

**Un tutorial sobre el uso de Modeltest3.7 para la selección de modelos usando LRTs, AICs y BICs**

- Conviene que leas este tutorial después de haber estudiado el tutorial de manejo de PAUP\* desde la línea de comandos y el tema de teoría sobre el uso del criterio de máxima verosimilitud en filogenética.
- **Modeltest** es una aplicación escrita por David Posada que **seleccionan el modelo mejor ajustado de la familia GTR** para un alineamiento de DNA, usando dos tipos de estrategias: **tests pareados de razones de verosimilitud (LRTs, hLRTs = hierarchical LRTs)** y **criterios de información (AIC y BIC)**. Para ello **necesita que PAUP\* calcule los -lnL scores** de un subconjunto (56) de todos los posibles modelos de la familia GTR (203). Estos scores de -lnL se calculan corriendo un "batch file" de comandos PAUP\*. Lo primero que se estima es un árbol (rápido) NJ-JC69. Se usa la topología resultante para evaluar los distintos modelos y obtener estimas de ML de los parámetros correspond.

**Un tutorial sobre el uso de Modeltest3.7 para la selección de modelos usando LRTs, AICs y BICs**

**Modeltest 3.7** (June 2005)



- **modelblockpaup**

```
#NEXUS
[! ***** MODEL FIT BLOCK -- MODELTEST 3.7 *****]
[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;
```

**Un tutorial sobre el uso de Modeltest3.7 para la selección de modelos usando LRTs, AICs y BICs**

- **modelblockpaup** - continuación
- ```
[!***** BEGIN TESTING 56 MODELS OF EVOLUTION *****]
BEGIN PAUP;
Default lscores longfmt=yes;
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 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;
LOG STOP;END;
```

**Un tutorial sobre el uso de Modeltest3.7 para la selección de modelos usando LRTs, AICs y BICs**

- **¿ Cómo ejecuto el archivo modelblockpaup ?**
- Existen básicamente dos opciones:
- I.- Interactivamente:**
    - paup myfile.nex ; [hacer ajustes deseados]  
exe '/path/to/modelblockpaup'
    - copiar el modelblock al final de nuestro archivo myfile.nex y ejecutar luego paup myfile.nex ;  
En UNIX/Linux se puede hacer fácilmente con el comando:  
cat myfile.nex modelblockpaup > myfile\_modelblock.nex
  - II.- NO- interactivamente (desde una terminal UNIX/Linux ):**  
cat modelblockpaup | paup -n myfile.nex

**Un tutorial sobre el uso de Modeltest3.7 para la selección de modelos usando LRTs, AICs y BICs**

• ¿ Qué hago después de que PAUP\* ha ejecutado el modelblock ?

- PAUP\* hará generados archivos: model.scores y modelfit.log. Modeltest trabajará sobre los valores de -lnL guardados en orden en el archivo model.scores.

Existen de nuevo dos opciones para correr modeltest:

**I.- Interactivamente:**

ejecuta el programa modeltest y verás la siguiente consola similar a esta:



**Un tutorial sobre el uso de Modeltest3.7 para la selección de modelos usando LRTs, AICs y BICs**

• ¿ Qué hago después de que PAUP\* ha ejecutado el modelblock ?

**II.- NO-interactivamente:**

ejecuta el programa modeltest desde la línea de comandos así:

**modeltest < model.scores > myfile\_modeltest.out**

• ¿ Qué hago después de correr model.scores con modeltest ?

- interpretación de la salida de modeltest: **1. hLRTs**

\* HIERARCHICAL LIKELIHOOD RATIO TESTS (hLRTs) \*

Confidence level = 0.01

**Equal base frequencies**

Null model = JC -lnLO = 6424.2026  
 Alternative model = F81 -lnL1 = 6284.9956  
 2(lnL1-lnLO) = 278.4141 df = 3  
 P-value = <0.000001

**Ti=Tv**

Null model = F81 -lnLO = 6284.9956  
 Alternative model = HKY -lnL1 = 5981.7202  
 2(lnL1-lnLO) = 606.5508 df = 1  
 P-value = <0.000001

**Equal Ti rates**

Null model = HKY -lnLO = 5981.7202  
 Alternative model = TrN -lnL1 = 5978.8550  
 2(lnL1-lnLO) = 5.7305 df = 1  
 P-value = 0.016673

**Equal Tv rates**

Null model = HKY -lnLO = 5981.7202  
 Alternative model = K81uf -lnL1 = 5973.2393  
 2(lnL1-lnLO) = 16.9619 df = 1  
 P-value = 0.000038

(continúa en la siguiente página)

• ¿ Qué hago después de correr model.scores con modeltest ?

- interpretación de la salida de modeltest: **1. hLRTs** (Continuación)

**Only two Tv rates**

Null model = K81uf -lnLO = 5973.2393  
 Alternative model = TVM -lnL1 = 5938.5615  
 2(lnL1-lnLO) = 69.3555 df = 2  
 P-value = <0.000001

**Equal rates among sites**

Null model = TVM -lnLO = 5938.5615  
 Alternative model = TVM+G -lnL1 = 5709.6323  
 2(lnL1-lnLO) = 457.8584 df = 1  
 Using mixed chi-square distribution  
 P-value = <0.000001

**No Invariable sites**

Null model = TVM+G -lnLO = 5709.6323  
 Alternative model = TVM+I+G -lnL1 = 5709.6323  
 2(lnL1-lnLO) = 0.0000 df = 1  
 Using mixed chi-square distribution

**P-value = >0.999999 es decir, no rechaza la H<sub>0</sub> !!! El modelo seleccionado es TVM+G**

• ¿ Qué hago después de correr model.scores con modeltest ?

- interpretación de la salida de modeltest: **1. hLRTs** (Continuación)

**Model selected: TVM+G**

-lnL = 5709.6323

K = 8

Base frequencies:

freqA = 0.3581

freqC = 0.3186

freqG = 0.0846

freqT = 0.2387

Substitution model:

Rate matrix

R(a) [A-C] = 3.9989

R(b) [A-G] = 40.5788

R(c) [A-T] = 3.4119

R(d) [C-G] = 2.3909

R(e) [C-T] = 40.5788

R(f) [G-T] = 1.0000

Among-site rate variation

Proportion of invariable sites = 0

Variable sites (G)

Gamma distribution shape parameter = 0.3752

• ¿ Qué hago después de correr model.scores con modeltest ?

- interpretación de la salida de modeltest: **1. hLRTs** (Continuación)

--

PAUP\* Commands Block: If you want to implement the previous estimates as likelihood settings in PAUP\*, attach the next block of commands after the data in your PAUP file:

```
[!
Likelihood settings from best-fit model (TVM+G) selected by hLRT
in Modeltest 3.7 on Sat May 20 17:12:56 2006
]
```

```
BEGIN PAUP;
Lset Base=(0.3581 0.3186 0.0846) Nst=6 Rmat=(3.9989 40.5788
3.4119 2.3909 40.5788) Rates=gamma Shape=0.3752 Pinvar=0;
END;
```

--

• ¿ Qué hago después de correr model.scores con modeltest ?

- interpretación de la salida de modeltest: **2. AIC = -2 ln L + 2 K; Akaike 1974**  
(cantidad de información perdida cuando la realidad es aproximada por un modelo)

```
-----
* AKAIKE INFORMATION CRITERION (AIC)
*
-----
```

**Model selected: TrN+G**

-lnL = 5710.5513

K = 6

AIC = 11433.1025

Base Frequencies:

freqA = 0.3581

freqC = 0.3252

freqG = 0.0765

freqT = 0.2402

Substitution model:

Rate matrix

R(a) [A-C] = 1.0000

R(b) [A-G] = 16.0043

R(c) [A-T] = 1.0000

R(d) [C-G] = 1.0000

R(e) [C-T] = 11.6796

R(f) [G-T] = 1.0000

Among-site rate variation

Proportion of invariable sites = 0

Variable sites (G)

Gamma distribution shape parameter = 0.3566

• ¿ Qué hago después de correr model.scores con modeltest ?

- interpretación de la salida de modeltest: **2. AIC** (continuación)

PAUP\* Commands Block: If you want to implement the previous estimates as likelihood settings in PAUP\*, attach the next block of commands after the data in your PAUP file:

```
[!
Likelihood settings from best-fit model (TrN+G) selected by
AIC in Modeltest 3.7 on Sat May 20 17:12:56 2006
]
```

```
BEGIN PAUP;
Lset Base=(0.3581 0.3252 0.0765) Nst=6 Rmat=(1.0000 16.0043
1.0000 1.0000 11.6796) Rates=gamma Shape=0.3566 Pinvar=0;
END;
```

• ¿ Qué hago después de correr model.scores con modeltest ?

- interpretación de la salida de modeltest: **2. AIC** (continuación)

\* MODEL SELECTION UNCERTAINTY : Akaike Weights

| Model     | -lnL      | K  | AIC        | delta  | weight | cumWeight     |
|-----------|-----------|----|------------|--------|--------|---------------|
| TrN+G     | 5710.5513 | 6  | 11433.1025 | 0.0000 | 0.2463 | 0.2463        |
| HKY+G     | 5711.9385 | 5  | 11433.8770 | 0.7744 | 0.1672 | 0.4135        |
| TIM+G     | 5710.4355 | 7  | 11434.8711 | 1.7686 | 0.1017 | 0.5152        |
| TrN+I+G   | 5710.5513 | 7  | 11435.1025 | 2.0000 | 0.0906 | 0.6058        |
| TVM+G     | 5709.6323 | 8  | 11435.2646 | 2.1621 | 0.0835 | 0.6894        |
| K81uf+G   | 5711.8125 | 6  | 11435.6250 | 2.5225 | 0.0698 | 0.7591        |
| GTR+G     | 5708.9224 | 9  | 11435.8447 | 2.7422 | 0.0625 | 0.8217        |
| HKY+I+G   | 5711.9385 | 6  | 11435.8770 | 2.7744 | 0.0615 | 0.8832        |
| TIM+I+G   | 5710.4355 | 8  | 11436.8711 | 3.7686 | 0.0374 | 0.9206        |
| TVM+I+G   | 5709.6323 | 9  | 11437.2646 | 4.1621 | 0.0307 | <b>0.9513</b> |
| K81uf+I+G | 5711.8125 | 7  | 11437.6250 | 4.5225 | 0.0257 | 0.9770        |
| GTR+I+G   | 5708.9224 | 10 | 11437.8447 | 4.7422 | 0.0230 | 1.0000        |

Intervalo de credibilidad del 95 %

• ¿ Qué hago después de correr model.scores con modeltest ?

- interpretación de la salida de modeltest: **2. AIC** (continuación)

\* MODEL AVERAGING AND PARAMETER IMPORTANCE (using Akaike Weights) including all 56 models (índices normalizados y relativos de Akaike)

| Parameter  | Importance | Model-averaged estimates |
|------------|------------|--------------------------|
| fA         | 1.0000     | 0.3596                   |
| fC         | 1.0000     | 0.3223                   |
| fG         | 1.0000     | 0.0794                   |
| fT         | 1.0000     | 0.2387                   |
| TiTv       | 0.2287     | 5.4113                   |
| rAC        | 0.1998     | 3.7999                   |
| rAG        | 0.5615     | 19.9668                  |
| rAT        | 0.1998     | 3.2371                   |
| rCG        | 0.1998     | 2.3657                   |
| rCT        | 0.5615     | 14.9960                  |
| pinv(I)    | 0.0000     | 0.3717                   |
| alpha(G)   | 0.7311     | 0.3621                   |
| pinv(I G)  | 0.2689     | 0.0000                   |
| alpha(I G) | 0.2689     | 0.3621                   |

• Interpretación de la importancia de parámetros

- los params. de frec. son un componente esencial del modelo
- Ti/Tv también es significativa
- El pto. 2 se ratifica en la import. de rAG y rCT respecto a tasas de Tv
- El parámetro alpha (uso de distrib. gamma) es mucho más imp. que asumir sólo pinv.

Values have been rounded.  
 (I): averaged using only +I models.  
 (G): averaged using only +G models.  
 (IG): averaged using only +I+G models.

• Modeltest3.7: opciones adicionales en la línea de comandos

• Existen varias opciones que pueden ser especificadas en la línea de comandos o desde la consola de ModelTest. Los más importantes son:

- a: nivel alfa de significancia (por ej. -a0.05) (por defecto a=0.01);
- n: tamaño de muestra (no. de caracteres; por ej. -n745). Fuerza a usar AICc (por defecto usa AIC);
- t: no. de taxa. Fuerza la inclusión de no. de ramas como parámetros (p. ej. -t8) (por defecto no se cuentan);
- w: intervalo de confianza para promediado (p. ej. -w0.95) (opc. por defecto w=1);
- l: activa el modo de calculadora de LRTs (uso: -l);
- b: activa el uso de BIC en vez de AIC para todos los cálculos (opc. por defecto=AIC);
- ?: ayuda;

Modelos de base evaluados por Modeltest3.7

Table 1. Model names. Some models have no reference (TNeI, K81uf, TIMeI, TIM, TVMeI, TVM); they are just some variations of some existing models, and they were not developed, only named, by D. Posada.

| Model | Name                                                          |
|-------|---------------------------------------------------------------|
| JC    | Jukes and Cantor (Jukes and Cantor, 1969)                     |
| F81   | Felsenstein 81 (Felsenstein, 1981)                            |
| K80   | Kimura 80 (=K2P) (Kimura, 1980)                               |
| HKY   | Hasegawa, Kishino, Yano 85 (Hasegawa, Kishino and Yano, 1985) |
| TNeI  | Tamura-Nei equal frequencies                                  |
| TN    | Tamura-Nei (Tamura and Nei, 1993)                             |
| K81   | Two transversion-parameters model 1 (=K81-K2P) (Kimura, 1981) |
| K81uf | Two transversion-parameters model 1 unequal frequencies       |
| TIMeI | Transitional model equal frequencies                          |
| TIM   | Transitional model                                            |
| TVMeI | Transversional model equal frequencies                        |
| TVM   | Transversional model                                          |
| SYM   | Symmetrical model (Zharkikh, 1994)                            |
| GTR   | General time reversible (=REV) (Lanave, 1986)                 |

Modelos de base evaluados por Modeltest3.7

Table 2. Model parameters. The substitution codes are just two ways of indicating the substitution scheme. Any of these models can ignore rate variation or include invariable sites (+I), rate variation among sites (+G), or both (+I+G).

| Model | Free       | Base        | Substitution rates | Substitution code |        |
|-------|------------|-------------|--------------------|-------------------|--------|
|       | parameters | frequencies |                    | code 1            | code 2 |
| JC    | 0          | equal       | a=b=c=d=e=f        | 000000            | aaaaaa |
| F81   | 3          | unequal     | a=b=c=d=e=f        | 000000            | aaaaaa |
| K80   | 1          | equal       | a=c=d=f, b=e       | 010010            | abaaba |
| HKY   | 4          | unequal     | a=c=d=f, b=e       | 010010            | abaaba |
| TMef  | 2          | equal       | a=c=d=f, b, e      | 010020            | abaaca |
| TN    | 5          | unequal     | a=c=d=f, b, e      | 010020            | abaaca |
| K81   | 2          | equal       | a=f, c=d, b=e      | 012210            | abccba |
| K81uf | 5          | unequal     | a=f, c=d, b=e      | 012210            | abccba |
| TMef  | 3          | equal       | a=f, c=d, b, e     | 012230            | abccda |
| TM    | 6          | unequal     | a=f, c=d, b, e     | 012230            | abccda |
| TVMeI | 4          | equal       | a, c, d, f, b=e    | 012314            | abcdbb |
| TVM   | 7          | unequal     | a, c, d, f, b=e    | 012314            | abcdbb |
| SYM   | 5          | equal       | a, c, d, f, b, e   | 012345            | abcdef |
| GTR   | 8          | unequal     | a, c, d, f, b, e   | 012345            | abcdef |