Phylomethods Fall 2006 Dating – r8s, multidistribute

Jun G. Inoue

Software of Dating

Molecular Clock

- Relaxed Molecular Clock
 - ▶r8s
 - multidistribute

r8s



Michael J. Sanderson UC Davis

Estimation of rates and times based on ML, smoothing, or other methods.

Methods of r8s – 1

- LF (Langley-Fitch) "local molecular clock"
 - The user must indicate for every branch in the tree which of the parameters are associated with it based on the biological feature (such as life history, generation time).
- NPRS (Nonparametric Rate Smoothing)
 - Relaxes the assumption of a molecular clock by using a least squares smoothing of local estimates of substitution rates. It relies on the observation that closely related lineages tend to have similar rates.

Methods of r8s - 2

- PL (Penalized likelihood)
 - Combines a parametric model having a different substitution rate on every branch with a nonparametric roughness penalty.
 - Finds the combination of (*R*, *T*) of *R* and *T* that obeys any constraints on node times and that maximizes,

$$log(p(X|R, T)) - \lambda \Phi(R)$$
.

PL method

Maximize the following:

$$log(p(X|R, T)) - \lambda \Phi(R)$$

log likelihood

penalty function

- X: data
- R: rate
- T: time

- $\varPhi(R)$: penalty function
- λ : it determines the contribution of this penalty function

PL method

 $\begin{array}{l} log(p(X|R,\,T)) - \lambda \varPhi(R) \\ \text{log likelihood} & \text{penalty function} \end{array} \\ \end{array}$

 $\varPhi(R)$: penalty function

Penalty function is the sum of two parts (squared):

- Variance among rates for those branches that are directly attached to the root.
- Those branches that are not directly attached to the root.
- λ : contribution of penalty function
 - Cross-validation strategy selects the value of λ .
 - $\lambda \rightarrow$ small, extensive rate variation over time.
 - $\lambda \rightarrow$ large, clock like pattern of rates.

PL method

 $\begin{array}{ll} log(p(X|R,\,T)) & -\,\lambda\,\varPhi(R) \\ \hline \mbox{log likelihood} & \mbox{penalty function} \\ \hline \mbox{Note: r8s does not use sequences} \end{array}$

- Not fully evaluate when finding \tilde{R} and \tilde{T} .
 - Yields an integer-valued estimate of the number of nucleotide substitutions that have affected the sequence on each branch of the tree.
 - Integer-valued estimate for a brach as if it were a directly observed realization from a **Poisson distribution**. The mean value of the Poisson distribution is determined from R and T by the product of the average rate and time duration of the branch.

r8s – infile



Tree with branch length from PAUP*

Total numbers of substitutions

Naming nodes

Fixing times for some nodes

r8s – infile



r8s – terminal



r8s – outfile

"/250M	B/manus	crips/Seminar o	f DivTimEs	t/Coe29b.	out" (InDesi	gn Classic書類	- 標準) 2006/1	0/31 5:28:19 PM			
8 ←	₀ (⇔	9 B 8	8	MacRom	ian 🔻	LF (UNIX) 🔻	4				
169 170	١	lode	(Fix	Constro [Mod]	aints N	Min	Rate Max Ad	es ge Estimat	ed Lo	ocal	
171	2222										
172	[*]	node1	*	[0]	-	(7 ,7);	450.00			18 - X2001	
173		(4)	12	[0]			31.01	3.7446e-04	3.5337	'e-04	
174		Lach	*	[0]			0.00	3.7532e-04	4.4341	Le-04	
175		Lame	*	[0]	02220	- 22	0.00	3.7333e-04	2.8415	6e-04	
176	[*]	node3		[0]		450.0	0 421.57	3.9001e-04	5.1434	le-04	1
177		(8)		[0]			104.67	4.2073e-04	6.2151	Le-04	
178		(9)		[0]			66.66	4.2395e-04	5.0952	2e-04	
179		Poop	*	[0]	9 <u>298</u> 9	9 <u>199</u> 0	0.00	4.2626e-04	5.2296	6e-04	
180		Pose	*	[0]	(===)	(==)	0.00	4.2368e-04	4.1256	5e-04	
181		Caca	٠	[0]	- -	- -	0.00	4.2507e-04	5.4028	3e-04	
182		(13)		[0]	2 <u>424</u> 0		365.23	3.6178e-04	2.8533	8e-04	
183	E*7	node7		[0]	155.00) (339.32	3.3150e-04	1.8683	8e-04	
INT 10/	- F*-	Dodo0		FAT	141 00		11 100	2 11270 04	7 1000	01	

Node Ages Rates Constraints

r8s – outfile



ASCII tree

Tree in parentheses notation including branch length information

multidistribute



Jeffrey L. Thorne North Carolina State University

Estimation of rates and times using Bayesian method.

The multidistribute consists of two programs:

- **estbranches** estimates branch lengths and their variancecovariance matrix.
- **multidivtime** estimates rates and times using Bayesian method.



Posterior distribution

The Metropolis-Hastings algorithm is used to approximate $p(T, R, \nu | X, C)$.

$$p(T, R, \nu \,|\, X, C) = \frac{p(X, T, R, \nu \,|\, C)}{p(X \,|\, C)}$$

- T: Internal node times (T₀, T₁,....,T_k)
- R: Rates (Ro, R1,....,Rk)
- ν : Parameter of rate autocorrelation (see next slide)
- C: Constraints on node times

X: Data

Thorne and Kishino (2005)

Rate Autocorrelation $p(T, R, \nu | X, C) = \frac{p(X, T, R, \nu | C)}{p(X | C)}$

- $\boldsymbol{\mathcal{V}}$: Rate autocorrelation parameter
- v determines the prior distribution for the rates on different branches given the internal node times.
- High \rightarrow little rate autocorrelation Low \rightarrow strong rate autocorrelation

Rate Autocorrelation

• A rate *Ro* is sampled from the prior distribution (gamma distribution), p(Ro).

 $\log(R_2)$ and $\log(R_3)$

- Normal distribution
 - a mean equal to log (Ro) a variance equal to $\nu x \frac{t_1 + t_2}{2}$
- So, the prior distribution of lognormal change of rate of R_5 is sampled depends on, but will differ from, the prior distribution from which R_0 is sampled.
- In this way, the joint probability density of the rates at all nodes on the tree is defined for a given value of the rate at the root node.



Multilocus Data Analysis



- Each gene is modeled as experiencing its own independent rate trajectory, i.e., the distributions of evolutionary rates among genes are priori uncorrelated.
- Rates at the root node for individual genes (Ra0 and Rb0) are independent realizations from a gamma distribution.
- Two approaches for autocorrelation rate parameter v.

Va = Vb or $Va \neq Vb$.

Posterior Distribution



Thorne and Kishino (2005)

Posterior Distribution

Posterior distribution $p(T, R, \nu | X, C) =$

Likelihood Prior distribution $\frac{p(X|B)p(R|T,\nu)p(T|C)p(\nu)}{p(X|C)}$

Marginal probability of the data

Numerator

p(X | B): Likelihood approximated with a multivariate normal distribution centered about the ML estimates of branch lengths B.

 $p(R \mid T, v)$: The prior distribution of rate evolution. This probability is determined once the times T and the constant v is determined.

p(T | C): The density p(T | C) is identical to the density p(T) up to a proportionality constant to. The p(T) is selected on the basis of simplicity of the Yule process.

p(**v**): The prior distribution for the rate change parameter.

Thorne and Kishino (2005)

Metropolis-Hastings Algorithm

$$p(T, R, \nu | X, C) = \frac{p(X|B)p(R|T, \nu)p(T|C)p(\nu)}{p(X|C)}$$
Denominator
$$p(X|C) = \sum \int \int \int \int \int Posterior \ Probability \ Density \ dT dR d\nu$$

This is not pretty.

Metropolis-Hastings Algorithm

$$p(T, R, \nu | X, C) = \frac{p(X|B)p(R|T, \nu)p(T|C)p(\nu)}{p(X|C)}$$

$$r = min\left(1, \frac{p(T', R', \nu' \mid X)}{p(T, R, \nu \mid X)} \frac{J(T, R, \nu \mid T', R', \nu')}{J(T', R', \nu' \mid T, R, \nu)}\right)$$

Hastings ratio
$$\frac{p(T', R', \nu' \mid X)}{p(T, R, \nu \mid X)} = \frac{p(X \mid B')p(R' \mid T', \nu')p(T' \mid C)p(\nu')}{p(X \mid B)p(R \mid T, \nu)p(T \mid C)p(\nu)}$$

- Metropolis-Hastings Algorithm is adopted to obtain an approximately random sample from p(T, R, v | X).
- Repeating many cycles of this procedure of random proposals followed by acceptance or rejection produces a Markov chain with a stationary distribution that is the desired posterior distribution p(T, R, v, | X).

Metropolis-Hastings Algorithm

$$r = min\left(1, \frac{p(T', R', \nu' \mid X)}{p(T, R, \nu \mid X)} \frac{J(T, R, \nu \mid T', R', \nu')}{J(T', R', \nu' \mid T, R, \nu)}\right)$$

Hastings ratio
$$\frac{p(T', R', \nu' \mid X)}{p(T, R, \nu \mid X)} = \frac{p(X \mid B')p(R' \mid T', \nu')p(T' \mid C)p(\nu')}{p(X \mid B)p(R \mid T, \nu)p(T \mid C)p(\nu)}$$

Proposal step for v

$$r = \min\left(1, \frac{p(R' \mid T', \nu')p(\nu')\nu'}{p(R \mid T, \nu)p(\nu)\nu}\right)$$

 $\begin{aligned} & \text{Proposal step for } \mathcal{T} \\ & r = \min\left(1, \, \frac{p(R' \,|\, T', \nu') p(T')(T_y - T_0)}{p(R \,|\, T, \nu) p(T)(T'_y - T'_0)}\right) \end{aligned}$

Proposal step for R

$$r = min\left(1, \frac{p(X' \mid B')p(R' \mid T', \nu')R'_i}{p(X \mid B)p(R \mid T, \nu)R_i}\right)$$

"Mixing" Step
$$r = min\left(1, \frac{p(R' \mid T', \nu')p(T')}{p(R \mid T, \nu)p(T)} \frac{1}{M^1}\right)$$

multidistribute Flow Chart

Estimation of model parameters

Transform baseml output files to estbranches input files



F84 + G5 model

5 gamma categories

	modelinf.Gene1
/MainNeo/	
1 A	utomated file generated by paml2modelinf program
3 4	number of hmm categories
4 4	number of different categories of evolutionary processes
s F	ollowing lines match each hmm category to the category
6 O	f the evolutionary process it obeys
7	0 0 1 1
9	
10	3 3
11	
12 A	long sequence transition probs among hmm categories follow
14 0	.200000 0.200000 0.200000 0.200000 0.200000
15 0	.200000 0.200000 0.200000 0.200000 0.200000
16 0	.200000 0.200000 0.200000 0.200000 0.200000
17 0 18 N	.200000 0.200000 0.200000 0.200000 0.200000 ext line: Structure Erequencies (expected at first position)
19 0	.200000 0.200000 0.200000 0.200000 0.200000
20 n	um pred'n sumcats, next lines: whichsumcat,
21 h	ow many HMMcats fall in sumcat, labels of sumcats
22 5 23 0	1.0
24 1	1 1
25 2	1 2
26 3	1 3
27 4 28 C	ategory, base frequencies of A G T C for category, ratio of within to
29 b	etween category substitution rate for category, rel. rate of category
30	0 0.2724700 0.2465100 0.2350900 0.2459300 0.0000100 0.3481400
31	1 0.2724700 0.2465100 0.2350900 0.2459300 0.0000100 0.6354800 2 0 2724700 0 2465100 0 2350900 0 2459300 0 0000100 0 8926000
33	3 0.2724700 0.2465100 0.2350900 0.2459300 0.0000100 1.2140500
34	4 0.2724700 0.2465100 0.2350900 0.2459300 0.0000100 1.9097300
	Baca fraguancias
	Dase nequencies
	(same values among categories)

Modelinf.Gene1

Relative rates of categories



oest.Gene1



Tree topology with branch length

variance-covariance matrix of branch lengths



multicntrl.dat

Outfile of 3 genes from estbranches Outfile of 3 genes from estbranches MCMC generation setting A priori expected time between tip and root between tip and root Mean of prior distribution for rate at root node 0.97		\varTheta \varTheta 😁 📓 multicntrl.dat
Outfile of 3 genes from estbranches MCMC generation setting A priori expected time between tip and root Beam of prior distribution for rate at root node Mean of prior distribution for rate at root node Mean of prior for autocorrelate rate parameter, v Mean of prior for autocorrelate rate parameter for beta prior of proportional node depth rate parameter, v Mean of prior for autocorrelate rate parameter for beta prior match at maker parameter for beta prior distribution for rate at root node for the parameter for beta paramet		"/MainNeo/Users/inoue/7multidistribute/multicntrl.dat"((書類 - 標準) 2006/11/6 0:26:08 PM
A priori expected time between tip and root Mean of prior distribution for rate at root node Mean of prior for autocorrelate rate parameter, v Mode constraints Node constraints	Outfile of 3 genes from	□ ← ← へ ○ ↓ ↓ ↓ Shift JIS ▼ LF (UNIX) ▼ ↓ ↓
MCMC generation setting example.tree a priori expected time between tip and root Mean of prior distribution for rate at root node 0.5 orthi: a priori expected number of time units between tip and root 1.5 rtmsd: standard deviation of prior for time between tip and root 1.5 rtmsd: standard deviation of prior for rate at root node 0.6 burnin: How many cycles between tip and root 1.5 rtmsd: standard deviation of prior for rate at root node 1.6 mindb: parameter for brownian motion constant "nu" 7.1 the following lines are all needed (i.e., do not delete them) but you may not want to alter entries unless you are familiar with the computer code */ 2.10.0 thek: parameter in Markov chain proposal step 2.10.0 the parameter in Markov chain proposal step 2.10.0 the parameter in Markov chain proposal step 2.10.0 the parameter in warkov chain proposal step 2.10.0 the parameter in warkov chain proposal step 2.10.0 and will enerty be not work to constraints should follow on subsequent lines */ 3.1 anumber of constraints on node times 1.1 for number of tips which are not collected at time 0 3.1 anumber of tips which are not collected at time 0 3.2 anumber of tips which are not collected at time 0 3.2 anumber of tips which are not collected at time 0 3.2 anumber of tips which are not collected at time 0 3.2 anumber of tips which are not collected at time 0 4.2 and the times approximate prior, 0 means approximate posterior 3.2 anumber of tips which are not collected at time 0 4.2 and the times approximate prior, 0 means approximate posterior 4.2 and the times approximate prior, 0 means approximate posterior 5 commonbrown: 1 if all genes have same tendency to change rate, 0 otherwise	estoranches	<pre>1 /* the following lines are all needed in multicntrl.dat 2 do not add or delete lines but change entry on left of each 3 line as you see fit */</pre>
A priori expected time between tip and root Mean of prior distribution for rate at root node Mean of prior for autocorrelate rate parameter, v Node constraints Node constraints	MCMC generation setting	4 example.tree
A priori expected time between tip and root Mean of prior distribution for rate at root node Mean of prior for autocorrelate rate parameter, v Node constraints Node constraints Node constraints Agent of the constraints Node constraints Agent of the constraints Node constraints Agent of the constraints Agent of the constraints Node constraints		oest.Gene2
Detween up and root 1100 samprag: How many cycles between samples of the Markov chain? Mean of prior distribution for rate at root node 1.5 rtms: a priori expected number of time units between tip and root 0.07 trates: standard deviation of prior for trate at root node 0.07 trates: standard deviation of prior for rate at root node 0.47 brownsd: std. deviation of prior for brownian motion constant "nu" 0.47 trates: standard deviation of prior for brownian motion constant "nu" Mean of prior for autocorrelate 0.4 brownsd: std. deviation of prior for brownian motion constant "nu" 1.6 minab: parameter, r 0.5 othk: parameter in Markov chain proposal step 2.0 5 othk: parameter in Markov chain proposal step 2.6 bigtime: number higher than time units between tip and root could 25 othk: parameter in Markov chain proposal step 2.8 bigtime: number higher than time units between tip and root could 26 /* the program will expect the entry below to be the number of constraints and then the specified number of constraints should follow on subsequent lines */ 30 0 nodate: 1 means approximate proximate posterior 31 0 notate: 1 means approximate posterior 32 0 commonbrown: 1 if all genes have same tendency to change rate, 0 otherwise	A priori expected time	9 10000 numsamps: How many times should the Markov chain be sampled?
Mean of prior distribution for rate at root node Mean of prior for distribution for rate at root node 0.07 rtrate: mean of prior for time between tip and root 0.08 rtrates: standard deviation of prior for rate at root node 0.07 rtrates: standard deviation of prior for rate at root node 0.07 rtrates: standard deviation of prior for rate at root node 0.07 brownsen: mean of prior for brownian motion constant "nu" 0.4 brownsel: std. deviation of prior for brownian motion constant "nu" 0.4 brownsel: std. deviation of prior of proposal step 1.0 minab: parameter in Markov chain proposal step 20 0.5 thek: parameter in Markov chain proposal step 21 0.6 brightime: number higher than time units between tip and root could 25 of constraints on node times of constraints on node times 26 other program will expect the entry below to be the number of constraints and then the specified number of constraints should follow on subsequent lines */ 29 number of tips which are not collected at time 0 30 number of tips which are not collected at time 0 40 commonbrown: 1 if all genes have same tendency to change rate, 0 otherwise	between tip and root	1 100 sampfreq: How many cycles between samples of the Markov chain? 11 100002 burnin: How many cycles before the first sample of Markov chain?
Mean of prior distribution for rate at root node is 0.07 rtrates: mean of prior distribution for rate at root node is 0.07 rtrates: standard deviation of prior for rate at root node Mean of prior for autocorrelate 0.4 browmean: mean of prior for brownian motion constant "nu" Mean of prior for autocorrelate 0.4 browmsd: std. deviation of prior for brownian motion constant "nu" Mean of prior for autocorrelate 0.4 browmean: each of prior for brownian motion constant "nu" 0.4 browmsd: std. deviation of prior for brownian motion constant "nu" 0.4 10 browmsd: std. deviation of prior on proportional node depth 11 0.4 browmeanter for beta prior on proportional node depth 12 0.5 othk: parameter in Markov chain proposal step 12 0.5 thek: parameter in Markov chain proposal step 12 0.5 bigitme: number nof constraints between tip and root could be in your wildest imagination 12 0.6 bigitme: number of constraints should follow on subsequent lines */ 13 bigitme: number of constraints on node times 14 0 nodata: 1 means approximate prior, 0 means approximate posterior 14 0 </td <td></td> <td>1.5 rttm: a priori expected number of time units between tip and root 13 1.5 rttmsd: standard deviation of prior for time between tip and root</td>		1.5 rttm: a priori expected number of time units between tip and root 13 1.5 rttmsd: standard deviation of prior for time between tip and root
Node constraints Node constraints 16 0.4 browmean: mean of prior for brownian motion constant "nu" 0.4 browmean: mean of prior for brownian motion constant "nu" 0.4 browmean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.4 brownean: mean of prior for brownian motion constant "nu" 0.1 means proor proposal step 0.2 bigtime: number higher than time units between tip and root could be in your wildest imagination 0.4 brownean: expension of constraints should follow on subsequent lines */ 0.1 mumber of constraints on node times 0.1 moduta: 1 means approximate prior, 0 means approximate posterior 0 commonbrown: 1 if all genes have same tendency to change rate, 0 otherwise	Mean of prior distribution for	 14 0.07 rtrate: mean of prior distribution for rate at root node 15 0.07 rtratesd: standard deviation of prior for rate at root node
Node constraints Node constraints 18 /* the following lines are all needed (i.e., do not delete them) but you may not want to alter entries unless you are familiar with the computer code */ 20 1.0 minab: parameter for beta prior on proportional node depth 21 0.1 newk: parameter in Markov chain proposal step 22 0.5 othk: parameter in Markov chain proposal step 23 0.5 thek: parameter in Markov chain proposal step 24 100.0 bigtime: number higher than time units between tip and root could be in your wildest imagination 26 /* the program will expect the entry below to be the number of constraints and then the specified number of constraints should follow on subsequent lines */ 23 number of constraints on node times 30 1 5 1.0 30 1 5 1.0 31 0 number of tips which are not collected at time 0 34 0 nodata: 1 means approximate prior, 0 means approximate posterior 35 0 comonbrown: 1 if all genes have same tendency to change rate, 0 otherwise	rate at root node	16 0.4 brownmean: mean of prior for brownian motion constant "nu" 0.4 brownsd: std. deviation of prior for brownian motion constant "nu"
Node constraints Node constraints 1.0 minab: parameter for beta prior on proportional node depth 20 1.0 minab: parameter in Markov chain proposal step 20 0.5 othk: parameter in Markov chain proposal step 20 100.0 bigtime: number of constraints on node times 30 1 5 1.0 30 1 5 1.0 30 1 5 1.0 30 1 5 1.		18 /* the following lines are all needed (i.e., do not delete them) but you may 19 not want to alter entries unless you are familiar with the computer code */
Node constraints Node constraints Node constraints Node constraints Node constraints	Mean of prior for autocorrelate	20 1.0 minab: parameter for beta prior on proportional node depth 21 0.1 newk: parameter in Markov chain proposal step
 100.0 bigtime: number higher than time units between tip and root could be in your wildest imagination /* the program will expect the entry below to be the number of constraints and then the specified number of constraints should follow on subsequent lines */ 3 number of constraints on node times 15 1.0 1.7 1.1 2.7 1.4 3.0 number of tips which are not collected at time 0 0 nodata: 1 means approximate prior, 0 means approximate posterior 0 commonbrown: 1 if all genes have same tendency to change rate, 0 otherwise 		22 0.5 othk: parameter in Markov chain proposal step 23 0.5 thek: parameter in Markov chain proposal step
Node constraints 7* the program will expect the entry below to be the number of constraints and then the specified number of constraints should follow on subsequent lines */ 30 151.0 31 171.1 32 0 33 0 34 0 35 0 36 0 36 0 37 1 38 1 39 1 30 1 31 1 32 1 34 1 35 1 36 1 36 1		24 100.0 bigtime: number higher than time units between tip and root could be in your wildest imagination
Node constraints Node c		26 /* the program will expect the entry below to be the number of constraints
Node constraints Node c		27 and then the specified number of constraints should follow on 28 subsequent lines */
Node constraints 3 L 7 1.1 32 U 7 1.4 33 0 number of tips which are not collected at time 0 34 0 nodata: 1 means approximate prior, 0 means approximate posterior 35 0commonbrown: 1 if all genes have same tendency to change rate, 0 otherwise 36 段落35/36		29 3 number of constraints on node times 30 L 5 1.0
33 0 number of tips which are not collected at time 0 34 0 nodata: 1 means approximate prior, 0 means approximate posterior 35 0 commonbrown: 1 if all genes have same tendency to change rate, 0 otherwise 36 段落35/36	Node constraints	3 L 7 1.1 32 U 7 1.4
35 0commonbrown: 1 if all genes have same tendency to change rate, 0 otherwise 36 段落35/36		33 0 number of tips which are not collected at time 0
36 段落35/36		35 0 commonbrown: 1 if all genes have same tendency to change rate, 0 otherwise
		36 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2



multidistribute – outfile

• out.Gene1

summarizes posterior means of divergence times and rates, and other information

Itree.Gene1

contains the tree definition of the chronogram

ratio.Gene1 contains relation

contains relative probability ratios for the parameter values sampled by MCMC

6

node.Gene1

contains detailed information about the MCMC run

• samp.Gene1

contains samples from the Markov chain: they can be useful for exploring convergence

Rutschmann (2005)

node.Gene1



tree.Gene1



Time scaled tree is shown using TreeView.

Programs for molecular clock-dating that relax the clock assumption

Method, software	Penalized likelihood in 'r8s'	Multidivtime	PhyBase	BEAST
Platforms	Unix (MacOSX), Linux	Unix (MacOSX), Linux, Windows	Unix (MacOSX), Linux, Windows	Unix (MacOSX), Linux, Windows, requires Java
Optimization	Smoothing (as in regression analysis)	Bayesian MCMC	Bayesian MCMC	1.4 Bayesian MCMC
Input data	Tree with branch lengths Value for the penalty par- ameter	Tree Sequences Several priors	Tree Sequences Several priors	Sequences Several priors NO topology
Model of rate evolution	Autocorrelation between ancestral and descendent branches	Autocorrelation, with rates drawn from a log- normal distribution	Autocorrelation, with rates drawn from six different distributions (see text)	Various models are implemented and can be fitted to the data
Allows polytomies in input topology	Yes (collapse option)	Yes	No	Not applicable
Handles multiple simul- taneous calibrations	Yes	Yes	No	Yes
Handles multiple data- sets with different rates	No	Yes	No	Yes
Provides error estimates	No (separate bootstrap- ping is required)	Bayesian 95% credibility intervals	Credibility intervals must be calculated by the user	Bayesian 95% credibility intervals
Relative speed	Very fast	Medium	Slow	Fast

References

Aris-Brosou S, Yang X. 2002. Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny. Syst Biol 51: 703–714.

Kishino H, Thorne JL. 2002. 分子進化速度のベイズ型階層モデル (in Japanese). 統計数理 50: 17-31.

Kishino H, Thorne JL, Bruno WJ. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. MBE 18: 352–361.

Renner SS. 2005. Relaxed molecular clocks for dating historical plant dispersal events. Trends in Plant Science 10: 550-558.

Rutschmann F. Bayesian molecular dating using PAML/multidivtime: A step-by-step manual, version 1.5 (July 2005). ftp://statgen.ncsu.edu/pub/thorne/bayesiandating1.5.pdf

Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. MBE 15: 1647–1657.

Thorne JL, Kishino H. 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst Biol 51: 689–702.

Thorne JL and Kishino K. 2005. Estimation of divergence times from molecular sequence data. In: Statistics for Biology and Health. Nielsen R. (Ed). Springer, NY.

Thorne JL. 2003. http://statgen.ncsu.edu/thorne/multidivtime.html

Yang Z. 2003. http://workshop.molecularevolution.org/people/faculty/files/ziheng_2003.txt

Yang Z, Yoder AD. 2003. Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. Syst Biol 52: 705–716.

Yoder AD, Yang ZH. 2004. Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context. Mol Ecol 13: 757-773.