

Package ‘TreePar’

February 14, 2012

Type Package

Title Estimating speciation and extinction rates based on phylogenies

Version 2.2

Date 2011-11-01

Author Tanja Stadler

Maintainer Tanja Stadler <tanja.stadler@env.ethz.ch>

Depends ape, Matrix, subplex, TreeSim

Description For a given species phylogeny on present day data which is calibrated to calendar-time, a method for estimating maximum likelihood speciation and extinction processes is provided. The method allows for non-constant rates. Rates may change (i) as a function of time, i.e. rate shifts at specified times or mass extinction events (implemented as `bd.shifts.optim`) or (ii) as a function of the number of species, i.e. density-dependence (implemented as `bd.densdep.optim`). Note that the method takes into account the whole phylogeny, in particular it accounts for the “pull of the present” effect. For a given phylogeny on higher taxa, but where the number of species is known within each higher taxa, speciation and extinction rates can be estimated, under the assumption that these rates remained constant (implemented as `bd.groups.optim`).

License GPL-2

LazyLoad yes

Repository CRAN

Date/Publication 2011-11-02 17:17:24

R topics documented:

TreePar-package	2
bd.densdep.optim	3
bd.groups.optim	4
bd.shifts.optim	6
getx	8
treemrcashifts	9

Index	11
--------------	-----------

TreePar-package	<i>Estimating speciation and extinction rates based on phylogenies</i>
-----------------	--

Description

For a given species phylogeny on present day data which is calibrated to calendar-time, a method for estimating maximum likelihood speciation and extinction processes is provided. The method allows for non-constant rates. Rates may change (i) as a function of time, i.e. rate shifts at specified times or mass extinction events (implemented as `bd.shifts.optim`) or (ii) as a function of the number of species, i.e. density-dependence (implemented as `bd.densdep.optim`). Note that the method takes into account the whole phylogeny, in particular it accounts for the "pull of the present" effect.

For a given phylogeny on higher taxa, but where the number of species is known within each higher taxa, speciation and extinction rates can be estimated, under the assumption that these rates remained constant (implemented as `bd.groups.optim`).

Details

Package:	TreePar
Type:	Package
Version:	2.2
Date:	2011-11-01
License:	GPL-2
LazyLoad:	yes

Author(s)

Tanja Stadler
 <<http://www.tb.ethz.ch/people/tstadler>>

References

T. Stadler: Mammalian phylogeny reveals recent diversification rate shifts. PNAS 108 (15), 6187-6192, 2011.

R.S. Etienne, B. Haegeman, T. Stadler, T. Aze, P.N. Pearson, A. Purvis and A.B. Phillimore: Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proc. Roy. Soc. B, doi: 10.1098/rspb.2011.1439, 2011.

T. Stadler, F. Bokma: Estimating speciation and extinction rates for phylogenies of higher taxa. Submitted, 2011.

See Also

ape TreeSim

bd.densdep.optim	<i>bd.densdep.optim: Estimating speciation and extinction rates in phylogenies under a density-dependent speciation model</i>
------------------	---

Description

bd.densdep.optim estimates the maximum likelihood speciation and extinction rates under a density-dependent speciation model. Speciation rate is a function of the number of species N , $\lambda(N) = \max(0, \lambda(1-N/K))$, and extinction rate is μ (constant).

Usage

```
bd.densdep.optim(x,minK,maxK,discrete=TRUE,continuous=FALSE,lambdainit=2,muinit=1,Kinit=0,Yule=FALSE)
```

Arguments

x	Vector of speciation times in the phylogeny. Time is measured increasing going into the past with the present being time 0. x can be obtained from a phylogenetic tree using <code>getx(TREE)</code> .
minK	Minimal value of K (when <code>discrete=TRUE</code>). Default is <code>minK = (number of species)</code> .
maxK	Maximal value of K (when <code>discrete=TRUE</code>). Default is <code>maxK = 1.5(number of species)</code> .
discrete	If <code>discrete=TRUE</code> , the likelihood function is maximized with K being an integer and the minimal size being <code>minK</code> and the maximal size being <code>maxK</code> .
continuous	If <code>continuous=TRUE</code> , the likelihood function is maximized with K being a continuous parameter. The function <code>subplex</code> is used for optimization and sometimes gets stuck at a non-optimal K. Thus I always recommend to also calculate with <code>discrete=TRUE</code> .
lambdainit	Initial lambda value for optimization when K is continuous (default is 2).
muinit	Initial mu value for optimization when K is continuous (default is 1).
Kinit	Initial K value for optimization when K is continuous (default is <code>(number of species)+1</code>).
Yule	<code>Yule=FALSE</code> is default. <code>Yule=TRUE</code> sets $\mu=0$, i.e. no extinction.

muset	muset=0 (default) maximizes over the whole parameter range. muset>0 means that the optimization is done over all $\mu > \text{muset}$. muset<0 fixes $\mu = -\text{muset}$.
rho	rho=1 is default meaning all species are sampled. rho<1 assumes that the phylogeny is incomplete, and each species is included with probability rho.
model	model=-1 (default) is the density-dependent model. model=0 (only relevant for testing purposes) assumes that lambda is constant for number of species < K, and 0 for number of species $\geq K$. model=0 is used for testing / comparing to constant rate model implemented in bd.shifts.optim.

Value

res	Maximum likelihood speciation and extinction rate and the saturation value K; the first entry, res[[1]], is the result when K being discrete (0 if discrete=FALSE) and the second entry, res[[2]], is the result when K being continuous (0 if continuous=FALSE). \$par is (lambda,mu,K). \$value is -log likelihood. The likelihood is calculated assuming there were two lineages at the time of the root. The likelihood is NOT conditioned on survival of the two lineages. Likelihood-values from bd.shifts.optim are directly comparable (eg AIC). Likelihood-values from laser are comparable after the transformation $-\$value + (\sum(\log(2:\text{length}(x))) - (\text{length}(x)-1)*\log(2))$.
-----	--

Note

bd.densdep.optim(x, Yule=TRUE, discrete=FALSE, cont=TRUE) in TreePar and DDL(x) in Laser return the same results (up to transforming the -log likelihood (\$value) from TreePar via $-\$value + (\sum(\log(2:\text{length}(x))) - (\text{length}(x)-1)*\log(2))$)

Author(s)

Tanja Stadler

References

R.S. Etienne, B. Haegeman, T. Stadler, T. Aze, P.N. Pearson, A. Purvis and A.B. Phillimore: Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Submitted to Proc. Roy. Soc. B.

bd.groups.optim	<i>bd.groups.optim: Estimating speciation and extinction rates in higher-taxonomic-level phylogenies</i>
-----------------	--

Description

bd.groups.optim estimates the maximum likelihood speciation and extinction rates under a constant rate speciation and extinction model based on a phylogeny on higher taxa, where the number of species within each taxa are known.

Usage

```
bd.groups.optim(phy, S, xcut=0, lambda=0, mu=0)
```

Arguments

phy	Phylogenetic tree on higher taxa.
S	Vector of number of species for each higher taxa.
xcut	Each species extant at cutoff time xcut in the past defines a higher taxa (implying that xcut must be after the most recent branching time in phi). xcut=0 is used if no xcut is specified and means that higher taxa are not defined using a cutoff, but higher taxa are randomly selected clades in the species phylogeny.
lambda	lambda=0 default. If lambda>0, then the likelihood for the given lambda and mu is calculated instead of finding the optimal parameter.
mu	see lambda

Value

res	IF lambda>0: return value is - log likelihood for the specified lambda and mu. ELSE result is maximum likelihood estimate:
res[[1]]	First row: Maximum likelihood estimate for turnover (mu/lambda) with 95 per cent confidence intervals. Second row: Maximum likelihood estimate for diversification rate (lambda-mu) with 95 per cent confidence intervals.
res[[2]]\$value	- Log Likelihood value
res[[2]]\$convergence	Convergence information. 0 is successful. For other output see man for subplex.

Note

The likelihood is calculated assuming there were two lineages at the time of the root. The likelihood is NOT conditioned on survival of the two lineages. Likelihood-values from bd.shifts.optim are directly comparable (eg AIC). Likelihood-values from laser are comparable after the transformation $-\text{res} + (\sum(\log(2:\text{length}(x))) - (\text{length}(x) - 1) * \log(2))$.

Author(s)

Tanja Stadler

References

T. Stadler, F. Bokma: Estimating speciation and extinction rates for phylogenies of higher taxa. Submitted, 2011.

Examples

```
### An example from Paradis (2003) using the avian orders:
data(bird.orders)
### Number of species in each order from Sibley and Monroe (1990):
S <- c(10, 47, 69, 214, 161, 17, 355, 51, 56, 10, 39, 152, 6, 143, 358, 103, 319, 23, 291, 313, 196, 1027, 5712)
xcut<-95*0.207407
out<-bd.groups.optim(bird.orders, S, xcut)
### Not accounting for the cutoff when defining the bird taxa estimates 0 extinction:
out1<-bd.groups.optim(bird.orders, S, 0)
### xcut=0 produces almost the same results as the following ape function
bd.ext(bird.orders, S)
```

bd.shifts.optim	<i>bd.shifts.optim: Estimating speciation and extinction rate changes and mass extinction events in phylogenies</i>
-----------------	---

Description

bd.shifts.optim estimates the maximum likelihood speciation and extinction rates together with the rate shift times $t=(t_1, t_2 \dots, t_m)$ in a (possibly incomplete sampled) phylogeny. At the times t , the rates are allowed to change and the species may undergo a mass extinction event.

Usage

```
bd.shifts.optim(x, sampling, grid, start, end, maxitk = 5, yule = FALSE, ME = FALSE, all = FALSE, positiv
```

Arguments

x	Vector of speciation times in the phylogeny. Time is measured increasing going into the past with the present being time 0. x can be obtained from a phylogenetic tree using <code>getx(TREE)</code> .
sampling	Vector of length m. <code>sampling_i</code> is the probability of a species surviving the mass extinction at time t_i . <code>sampling_1</code> is the probability of an extant species being sampled. <code>sampling_1=1</code> means that the considered phylogeny is complete. <code>sampling_i=1</code> ($i>1$) means that at time t_i , a rate shift may occur but no species go extinct. If <code>ME=TRUE</code> , all entries but <code>sampling_1</code> will be discarded as they are estimated (however, input a vector <code>sampling</code> of the appropriate length such that the program knows how many mass extinction events you want to allow for).
grid, start, end	The model parameters are optimized for different fixed rate shift times. The fixed rate shift times are specified by being at <code>(start, start+grid, start+2*grid .. end)</code> . I calculate the likelihood for the different rate shift times t instead of optimizing t with the function <code>optim</code> used for the other parameters, as the optimization performed poor for t (namely getting stuck in local optima).
yule	<code>yule=TRUE</code> sets the extinction rates to zero.

maxitk	Integer value defining how many iterations shall be done in the optimization. Default is 5, but needs to be increased if too many warnings "convergence problem" appear.
ME	ME=FALSE (default) uses the mass extinction fractions specified in sampling and does not estimate them. If ME=FALSE is used with sampling=c(1,1, .. , 1), no mass extinction events are considered.
all	Only relevant when ME=TRUE. all=FALSE (default and recommended) estimates one speciation and one extinction rate for the whole tree, and estimates the intensities sampling_i (i>1) of mass extinction events. all=TRUE allows for varying speciation and extinction rates. Since the parameters might correlate, all=TRUE is not recommended.
posdiv	posdiv=FALSE (default) allows the speciation - extinction rate to be negative, i.e. allows for periods of declining diversity. posdiv=TRUE forces the speciation - extinction rate to be positive.
miniall	If you run the bd.shifts.optim for k shifts, but you now want to have K>k shifts, then set for continuing the analysis: update sampling, and set miniall=res[[2]] where res[[2]] is the output from the run with k shifts.

Value

res[[1]][[i]]	List of maximum likelihood parameter estimates for each fixed t where i-1 shifts are allowed to occur (i in 1:m).
res[[2]][[i]]	Maximum likelihood parameter estimates for i-1 shifts (i in 1:m): First entry is the (-log likelihood) value. The next i entries are the turnover (extinction/speciation) estimates, for the successive intervals going back in time. The next i entries are the diversification rate estimates (speciation-extinction). The next i-1 entries are the sampling estimates (if ME=TRUE). The last i-1 entries are the shift times. (Note: if ME=TRUE and all==FALSE, the second entry is the turnover, the third the diversification rate, followed by the sampling estimates).
res[[3]]	Vector of time points where the function was evaluated.
res[[4]]	Array specifying the time points when there was a convergence problem: a row of res[[4]] with entry (i,t_i) means that when adding the i-th shift at time t_i, a convergence problem was encountered.

Note

The likelihood is calculated assuming there were two lineages at the time of the root. The likelihood is NOT conditioned on survival of the two lineages. Likelihood-values from bd.densdep.optim are directly comparable (eg AIC). Likelihood-values from laser are comparable after the transformation $-\text{res} + (\sum(\log(2:\text{length}(x))) - (\text{length}(x) - 1) * \log(2))$.

Author(s)

Tanja Stadler

References

T. Stadler: Mammalian phylogeny reveals recent diversification rate shifts. PNAS 108 (15), 6187-6192, 2011.

Examples

```
set.seed(1)

# First we simulate a tree, and then estimate the parameters for the tree:
# Number of species
nspecies <- 20
# At time 1 and 2 in the past, we have a rate shift:
time <- c(0,1,2)
# Mass extinction intensities 0.5 at time 1 in past, 0.4 at time 2 in past. Present day species are all sampled (rho_1)
rho <- c(1,0.5,0.4)
# speciation rates (between t_i,t_{i+1} we have speciation rate lambda_i):
lambda <- c(2,2,1)
# extinction rates (between t_i,t_{i+1} we have extinction rate mu_i):
mu <- c(1,1,0)
# Simulation of a tree:
tree<-sim.rateshift.taxa(nspecies,1,lambda,mu,frac=rho,times=time,complete=FALSE)
# Extracting the speciation times x:
x<-sort(getx(tree[[1]][[1]]),decreasing=TRUE)

# When estimating the shift times t for x, we allow the shift times to be 0.6, 0.8, 1, 1.2, .. ,2.4:
start <- 0.6
end <- 2.4
grid <- 0.2

# We estimate time, rho, lambda, mu:
resrho <- bd.shifts.optim(x,rho,grid,start,end,ME=TRUE)
resrho[[2]]
# We fix rho and estimate time, lambda, mu:
res <- bd.shifts.optim(x,rho,grid,start,end)
res[[2]]
# We fix rho=1 and mu=0 and then estimate time, lambda:
resyule <- bd.shifts.optim(x,rho,grid,start,end,yule=TRUE)
resyule[[2]]
```

getx

getx: Getting the vector of speciation times x from a phylogenetic tree

Description

getx calculates the input vector x for the rate estimation functions from a phylogenetic tree (which may have polytomies)

Usage

```
getx(datatree)
```

Arguments

datatree phylogenetic tree

Value

x Vector of speciation times

Author(s)

Tanja Stadler

References

#T. Stadler: Mammalian phylogeny reveals recent diversification rate shifts. In press, PNAS, 2011.

treemrcashifts	<i>treemrcashifts: calculates the likelihood of a tree</i>
----------------	--

Description

treemrcashifts: calculates the likelihood of a tree given the speciation and extinction rates and shift times.

Usage

```
treemrcashifts(x, t, l, mu, sampling, posdiv=FALSE)
```

Arguments

x	Vector of speciation times in the phylogeny. Time is measured increasing going into the past with the present being time 0. x can be obtained from a phylogenetic tree using <code>getx(TREE)</code> .
t	The time of rate shifts (<code>t[1]=0</code> is required, being the present)
l, mu	Vectors of the same length as t. <code>l[i]</code> (resp. <code>mu[i]</code>) specifies the speciation (resp. extinction rate) prior to <code>t[i]</code> .
sampling	Vector of length m. <code>sampling_i</code> is the probability of a species surviving the mass extinction at time <code>t_i</code> . <code>sampling_1</code> is the probability of an extant species being sampled. <code>sampling_1=1</code> means that the considered phylogeny is complete. <code>sampling_i=1</code> ($i>1$) means that at time <code>t_i</code> , a rate shift may occur but no species go extinct.
posdiv	Not relevant when using <code>treemrcashifts</code> without optimizing (for <code>bd.shifts.optim</code> : <code>posdiv=FALSE</code> (default) allows the speciation - extinction rate to be negative, i.e. allows for periods of declining diversity. <code>posdiv=TRUE</code> forces the speciation - extinction rate to be positive).

Value

res -log likelihood of the tree given the parameters. The likelihood is calculated assuming there were two lineages at the time of the root. The likelihood is NOT conditioned on survival of the two lineages. Likelihood-values from *bd.densdep.optim* are directly comparable (eg AIC). Likelihood-values from *laser* are comparable after the transformation $-\text{res} + (\sum(\log(2:\text{length}(x))) - (\text{length}(x) - 1) * \log(2))$.

Author(s)

Tanja Stadler

References

T. Stadler: Mammalian phylogeny reveals recent diversification rate shifts. *PNAS* 108 (15), 6187-6192, 2011.

Index

`bd.densdep.optim`, 3

`bd.groups.optim`, 4

`bd.shifts.optim`, 6

`getx`, 8

`treemrcashifts`, 9

`TreePar (TreePar-package)`, 2

`TreePar-package`, 2