

Метод апостериорной вероятности,

Прикладная генетика для зоологов, лекция 8

Мюге Н.С.

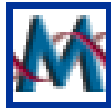
Три основных метода реконструкции филогении:



▶ Парсимония (Parsimony) (PAUP, MEGA, Phylip)



▶ Максимального правдоподобия (maximum likelihood) - (PAUP, Phylip)



▶ Обратной вероятности, байезиан (bayesian) – (MrBayes)



Томас Байес 1702-1761



Беркли утверждает, что Логика и Метафизика откроют математикам глаза и выведут их из всех затруднений... Но если склоки среди профессоров любой науки позорят саму науку, а Логика и Метафизика намного более склочны, нежели математика, то почему же, раз я наполовину слеп, я должен выбирать себе в проводники того, кто вообще ничего не видит?

Преподобный Томас Байес, "В защиту математиков..."



Академик (член королевского общества), не опубликовавший ни одной работы по математике

- ▶ Единственная работа отца Байеса, опубликованная им под своим именем (в 1731 году):

"Благость господня, или попытка доказать, что конечной целью божественного провидения и направления является счастье его созданий".



-
- ▶ **Фундаментальное исследование Байеса в области теории вероятностей было изложено им в**

"Эссе о решении проблем в теории случайных событий".

Эту работу математика лишь после его смерти обнаружил друг Ричард Прайс, который и переслал статью в академию. В 1764 году это "Эссе" было опубликовано в "Трудах Лондонского Королевского общества", откуда и берет начало его мировая слава.



Теорема преподобного Байеса

Формула Байеса позволяет «переставить причину и следствие»: по известному факту события вычислить вероятность того, что оно было вызвано данной причиной.

События, отражающие действие «причин», в данном случае обычно называют **гипотезами**, так как они — *предполагаемые* события, повлекшие данное. Безусловную вероятность справедливости гипотезы называют **априорной** (насколько вероятна причина *вообще*), а условную - с учетом факта произошедшего события — **апостериорной** (насколько вероятна причина *оказалась с учетом данных о событии*).

Формула Байеса:

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

где

$P(A)$ — априорная вероятность гипотезы A (смысл такой терминологии см. ниже);

$P(A|B)$ — вероятность гипотезы A при наступлении события B (апостериорная вероятность);

$P(B|A)$ — вероятность наступления события B при истинности гипотезы A ;

$P(B)$ — вероятность наступления события B .



Формула полной вероятности

Важным следствием формулы Байеса является формула полной вероятности события, зависящего от нескольких несовместных гипотез (и только от них!).

$$P(B) = \sum_{i=1}^N P(A_i)P(B|A_i)$$

вероятность наступления события B , зависящего от ряда гипотез A_i , если известны степени достоверности этих гипотез (например, измерены экспериментально)



<http://mrbayes.csit.fsu.edu/>



MrBayes: Bayesian Inference of Phylogeny

Search results for 'mrbayes: (Нет результатов)'

[Home](#)

[Download](#)

[Manual](#)

[Online Help](#)

[Bug Report](#)

[Authors](#)

[Wiki](#)

[Links](#)

MrBayes is a program for the Bayesian estimation of phylogeny. Bayesian inference of phylogeny is based upon a quantity called the posterior probability distribution of trees, which is the probability of a tree conditioned on the observations. The conditioning is accomplished using Bayes's theorem. The posterior probability distribution of trees is impossible to calculate analytically; instead, MrBayes uses a simulation technique called Markov chain Monte Carlo (or MCMC) to approximate the posterior probabilities of trees.

The program takes as input a character matrix in a NEXUS file format. The output is several files with the parameters that were sampled by the MCMC algorithm. MrBayes can summarize the information in these files for the user. The program features include:



On-line help:

<http://mrbayes.csit.fsu.edu/Help/Help>

The online help was generated by Jeff Bates, Smithsonian Institution. jbates@lab.si.edu.

Commands that are available from the command line or from a MrBayes block include:

[About](#) -- Describes the program

[Acknowledgments](#) -- Shows program acknowledgments

[Charset](#) -- Assigns a group of sites to a set

[Charstat](#) -- Shows status of characters

[Citations](#) -- Appropriate citation of program

[Comparetree](#) -- Compares the trees from two tree files

[Constraint](#) -- Defines a constraint on tree topology

[Ctype](#) -- Assigns ordering for the characters

[Databreaks](#) -- Defines nucleotide pairs (doublets) for stem models

[Delete](#) -- Deletes taxa from the analysis

[Deroot](#) -- Deroots user tree

[Disclaimer](#) -- Describes program disclaimer

[Exclude](#) -- Excludes sites from the analysis

[Execute](#) -- Executes a file

[Help](#) -- Provides detailed description of commands



Execute

This command executes a file called `<file name>`. The correct usage is:

```
execute <file name>
```

For example,

```
execute replicase.nex
```

would execute the file named "[replicase.nex](#)". This file must be in the same directory as the executable.

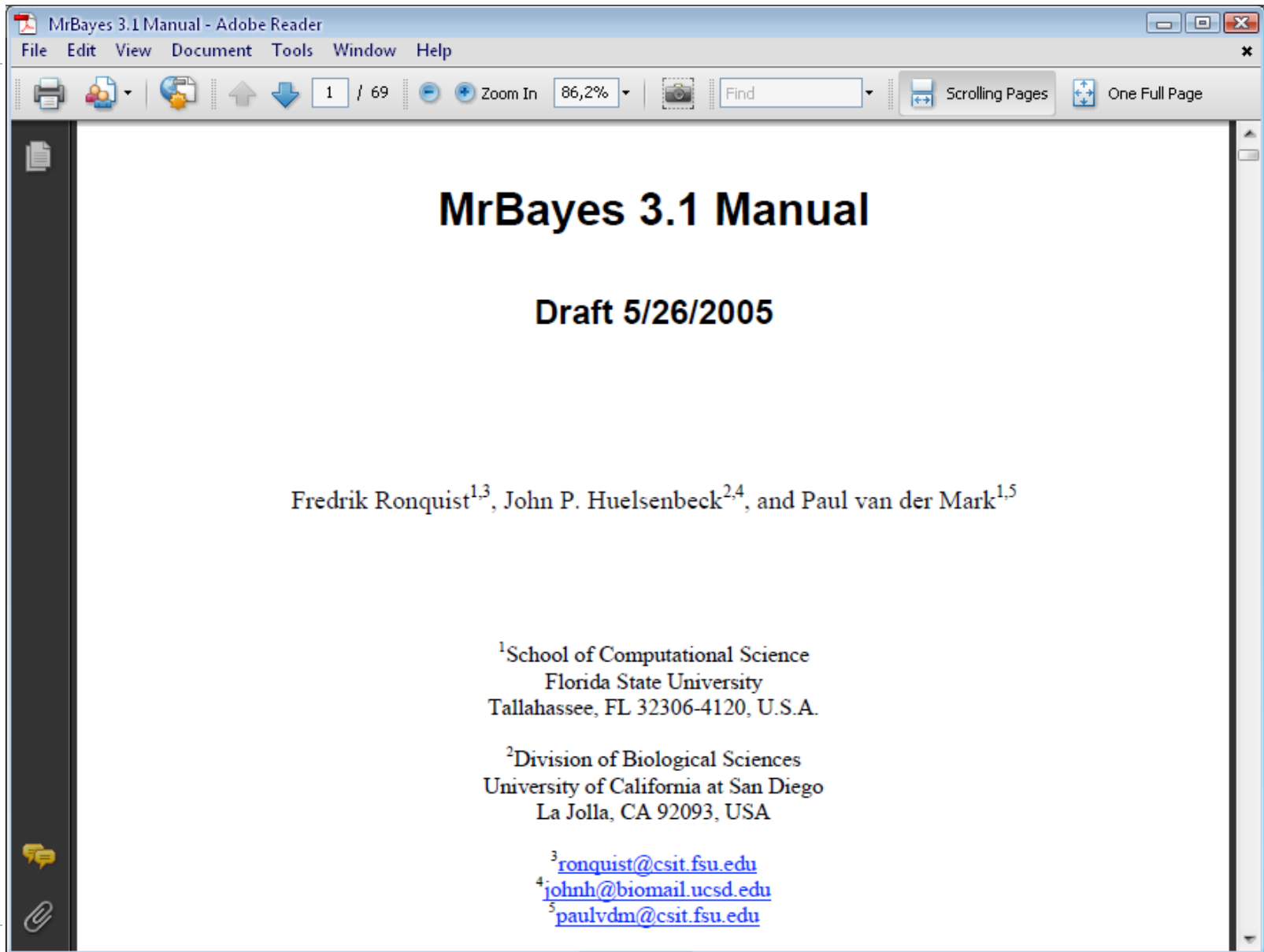
[Return to Help Menu.](#)



```
C:\MrBayes\mrbayes.exe

MrBayes v3.1.2
<Bayesian Analysis of Phylogeny>
by
Fredrik Ronquist and John P. Huelsenbeck
School of Computational Science
Florida State University
ronquist@csit.fsu.edu
Section of Ecology, Behavior and Evolution
Division of Biological Sciences
University of California, San Diego
johnh@biomail.ucsd.edu
Distributed under the GNU General Public License
Type "help" or "help <command>" for information
on the commands that are available.

MrBayes > _
```



MrBayes 3.1 Manual

Draft 5/26/2005

Fredrik Ronquist^{1,3}, John P. Huelsenbeck^{2,4}, and Paul van der Mark^{1,5}

¹School of Computational Science
Florida State University
Tallahassee, FL 32306-4120, U.S.A.

²Division of Biological Sciences
University of California at San Diego
La Jolla, CA 92093, USA

³ronquist@csit.fsu.edu

⁴johnh@biomail.ucsd.edu

⁵paulvdm@csit.fsu.edu

Типичный файл с данными для MrBayes

```
begin data;
dimensions ntax=12 nchar=898;
format datatype=dna interleave=no gap=-;
matrix
Saimiri_sciureus AAGCTTCATAGGAGC ... АСТАТСССТААГСТТ
Tarsius_syrichta AAGCTTCACCGGCGC ... АТТАТГССТААГСТТ
Lemur_catta AAGCTTCACCGGCGC ... АСТАТСТАТТАГСТТ
Macaca_fuscata AAGCTTCACCGGCGC ... ССТААСГСТААГСТТ
M_mulatta AAGCTTCACCGGCGC ... ССТААСАСТААГСТТ
M_fascicularis AAGCTTTACAGGTGC ... ССТААСАСТААГСТТ
M_sylvanus AAGCTTTTCCGGCGC ... ССТААСАТТААГСТТ
Homo_sapiens AAGCTTTTCTGGCGC ... ГСТСТСССТААГСТТ
Pan AAGCTTCTCCGGCGC ... ГСТСТСССТААГСТТ
Gorilla AAGCTTCTCCGGTGC ... АСТСТСССТААГСТТ
Pongo AAGCTTCACCGGCGC ... АСТСТСАСТААГСТТ
Hylobates AAGTTTCATTGGAGC ... АСТСТСССТААГСТТ
;
end;
```



-
- ▶ 1. At the MrBayes > prompt, type **execute primates.nex.**
 - ▶ 2. At the MrBayes > prompt, type **lset nst=6 rates=invgamma.**
 - ▶ 3.1. At the MrBayes > prompt, type **mcmc ngen=10000 samplefreq=10.**
 - ▶ 3.2. If the standard deviation of split frequencies is below 0.01 after 100,000 generations,
 - ▶ stop the run by answering **no when the program asks “Continue the analysis? (yes/no)”**. Otherwise, keep adding generations until the value falls below 0.01.
-



-
- ▶ 4.1. Summarize the parameter values by typing

sump burnin=250

(or whatever value corresponds to 25 % of your samples).
Make sure (PSRF) is reasonably close to 1.0 for all parameters.

- ▶ 4.2. Summarize the trees by typing

sumt burnin=250

(or whatever value corresponds to 25 % of your samples).



Вычисление модели нуклеотидной эволюции

- MrModeltest 2.3 <http://www.abc.se/~nylander/>

Latest update May 22 2008

Johan Nylander



@: Johan.Nylander-at-abc-dot-se

Downloads

Name	Version	What	Readme	Download
MrAIC	1.4.3	Perl script for selecting DNA substitution models using PHYML	README	WIN/MACOSX/MACOS9/UNIX
MrModeltest2	2.3	C program for selecting DNA substitution models using PAUP	README	WIN/MACOSX/MACOS9/UNIX
Mailfit	0.2	Perl script that runs PAUP and MrModeltest2 (or Modeltest) and sends (optionally) results via e-mail	README	WIN/MACOSX/MACOS9/UNIX
Burntrees	0.1.7	Perl script for manipulating MrBayes tree and parameter files. The script comes with an extra tool, catmb.pl , for concatenating files from separate MrBayes runs.	README	WIN/MACOSX/MACOS9/UNIX

Для анализа ML – выбрать и задать модель



MODELTEST: A tool to select the best-fit model of nucleotide substitution

© 1998-2006 David Posada

Current version is 3.7.

MODELTEST is program for the selection the model of nucleotide substitution that best fits the data. The program chooses among 56 models, and implements three different model selection frameworks: hierarchical likelihood ratio tests (hLRTs), Akaike information criterion (AIC), and Bayesian information criterion (BIC). The program also implements the assesment of model uncertainty and tools for model averaging and calculation of parameter importance, using the AIC or the BIC.

Operative systems

Executables are provided for macintosh and windows. Source code and a makefile are provided for compilation in any OS with a C compiler.

Links

These are some useful links related to Modeltest:

- [MTgui](#): a windows and linux interface for modeltest. By Paulo Nuin.
- [MrModeltest](#): a version of Modeltest modified for its use with MrBayes. By Johan Nylander.
- Instructions for [running Modeltest on Windows](#). By Bevan Weir
- [FindModel](#): web server to choose among 28 nucleotide models with the AIC at Los Alamos National Laboratory

← программа

← Пошаговые инструкции

← Modeltest online

Citation

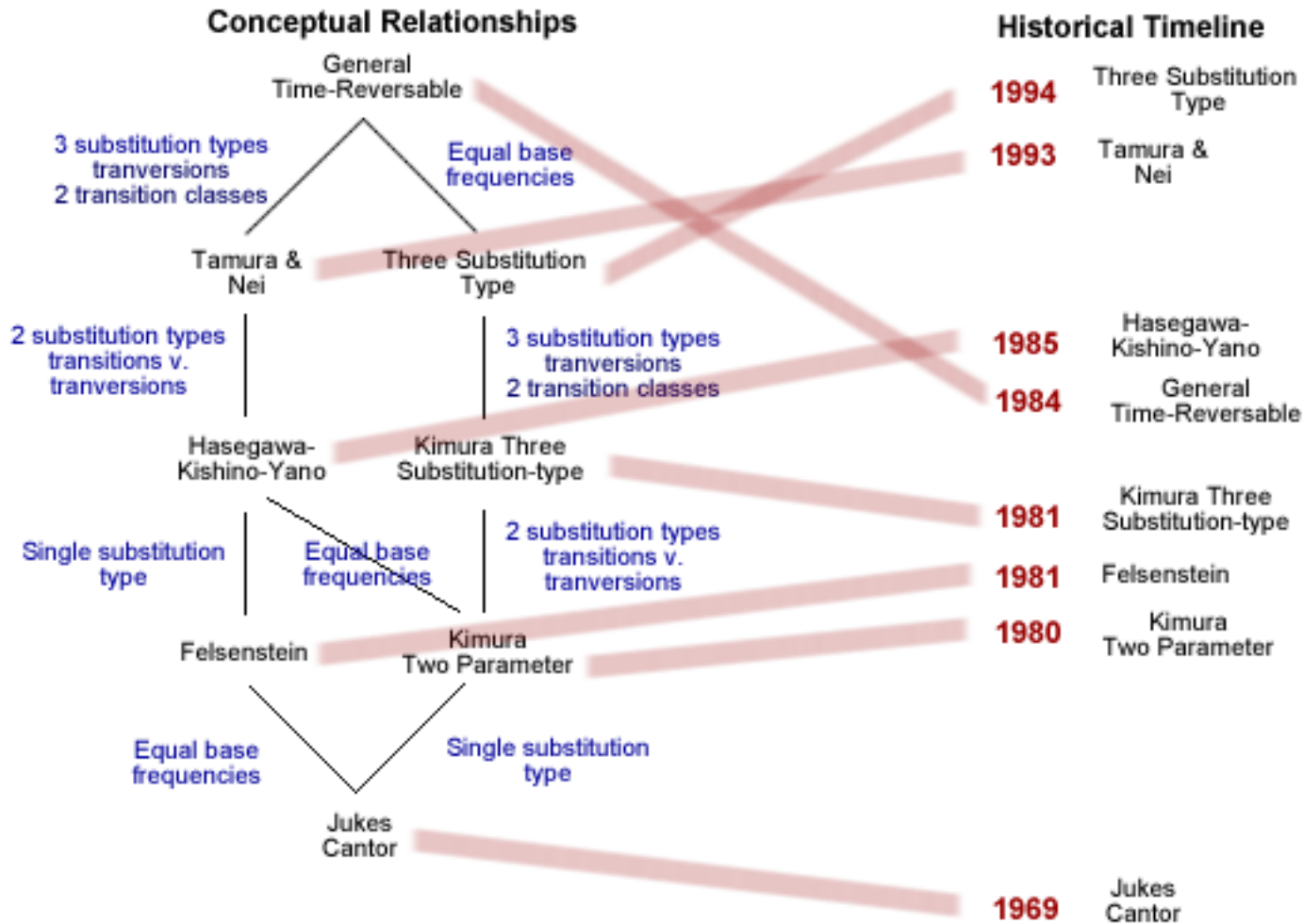
Posada D and Crandall KA 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14 (9): 817-818.

See also: Posada D and Buckley TR. 2004. Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Systematic Biology* 53: 793-801.

Modeltest registration form

▶ <http://darwin.uvigo.es/software/modeltest.html>

Эволюция моделей эволюции ДНК



▶ **JC69 model (Jukes and Cantor, 1969)**

- ▶ JC69 is the simplest substitution model. There are several assumptions. It assumes equal base frequencies (

$$\pi_T = \pi_C = \pi_A = \pi_G = \frac{1}{4}$$

- ▶) and equal mutation rates. The only parameter of this model is therefore μ , the overall substitution rate.



▶ **K80 model (Kimura, 1980)**

- ▶ The K80 model distinguishes between transitions (A <-> G, i.e. from purine to purine, or C <-> T, i.e. from pyrimidine to pyrimidine) and transversions (from purine to pyrimidine or vice versa) (α/β).
- ▶ It also assumes equal base frequencies

$$\pi_T = \pi_C = \pi_A = \pi_G = \frac{1}{4}$$



F81 model (Felsenstein 1981)

Unequal base frequencies ($\pi_T \neq \pi_C \neq \pi_A \neq \pi_G$)

$$\text{Rate matrix } Q = \begin{pmatrix} * & \pi_T & \pi_T & \pi_T \\ \pi_C & * & \pi_C & \pi_C \\ \pi_A & \pi_A & * & \pi_A \\ \pi_G & \pi_G & \pi_G & * \end{pmatrix}$$

HKY85 model (Hasegawa, Kishino and Yano 1985)

The HKY85 model distinguishes between [transitions](#) and [transversions](#) (α/β).

It allows unequal base frequencies ($\pi_T \neq \pi_C \neq \pi_A \neq \pi_G$).

$$\text{Rate matrix } Q = \begin{pmatrix} * & \kappa\pi_T & \pi_T & \pi_T \\ \kappa\pi_C & * & \pi_C & \pi_C \\ \pi_A & \pi_A & * & \kappa\pi_A \\ \pi_G & \pi_G & \kappa\pi_G & * \end{pmatrix}$$



GTR: Generalised time reversible

$$Q = \begin{pmatrix} -\left(\frac{\pi_1 x_1}{\pi_2} + \frac{\pi_1 x_2}{\pi_3} + \frac{\pi_1 x_3}{\pi_4}\right) & \frac{\pi_1 x_1}{\pi_2} & \frac{\pi_1 x_2}{\pi_3} & \frac{\pi_1 x_3}{\pi_4} \\ x_1 & -\left(x_1 + \frac{\pi_2 x_4}{\pi_3} + \frac{\pi_2 x_5}{\pi_4}\right) & \frac{\pi_2 x_4}{\pi_3} & \frac{\pi_2 x_5}{\pi_4} \\ x_2 & x_4 & -\left(x_2 + x_4 + \frac{\pi_3 x_6}{\pi_4}\right) & \frac{\pi_3 x_6}{\pi_4} \\ x_3 & x_5 & x_6 & -(x_3 + x_5 + x_6) \end{pmatrix}$$



Modeltest tutorial

Modeltest guide
MrBayes guide
Using DOS
Downloads
Related Links
Modeltest website
Maximum likelihood
PAUP*
ClustalX

Ads by Google

[Macrogen DNA sequencing](#)
\$5/reaction, 900bp, free shipping Worldwide, Shotgun seq.
www.macrogen.com/english

[Free Phylogeny Software](#)
Easy-to-use Bioinformatics Tools Advanced and Interactive Graphics
www.clcbio.com

[Free Online Book](#)
Future Human Evolution Eugenics in the 21st Century
www.whatwemaybe.org

[DNA-Based Diagnostic Mark](#)
The comprehensive market analysis report

The step by step guide to Modeltest

This guide focuses on using Modeltest to find the best parameters to construct a maximum likelihood phylogenetic tree. This is not a substitute for a bioinformatics education, make sure you know what you are doing. [Updated 23rd July 2006]

MODELTEST is a computer program by David Posada and Keith A Crondall. One use of this program is to calculate the best model for DNA evolution using Maximum Likelihood (ML). For more information read their [publication](#) (pdf) or visit the [Modeltest website](#). When I started to use this program I found there was little detailed information about to help newbies use it, so I wrote this tutorial to remind me of the steps, and to help some graduate students in my lab.

The latest version of Modeltest is 3.7 released on July 29, 2005

To use the Modeltest program you need the application itself (available from the authors website), and the phylogenetics package **PAUP*** (this is not free unfortunately, hopefully your university or company has a license). I will be referring to the windows command line version in this document, but these files will work on all versions. Finally you will need a fast computer, I use a 2.4 GHz Pentium 4 laptop, on this it takes about an hour to work through 50 taxa of 1300 bp. You also need two files: [modelblockPAUPb10.txt](#) and [ML-search.txt](#) which you can download from my website. [modelblockPAUPb10.txt](#) was written by David Posada and is a copy of the file found in the paupblock folder that is included in the download. [ML-search](#) was written by me, modify this for your own use.

I also assume that at this point you have a sequence alignment in the NEXUS format (this is critical). The easiest way to do this is use the "save as" command in [ClustalX](#) Open up this file in a text editor and replace the "." (gap) symbol with the missing data symbol "?" where appropriate. It is getting out of the scope of this document to explain this, but typically this is at the ends of the alignment when you don't have full sequence for all taxa. One handy hint don't use the "-" symbol in taxa names, use "_" instead.

I use my own file-naming scheme to keep track of things. I suggest replacing the word "test" below with the name of your gene, i.e.: "p53.ML.search.nex"

Finally I highly recommend you read the documentation file: [modeltest3.6.pdf](#) found in the doc folder of the download

Step by step MODELTEST for the windows version


1. Start with a nexus file of your alignment in this document it will be called "test.aln.nex"
2. Add the PAUP*-block to the end of this file. The sample block is called "modelblockPAUPb10.txt" This has changed from previous versions, by adding the command "default lscores longfmt=yes;" this is due to a bug in PAUP* 4.10b. If you use the PAUP-block that comes with earlier versions of Modeltest it won't work!
3. Save this file as "test.model.nex" (keeping the original). Drag this new file on to the paupstar.exe executable. Or if you have a version with a graphical user interface (GUI), execute the file using the menu system.
4. All going well a DOS-like window will open up and PAUP* will begin to test your data against 56 different models of DNA evolution.

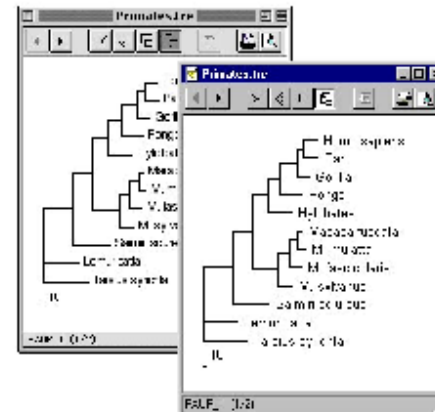
Программа для работы с деревьями- TreeView (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>)

Taxonomy and Systematics at Glasgow

TreeView

Tree drawing software for Apple Macintosh and Windows

 (and now [Linux and Unix](#))




TreeView is a simple program for displaying phylogenies on Apple Macintosh and Windows PCs. It has the following features:

Phylogeny Programs - Windows Internet Explorer

http://evolution.genetics.washington.edu/phylip/software.html

Phylogenetic programs - ... Phylogeny Programs



Phylogeny Programs

[Changes](#)
 [Waiting list](#)
 [Other lists](#)
 [Old programs](#)
 [Not listed](#)
 [???](#)

Here are 362 phylogeny packages and 50 [free servers](#), all that I know about. It is an attempt to be completely comprehensive. I have not made any attempt to exclude programs that do not meet some standard of quality or importance. Updates to these pages are made roughly weekly. [Here](#) is a "waiting list" of new programs waiting to have their full entries constructed. Many of the programs in these pages are available on the web, and some of the older ones are also available from [ftp server machines](#).

The programs listed below include both free and non-free ones; in some cases I do not know whether a program is free. I have listed as free those that I knew were free; for the others you have to ask their distributor. Usually when I say that a program is downloadable from a web site, this means that it is available free.

Email addresses in these pages have had the @ symbol replaced by (at) and also surrounded by invisible confusing tags and blank characters in hopes of foiling spambots that harvest email addresses.

If you discover any inaccuracies, or feel that I have left any important programs or facts out, or if links do not work properly, please e-mail me at: (joe (at) gs.washington.edu). You can also use the submission form [here](#) to submit new entries.

Owing to past NSF support of these pages, I am required to note that any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation (NSF supported these pages from 1995-2003).

List of packages arranged ...

... [by methods available](#)

... [by computer systems on which they work](#)

... [cross-referenced by method and by computer system.](#)

... [by ones which analyze particular kinds of data.](#)

<http://evolution.genetics.washington.edu/phylip/software.html>