (Felsenstein 1981), HKY (Hasegawa et al. 1985), TrN (Tamura and Nei 1993), K81uf (K81 unequal base frequencies; see above), TIM, TIV, and GTR (Rodríguez et al. 1990). G: shape parameter of the gamma distribution; I: proportion of invariable sites. df: degrees of freedom.
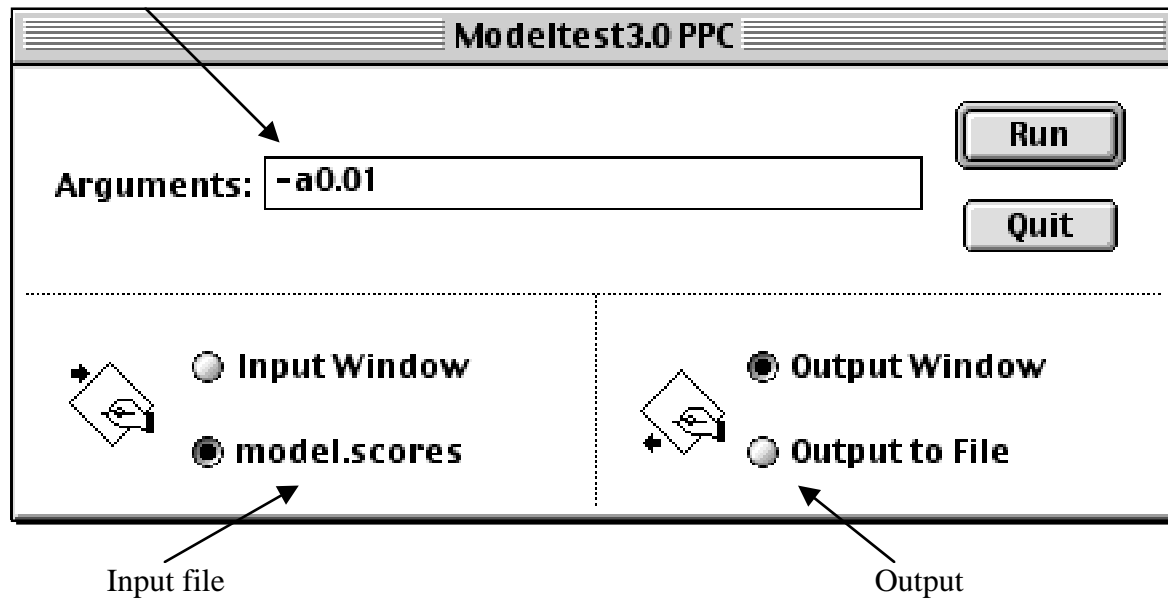
MACINTOSH

Command line

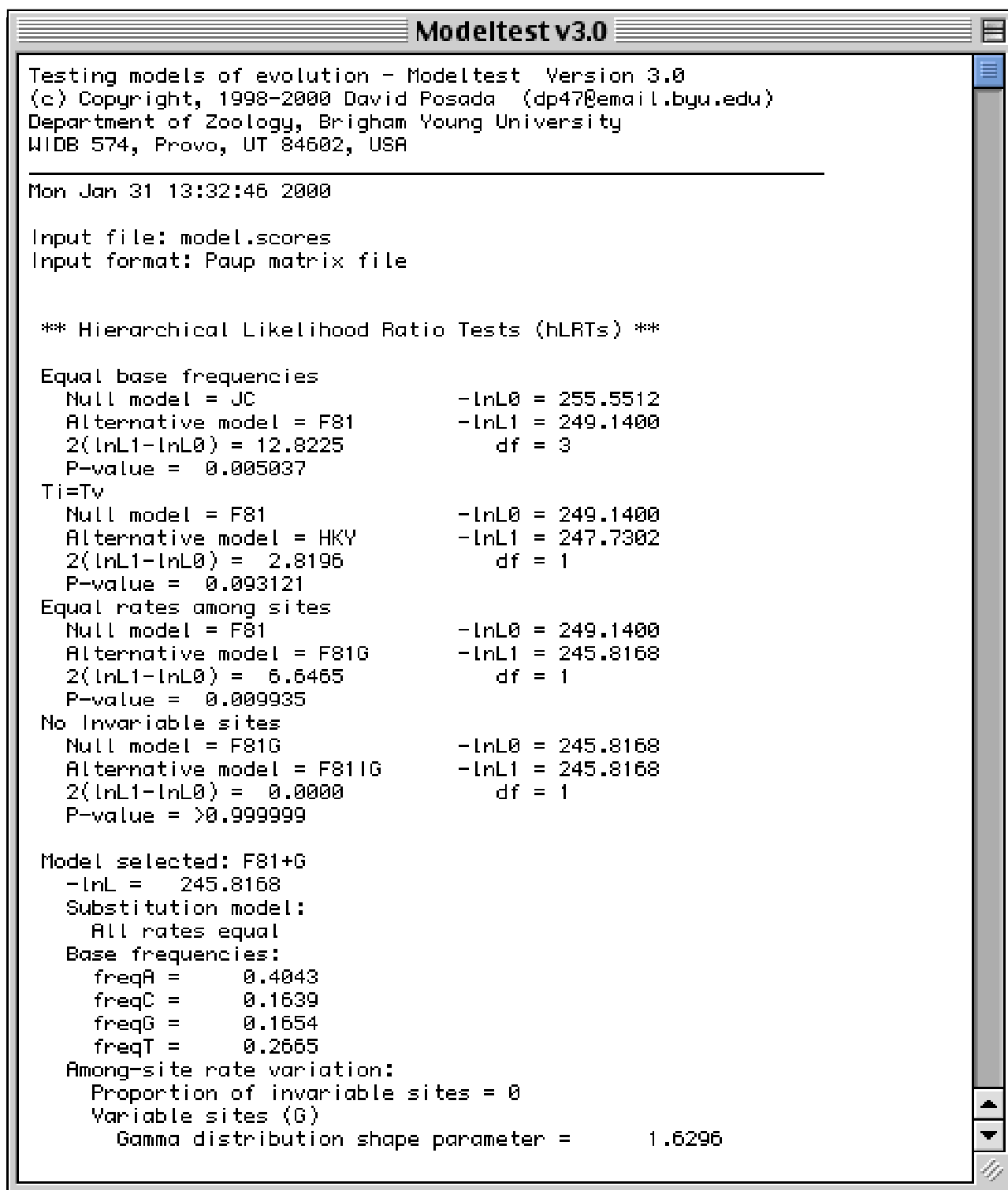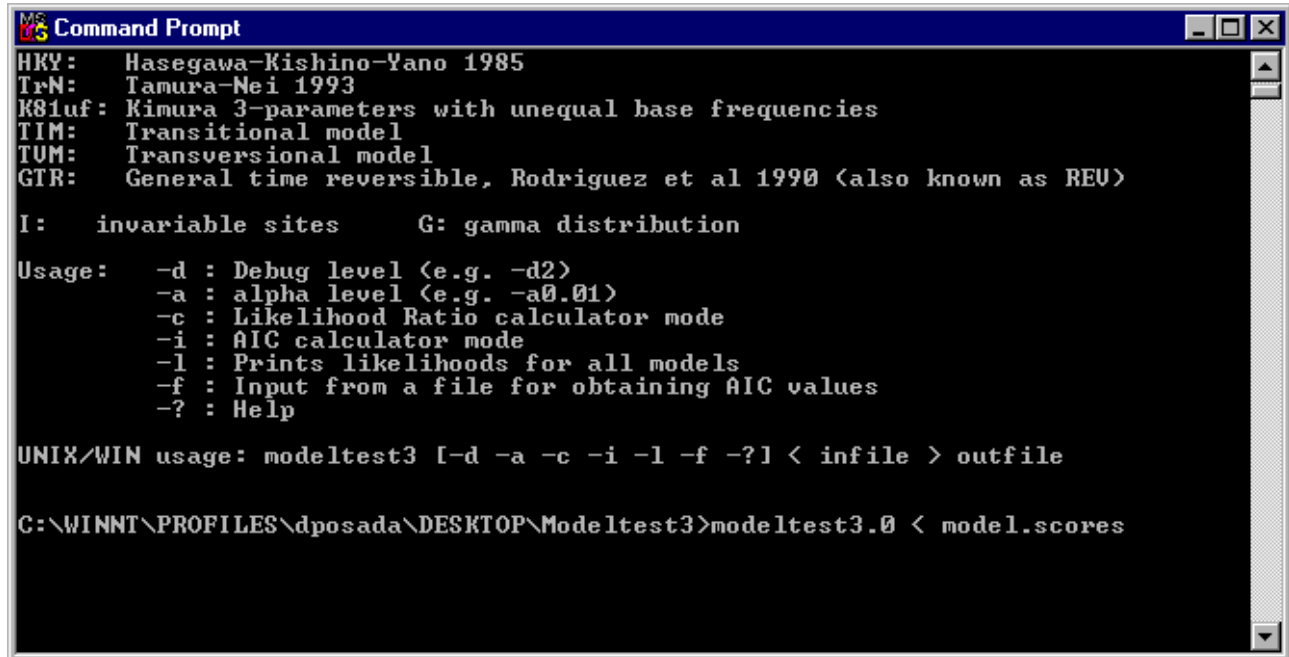Figure 2. Console Interface for Macintosh

```
═══════════════ Modeltest v3.0 ═══════════════

Testing models of evolution - Modeltest  Version 3.0
(c) Copyright, 1998-2000 David Posada  (dp47@email.byu.edu)
Department of Zoology, Brigham Young University
WIDB 574, Provo, UT 84602, USA
_____

Mon Jan 31 13:32:46 2000

Input file: model.scores
Input format: Paup matrix file


** Hierarchical Likelihood Ratio Tests (hLRTs) **

Equal base frequencies
  Null model = JC                    -lnL0 = 255.5512
  Alternative model = F81            -lnL1 = 249.1400
  2(lnL1-lnL0) = 12.8225               df = 3
  P-value =  0.005037
Ti=Tv
  Null model = F81                   -lnL0 = 249.1400
  Alternative model = HKY            -lnL1 = 247.7302
  2(lnL1-lnL0) =  2.8196               df = 1
  P-value =  0.093121
Equal rates among sites
  Null model = F81                   -lnL0 = 249.1400
  Alternative model = F81G           -lnL1 = 245.8168
  2(lnL1-lnL0) =  6.6465               df = 1
  P-value =  0.009935
No Invariable sites
  Null model = F81G                  -lnL0 = 245.8168
  Alternative model = F81IG          -lnL1 = 245.8168
  2(lnL1-lnL0) =  0.0000               df = 1
  P-value = >0.999999

Model selected: F81+G
  -lnL =    245.8168
  Substitution model:
    All rates equal
  Base frequencies:
    freqA =      0.4043
    freqC =      0.1639
    freqG =      0.1654
    freqT =      0.2665
  Among-site rate variation:
    Proportion of invariable sites = 0
    Variable sites (G)
      Gamma distribution shape parameter =      1.6296
```

Figure 3. Macintosh output

WINDOWS

```
Command Prompt                                                    _ □ ×
HKY:    Hasegawa-Kishino-Yano 1985
TrN:    Tamura-Nei 1993
K81uf: Kimura 3-parameters with unequal base frequencies
TIM:    Transitional model
TVM:    Transversional model
GTR:    General time reversible, Rodriguez et al 1990 (also known as REV)

I:    invariable sites      G: gamma distribution

Usage:    -d : Debug level (e.g. -d2)
          -a : alpha level (e.g. -a0.01)
          -c : Likelihood Ratio calculator mode
          -i : AIC calculator mode
          -l : Prints likelihoods for all models
          -f : Input from a file for obtaining AIC values
          -? : Help

UNIX/WIN usage: modeltest3 [-d -a -c -i -l -f -?] < infile > outfile


C:\WINNT\PROFILES\dposada\DESKTOP\Modeltest3>modeltest3.0 < model.scores
```
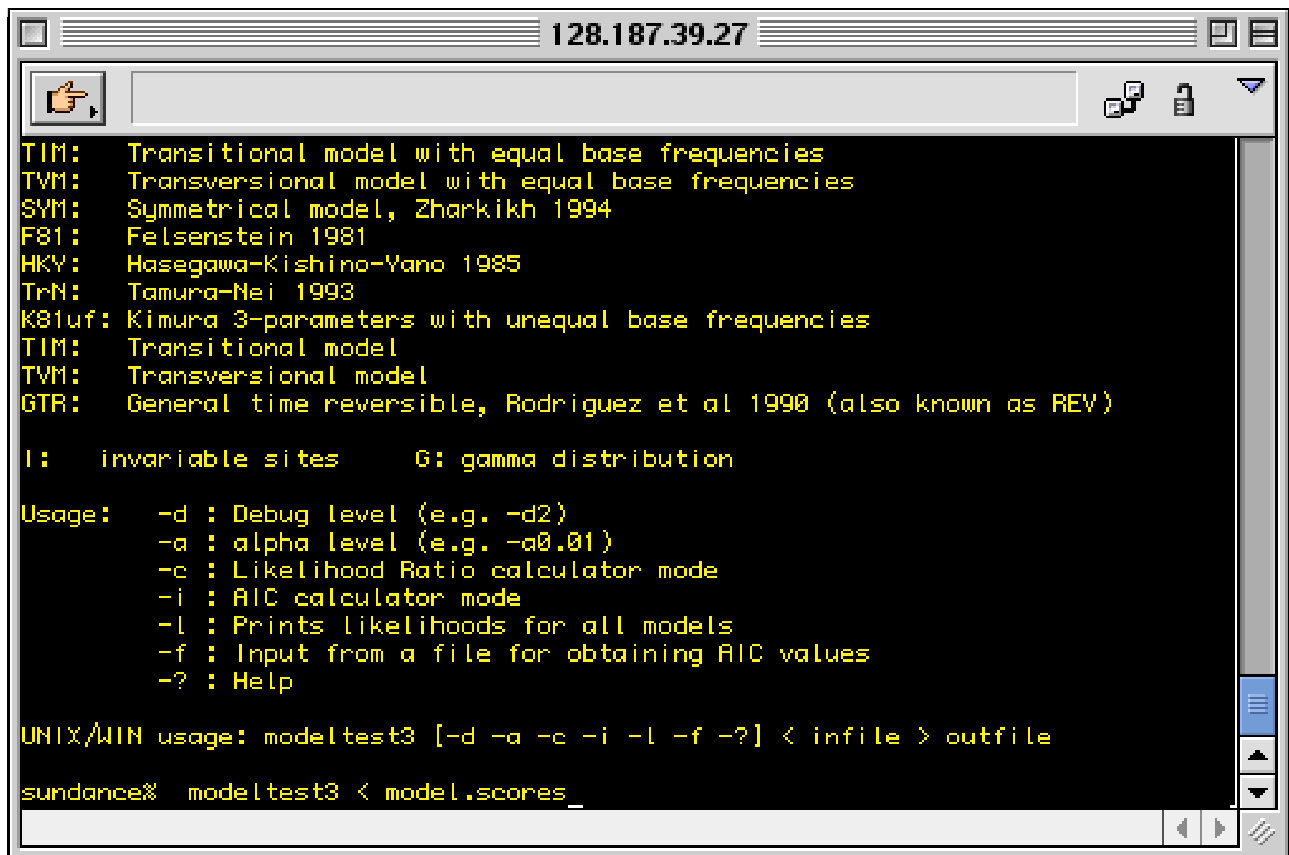
Figure 4. Command prompt in Windows

UNIX

```
128.187.39.27
TIM:    Transitional model with equal base frequencies
TVM:    Transversional model with equal base frequencies
SYM:    Symmetrical model, Zharkikh 1994
F81:    Felsenstein 1981
HKY:    Hasegawa-Kishino-Yano 1985
TrN:    Tamura-Nei 1993
K81uf: Kimura 3-parameters with unequal base frequencies
TIM:    Transitional model
TVM:    Transversional model
GTR:    General time reversible, Rodriguez et al 1990 (also known as REV)

I:    invariable sites      G: gamma distribution

Usage:    -d : Debug level (e.g. -d2)
          -a : alpha level (e.g. -a0.01)
          -c : Likelihood Ratio calculator mode
          -i : AIC calculator mode
          -l : Prints likelihoods for all models
          -f : Input from a file for obtaining AIC values
          -? : Help

UNIX/WIN usage: modeltest3 [-d -a -c -i -l -f -?] < infile > outfile

sundance%  modeltest3 < model.scores
```

Figure 5. Unix console

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Contr 19:716-723

Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. J Mol Evol 17:368-376

Felsenstein J (1988) Phylogenies from molecular sequences: inference and reliability. Annu Rev Genet 22:521-565

Goldman N (1993a) Simple diagnostic statistical test of models of DNA substitution. J Mol Evol 37:650-661

Goldman N (1993b) Statistical tests of models of DNA substitution. J Mol Evol 36:182-198

Goldman N, Whelan S (2000) Statistical tests of gamma-distributed rate heterogeneity in models of sequence evolution in phylogenetics. Mol Biol Evol 17:975-978

Hasegawa M (1990a) Mitochondrial DNA evolution in primates: transition rate has been extremely low in the lemur. J Mol Evol 31:113-121

Hasegawa M (1990b) Phylogeny and molecular evolution in primates. Jpn J Genet 65:243-266

Hasegawa M, Kishino H, Saitou N (1991) On the maximum likelihood method in molecular phylogenetics. J Mol Evol 32:443-445

Hasegawa M, Kishino K, Yano T (1985) Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160-174

Huelsenbeck JP, Crandall KA (1997) Phylogeny estimation and hypothesis testing using maximum likelihood. Annu Rev Ecol Syst 28:437-466

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HM (ed) Mammalian Protein Metabolism. Academic Press, New York, NY, p 21-132

Kendall M, Stuart A (1979) The Advanced Theory of Statistics. Charles Griffin, London

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111-120

Kimura M (1981) Estimation of evolutionary distances between homologous nucleotide sequences. Proc Natl Acad Sci USA 78:454-458

Muse S (1999) Modeling the molecular evolution of HIV sequences. In: Crandall KA (ed) The Evolution of HIV. Johns Hopkins University Press, Baltimore, MD, p 122-152

Ohta T (1992) Theoretical study of near neutrality. II. Effect of subdivided population structure with local extinction and recolonization. Genetics:917-923

Rodríguez F, Oliver JF, Marín A, Medina JR (1990) The general stochastic model of nucleotide substitution. J Theor Biol 142:485-501

Swofford DL (1998) PAUP* Phylogenetic analysis using parsimony and other methods. Sinauer Associates, Sunderland, MA

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512-526

Whelan S, Goldman N (1999) Distributions of statistics used for the comparison of models of sequence evolution in phylogenetics. Mol Biol Evol 16:1292-1299

Yang Z (1996) Maximum-likelihood models for combined analyses of multiple sequence data. J Mol Evol 42:587-596

Yang Z, Goldman N, Friday A (1995) Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. Syst Biol 44:384-399

Zharkikh A (1994) Estimation of evolutionary distances between nucleotide sequences. J Mol Evol 39:315-329

APENDIX 1: RUNNING THE PAUP COMMANDS BLOCK

The input of Modeltest is likelihood scores corresponding to the specific data set and each one of 40 models. The easiest way of obtaining these scores from an alignment of DNA sequences is using PAUP*. A block of commands for PAUP* is provided below in the commands file ("modelblock3"). Follow these steps:

1) Open your data file and execute it in PAUP
2) Open the command file (modelblock) and execute it
3) Paup starts to work in the data following the commands. Once is finished you will see a file called "model.scores" in the same directory as the command file
4) Run the file "model.scores" through modeltest

Alternatively, if you are familiar with PAUP*, you can add the PAUP* commands after your data block directly in your data file and execute it.

The PAUP* commands included in the "modelblock" file are described below. In brackets there are some comments that don't affect the PAUP run. The output obtained (model.scores) from the execution of this block of commands can be directed as input for MODELTEST. **Note that this file is now different from that in versions previous to 3.0.**

#NEXUS

[! ***** MODELFIT BLOCK -- MODELTEST 3.0 *****]

[The following command will calculate a NJ tree using the JC69 model of evolution]

BEGIN PAUP;
        log file= modelfit.log replace;
        DSet distance=JC objective=ME base=equal rates=equal pinv=0
        subst=all negbrlen=setzero;
        NJ showtree=no breakties=random;
End;

[!
***** BEGIN TESTING 56 MODELS OF EVOLUTION *****]

BEGIN PAUP;
Set criterion=like;

[!
** Model 1 of 56 * Calculating JC **]
lscores 1/ nst=1 base=equal rates=equal pinv=0
scorefile=model.scores replace;
[!
** Model 2 of 56 * Calculating JC+I **]
lscores 1/ nst=1 base=equal rates=equal pinv=est
scorefile=model.scores append;

[!
** Model 3 of 56 * Calculating JC+G **]
lscores 1/ nst=1 base=equal rates=gamma shape=est pinv=0
scorefile=model.scores append;
[!
** Model 4 of 56 * Calculating JC+I+G **]
lscores 1/ nst=1 base=equal rates=gamma shape=est pinv=est
scorefile=model.scores append;

[!
** Model 5 of 56 * Calculating F81 **]
lscores 1/ nst=1 base=est rates=equal pinv=0
scorefile=model.scores append;
[!
** Model 6 of 56 * Calculating F81+I **]
lscores 1/ nst=1 base=est rates=equal pinv=est
scorefile=model.scores append;
[!
** Model 7 of 56 * Calculating F81+G **]
lscores 1/ nst=1 base=est rates=gamma shape=est pinv=0
scorefile=model.scores append;
[!
** Model 8 of 56 * Calculating F81+I+G **]
lscores 1/ nst=1 base=est rates=gamma shape=est pinv=est
scorefile=model.scores append;

[!
** Model 9 of 56 * Calculating K80 **]
lscores 1/ nst=2 base=equal tratio=est rates=equal pinv=0
scorefile=model.scores append;
[!
** Model 10 of 56 * Calculating K80+I **]
lscores 1/ nst=2 base=equal tratio=est rates=equal pin=est
scorefile=model.scores append;
[!
** Model 11 of 56 * Calculating K80+G **]
lscores 1/ nst=2 base=equal tratio=est rates=gamma shape=est pinv=0
scorefile=model.scores append;
[!
** Model 12 of 56 * Calculating K80+I+G **]
lscores 1/ nst=2 base=equal tratio=est rates=gamma shape=est pinv=est
scorefile=model.scores append;

[!
** Model 13 of 56 * Calculating HKY **]
lscores 1/ nst=2 base=est tratio=est rates=equal pinv=0
scorefile=model.scores append;

[!
** Model 14 of 56 * Calculating HKY+I **]
lscores 1/ nst=2 base=est tratio=est rates=equal pinv=est
scorefile=model.scores append;
[!
** Model 15 of 56 * Calculating HKY+G **]
lscores 1/ nst=2 base=est tratio=est rates=gamma shape=est pinv=0
scorefile=model.scores append;
[!
** Model 16 of 56 * Calculating HKY+I+G **]
lscores 1/ nst=2 base=est tratio=est rates=gamma shape=est pinv=est
scorefile=model.scores append;

[!
** Model 17 of 56 * Calculating TrNef **]   [a b c d e f]
lscores 1/ nst=6 base=equal rmat=est rclass=(a b a a e a) rates=equal pinv=0
scorefile=model.scores append;
[!
** Model 18 of 56 * Calculating TrNef+I **]
lscores 1/ nst=6 base=equal rmat=est rates=equal pinv=est
scorefile=model.scores append;
[!
** Model 19 of 56 * Calculating TrNef+G **]
lscores 1/ nst=6 base=equal rmat=est rates=gamma shape=est pinv=0
scorefile=model.scores append;
[!
** Model 20 of 56 * Calculating TrNef+I+G **]
lscores 1/ nst=6 base=equal rmat=est rates=gamma shape=est pinv=est
scorefile=model.scores append;

[!
** Model 21 of 56 * Calculating TrN **]
lscores 1/ nst=6 base=est rmat=est rates=equal pinv=0
scorefile=model.scores append;
[!
** Model 22 of 56 * Calculating TrN+I **]
lscores 1/ nst=6 base=est rmat=est rates=equal pinv=est
scorefile=model.scores append;
[!
** Model 23 of 56 * Calculating TrN+G **]
lscores 1/ nst=6 base=est rmat=est rates=gamma shape=est pinv=0
scorefile=model.scores append;
[!
** Model 24 of 56 * Calculating TrN+I+G **]
lscores 1/ nst=6 base=est rmat=est rates=gamma shape=est pinv=est
scorefile=model.scores append;

[!
** Model 25 of 56 * Calculating K3P **]   [a b c d e f]
lscores 1/ nst=6 base=equal rmat=est rclass=(a b c c b a) rates=equal pinv=0
scorefile=model.scores append;
[!
** Model 26 of 56 * Calculating K3P+I **]
lscores 1/ nst=6 base=equal rmat=est rates=equal pinv=est
scorefile=model.scores append;
[!
** Model 27 of 56 * Calculating K3P+G **]
lscores 1/ nst=6 base=equal rmat=est rates=gamma shape=est pinv=0
scorefile=model.scores append;
[!
** Model 28 of 56 * Calculating K3P+I+G **]
lscores 1/ nst=6 base=equal rmat=est rates=gamma shape=est pinv=est
scorefile=model.scores append;

[!
** Model 29 of 56 * Calculating K3Puf **]
lscores 1/ nst=6 base=est rmat=est rates=equal pinv=0
scorefile=model.scores append;
[!
** Model 30 of 56 * Calculating K3Puf+I **]
lscores 1/ nst=6 base=est rmat=est rates=equal pinv=est
scorefile=model.scores append;
[!
** Model 31 of 56 * Calculating K3Puf+G **]
lscores 1/ nst=6 base=est rmat=est rates=gamma shape=est pinv=0
scorefile=model.scores append;
[!
** Model 32 of 56 * Calculating K3Puf+I+G **]
lscores 1/ nst=6 base=est rmat=est rates=gamma shape=est pinv=est
scorefile=model.scores append;

[!
** Model 33 of 56 * Calculating TIMef **]   [a b c d e f]
lscores 1/ nst=6 base=equal rmat=est rclass=(a b c c e a) rates=equal pinv=0
scorefile=model.scores append;
[!
** Model 34 of 56 * Calculating TIMef+I **]
lscores 1/ nst=6 base=equal rmat=est rates=equal pinv=est
scorefile=model.scores append;
[!
** Model 35 of 56 * Calculating TIMef+G **]

lscores 1/ nst=6 base=equal rmat=est rates=gamma shape=est pinv=0
scorefile=model.scores append;
[!
** Model 36 of 56 * Calculating TIMef+I+G **]
lscores 1/ nst=6 base=equal rmat=est rates=gamma shape=est pinv=est
scorefile=model.scores append;

[!
** Model 37 of 56 * Calculating TIM **]
lscores 1/ nst=6 base=est rmat=est rates=equal pinv=0
scorefile=model.scores append;
[!
** Model 38 of 56 * Calculating TIM+I **]
lscores 1/ nst=6 base=est rmat=est rates=equal pinv=est
scorefile=model.scores append;
[!
** Model 39 of 56 * Calculating TIM+G **]
lscores 1/ nst=6 base=est rmat=est rates=gamma shape=est pinv=0
scorefile=model.scores append;
[!
** Model 40 of 56 * Calculating TIM+I+G **]
lscores 1/ nst=6 base=est rmat=est rates=gamma shape=est pinv=est
scorefile=model.scores append;

[!
** Model 41 of 56 * Calculating TVMef **]   [a b c d e f]
lscores 1/ nst=6 base=equal rmat=est rclass=(a b c d b e) rates=equal pinv=0
scorefile=model.scores append;
[!
** Model 42 of 56 * Calculating TVMef+I **]
lscores 1/ nst=6 base=equal rmat=est rates=equal pinv=est
scorefile=model.scores append;
[!
** Model 43 of 56 * Calculating TVMef+G **]
lscores 1/ nst=6 base=equal rmat=est rates=gamma shape=est pinv=0
scorefile=model.scores append;
[!
** Model 44 of 56 * Calculating TVMef+I+G **]
lscores 1/ nst=6 base=equal rmat=est rates=gamma shape=est pinv=est
scorefile=model.scores append;

[!
** Model 45 of 56 * Calculating TVM **]
lscores 1/ nst=6 base=est rmat=est rates=equal pinv=0

scorefile=model.scores append;
[!
** Model 46 of 56 * Calculating TVM+I **]
lscores 1/ nst=6 base=est rmat=est rates=equal pinv=est
scorefile=model.scores append;
[!
** Model 47 of 56 * Calculating TVM+G **]
lscores 1/ nst=6 base=est rmat=est rates=gamma shape=est pinv=0
scorefile=model.scores append;
[!
** Model 48 of 56 * Calculating TVM+I+G **]
lscores 1/ nst=6 base=est rmat=est rates=gamma shape=est pinv=est
scorefile=model.scores append;

[!
** Model 49 of 56 * Calculating SYM **]   [a b c d e f]
lscores 1/ nst=6 base=equal rmat=est rclass= (a b c d e f) rates=equal pinv=0
scorefile=model.scores append;
[!
** Model 50 of 56 * Calculating SYM+I **]
lscores 1/ nst=6 base=equal rmat=est rates=equal pinv=est
scorefile=model.scores append;
[!
** Model 51 of 56 * Calculating SYM+G **]
lscores 1/ nst=6 base=equal rmat=est rates=gamma shape=est pinv=0
scorefile=model.scores append;
[!
** Model 52 of 56 * Calculating SYM+I+G **]
lscores 1/ nst=6 base=equal rmat=est rates=gamma shape=est pinv=est
scorefile=model.scores append;

[!
** Model 53 of 56 * Calculating GTR **]
lscores 1/ nst=6 base=est rmat=est rates=equal pinv=0
scorefile=model.scores append;
[!
** Model 54 of 56 * Calculating GTR+I **]
lscores 1/ nst=6 base=est rmat=est rates=equal pinv=est
scorefile=model.scores append;
[!
** Model 55 of 56 * Calculating GTR+G **]
lscores 1/ nst=6 base=est rmat=est rates=gamma shape=est pinv=0
scorefile=model.scores append;
[!
** Model 56 of 56 * Calculating GTR+I+G **]
lscores 1/ nst=6 base=est rmat=est rates=gamma shape=est pinv=est

scorefile=model.scores append;

[!
NOTE:The molecular clock hypothesis is not included in this PAUP block for reasons of
computing time. This hypothesis may be tested once we have chosen one of the above
models, simply calculating the log likelihood score of the chosen model with the
molecular clock enforced and comparing it with the log likelihood previously obtained
without enforcing the molecular clock. In this case, the molecular clock is the
null hypothesis. The number of degrees of freedom is the number of OTUs - 2]

LOG STOP;

END;

[
lscores: tree/s
nst: number of substitution types
rmat: rate matrix
base= nucleotide frequencies
rates= rate of evolution for variable sites (same
shape= alpha parameter of the gamma distribution
pinv= proportion of invariable sites
]

[The likelihood ratio test statistic is calculated as twice the difference between
the log likelihood scores of the two models constrasted. When the model representing
the null hypothesis is a special case of the alternate model, this statistic
fits a chi-square distribution with a number of degrees of freedom equal to the
number of parameters that freely vary between the two models. A detailed explanation
of the test of the model of DNA substitution can be found in Huelsenbeck and Crandall
Annu. Rev. Ecol. Syst. 1997. 28: 437-466]

[!*** END OF MODELTEST BLOCK ***]