

Package ‘DDD’

September 24, 2012

Type Package

Title Diversity-dependent diversification

Version 1.4

Date 2012-09-24

Depends R (>= 2.14.2), optimx, deSolve, laser

Suggests BB, Rcgmin, Rvmmmin, minqa, numDeriv, ucminf

Author Rampal S. Etienne & Bart Haegeman

Maintainer Rampal S. Etienne <r.s.etienne@rug.nl>

License GPL-2

Description DDD implements maximum likelihood methods based on the diversity-dependent birth-death process to test whether speciation or extinction are diversity-dependent.
See Etienne et al. 2012, Proc. Roy. Soc. B 279: 1300-1309, doi: 10.1098/rspb.2011.1439.
DDD also implements a maximum likelihood method to detect key innovations defined as decoupling of diversity-dependence.
See Etienne & Haegeman 2012, Am. Nat. In press, doi: 10.1086/667574.

R topics documented:

DDD-package	2
dd_KI_loglik	3
dd_KI_ML	5
dd_loglik	7
dd_ML	8
dd_SR_loglik	10
dd_SR_ML	11
Index	14

Description

This package is a likelihood-based statistical package to estimate parameters under diversity-dependent diversification models.

New in version 0.3:

- Models with shifts in the parameters at a time tshift.

New in version 0.4:

- Bug fix of prefactor in likelihood - bug did not affect parameter estimates.
- There is an option to condition or not on non-extinction of the phylogeny.
- The functions dd_loglik and dd_SR_loglik now accept untransformed parameters; the transformation is done in the auxiliary functions.
- Improved help pages.
- Auxiliary functions no longer in help pages.

New in version 1.0:

- Another bug fix of prefactor in likelihood - bug did not affect parameter estimates.
- Models with key innovations defined as decoupling of diversity-dependent dynamics.
- Option to compute the likelihood of a set of branching times or of the phylogeny; these differ by a constant factor.

New in version 1.1

- A model with diversity-dependence in speciation and extinction rate, currently only available for the likelihood without shifts or decoupling.

New in version 1.2

- Possibility of automatically exploring multiple initial conditions for the SR models
- Suppressing unnecessary output
- Flushing computed likelihood values immediately (even with buffer on)
- Bug fix in default value of the initial time of decoupling

New in version 1.3

- Fix of default length of idparsopt
- Fix of conversion problems for small branching times
- Default initial parameter values in dd_ML based on birth-death model

New in version 1.4

- Bug fix for when there is a large number of missing species

Details

Package:	DDD
Type:	Package
Version:	1.4
Date:	2012-09-24
License:	GPL 2.0

dd_loglik computes the likelihood of any of 4 different models of diversity-dependent diversification, given a set of parameters and a data set of phylogenetic branching times.

dd_ML finds the parameters that maximizes the likelihood computed by dd_loglik.

dd_SR_loglik is the same as dd_loglik except that it allows a shift in the parameters at time tshift.

dd_SR_ML finds the parameters that maximizes the likelihood computed by dd_SR_loglik.

dd_KI_loglik is the same as dd_loglik except that it allows decoupling of a subclade's dynamics from the main clade's dynamics potentially accompanied by a shift in parameters at time t_d.

dd_KI_ML finds the parameters that maximizes the likelihood computed by dd_KI_loglik.

Author(s)

Rampal S. Etienne & Bart Haegeman Maintainer: Rampal S. Etienne (r.s.etienne@rug.nl)

References

Etienne et al. 2012, Proc. Roy. Soc. B 279: 1300-1309, doi: 10.1098/rspb.2011.1439

Etienne & Haegeman 2012, Am. Nat. In press, doi: 10.1086/667574

See Also

laser

Examples

```
cat("Estimate the parameters of the diversity-dependent model")
cat("for a set of branching times with the default settings:")
brts = 1:20
dd_ML(brts)
```

```
cat("Estimating the parameters of the diversity-dependent model with shifting rates")
cat("for a set of branching times brts with the default settings:")
brts = 1:20
dd_SR_ML(brts)
```

```
cat("Estimating the parameters of the diversity-dependent model with decoupling")
cat("for a set of branching times with the default settings:")
brtsM = 1:20
brtsS = seq(0.2,1.5,0.3)
tsplit = 5
dd_ML(brtsM,brtsS,tsplit)
```

dd_KI_loglik

Loglikelihood for diversity-dependent diversification models with decoupling of a subclade from a main clade at time $t = t_d$

Description

This function computes loglikelihood of a diversity-dependent diversification model for a given set of branching times and parameter values where the diversity-dependent dynamics of a subclade decouple from the dynamics of the main clade at time t_d, potentially accompanied by a shift in parameters.

Usage

```
dd_KI_loglik(pars1, pars2, brtsM, brtsS, missnumspec)
```

Arguments

pars1	<p>Vector of parameters:</p> <p>pars1[1] corresponds to lambda_M (speciation rate) of the main clade pars1[2] corresponds to mu_M (extinction rate) of the main clade pars1[3] corresponds to K_M (clade-level carrying capacity) of the main clade pars1[4] corresponds to lambda_M (speciation rate) of the subclade pars1[5] corresponds to mu_S (extinction rate) of the subclade pars1[6] corresponds to K_S (clade-level carrying capacity) of the subclade pars1[7] corresponds to t_d (the time of decoupling)</p>
pars2	<p>Vector of model settings:</p> <p>pars2[1] sets the maximum number of species for which a probability must be computed. This must be larger than 1 + length(brts).</p> <p>pars2[2] sets the model of diversity-dependence: - pars2[2] == 1 linear dependence in speciation rate - pars2[2] == 2 exponential dependence in speciation rate - pars2[2] == 3 linear dependence in extinction rate - pars2[2] == 4 exponential dependence in extinction rate</p> <p>pars2[3] sets the conditioning on non-extinction of the phylogeny (TRUE or FALSE) pars2[4] sets the time of splitting of the branch that will decouple</p>
brtsM	A set of branching times of the main clade in the phylogeny, all positive
brtsS	A set of branching times of the subclade in the phylogeny, all positive
missnumspec	The number of species that are missing in the phylogeny

Value

The loglikelihood

Author(s)

Rampal S. Etienne & Bart Haegeman

References

- Etienne et al. 2012. Proc. Roy. Soc. B 279: 1300-1309, doi: 10.1098/rspb.2011.1439 - Etienne & Haegeman 2012. Am. Nat. In press.

See Also

[dd_KI_ML](#), [dd_loglik](#) [dd_SR_loglik](#)

Examples

```
pars1 = c(0.25,0.12,25.51,1.0,0.16,8.61,9.8)
pars2 = c(200,1,TRUE,18.8)
missnumspec = 0
brtsM = c(25.2,24.6,24.0,22.5,21.7,20.4,19.9,19.7,18.8,17.1,15.8,11.8,9.7,8.9,5.7,5.2)
brtsS = c(9.6,8.6,7.4,4.9,2.5)
dd_KI_loglik(pars1,pars2,brtsM,brtsS,missnumspec)
```

dd_KI_ML

Maximization of the loglikelihood under a diversity-dependent diversification model with decoupling of a subclade's diversification dynamics from the main clade's dynamics

Description

This function computes the maximum likelihood estimates of the parameters of a diversity-dependent diversification model with decoupling of the diversification dynamics of a subclade from the dynamics of the main clade for a given set of phylogenetic branching times of main clade and subclade and the time of splitting of the lineage that will form the subclade. It also outputs the corresponding loglikelihood that can be used in model comparisons.

Usage

```
dd_KI_ML(brtsM, brtsS, tsplit, initparsopt = c(0.5, 0.1, 2 * (1 + length(brtsM)), 2 * (1 + length(brtsS))/2), parsfix = NULL, idparsopt = c(1:3, 6:7), idparsfix = NULL, idparsnoshift = (1:7)[c(-idparsopt, (-1)^(length(idparsfix) != 0) * idparsfix)], res = 10 * (1 + length(c(brtsM, brtsS)) + missnumspec), ddmodel = 1, missnumspec = 0, cond = 0)
```

Arguments

brtsM	A set of branching times of the main clade in a phylogeny, all positive
brtsS	A set of branching times of the subclade in a phylogeny, all positive
tsplit	The branching time at which the lineage forming the subclade branches off, positive
initparsopt	The initial values of the parameters that must be optimized
parsfix	The values of the parameters that should not be optimized
idparsopt	The ids of the parameters that must be optimized, e.g. 1:7 for all parameters. The ids are defined as follows: id == 1 corresponds to lambda_M (speciation rate) of the main clade id == 2 corresponds to mu_M (extinction rate) of the main clade id == 3 corresponds to K_M (clade-level carrying capacity) of the main clade id == 4 corresponds to lambda_S (speciation rate) of the subclade id == 5 corresponds to mu_S (extinction rate) of the subclade id == 6 corresponds to K_S (clade-level carrying capacity) of the subclade id == 7 corresponds to t_d (the time of decoupling)
idparsfix	The ids of the parameters that should not be optimized, e.g. c(1,3,4,6) if lambda and K should not be optimized, but only mu. In that case idparsopt must be c(2,5,7). The default is to fix all parameters not specified in idparsopt.

idparsnoshift	The ids of the parameters that should not shift; This can only apply to ids 4, 5 and 6, e.g. idparsnoshift = c(4,5) means that lambda and mu have the same values before and after tshift
res	sets the maximum number of species for which a probability must be computed, must be larger than $1 + \max(\text{length}(\text{brtsM}), \text{length}(\text{brtsS}))$
ddmodel	sets the model of diversity-dependence: ddmodel == 1 : linear dependence in speciation rate ddmodel == 2 : exponential dependence in speciation rate ddmodel == 3 : linear dependence in extinction rate ddmodel == 4 : exponential dependence in extinction rate
missnumspec	The number of species that are in the clade but missing in the phylogeny
cond	Conditioning on non-extinction of the phylogeny (TRUE or FALSE)

Details

The output is a dataframe where \$par and \$fvalues are the first elements. There are more elements that are all from the optimization algorithm, optimx. The computed loglikelihood contains the factor $q! m!/(q + m)!$ where q is the number of species in the phylogeny and m is the number of missing species, as explained in the supplementary material to Etienne et al. 2012.

Value

par	gives the list of parameter estimates
fvalues	gives the maximum loglikelihood

Note

The optimization may get trapped in local optima. Try different starting values to search for the global optimum.

Author(s)

Rampal S. Etienne & Bart Haegeman

References

- Etienne et al. 2012, Proc. Roy. Soc. B 279: 1300-1309, doi: 10.1098/rspb.2011.1439
- Etienne & Haegeman 2012. Am. Nat. In press.

See Also

[dd_KI_loglik](#), [dd_ML](#), [dd_SR_ML](#), [optimx](#)

Examples

```
cat("Estimating parameters for two sets of branching times brtsM, brtsS with the default settings:")
brtsM = 1:20
brtsS = seq(0.1, 3.9, 0.5)
t_split = 5;
dd_KI_ML(brtsM, brtsS, tsplit)
```

dd_loglik

*Loglikelihood for diversity-dependent diversification models***Description**

This function computes loglikelihood of a diversity-dependent diversification model for a given set of branching times and parameter values.

Usage

```
dd_loglik(pars1, pars2, brts, missnumspec)
```

Arguments

pars1	Vector of parameters: pars1[1] corresponds to lambda (speciation rate) before the shift pars1[2] corresponds to mu (extinction rate) before the shift pars1[3] corresponds to K (clade-level carrying capacity) before the shift
pars2	Vector of model settings: pars2[1] sets the maximum number of species for which a probability must be computed. This must be larger than 1 + length(brts). pars2[2] sets the model of diversity-dependence: - pars2[2] == 1 linear dependence in speciation rate - pars2[2] == 2 exponential dependence in speciation rate - pars2[2] == 3 linear dependence in extinction rate - pars2[2] == 4 exponential dependence in extinction rate pars2[3] sets the conditioning on non-extinction of the phylogeny (TRUE or FALSE) pars2[4] sets whether the likelihood is for the branching times (0) or the phylogeny (1)
brts	A set of branching times of a phylogeny, all positive
missnumspec	The number of species that are in the clade but missing in the phylogeny

Value

The loglikelihood

Author(s)

Rampal S. Etienne & Bart Haegeman

References

Etienne et al. 2011, Proc. Roy. Soc. B, doi: 10.1098/rspb.2011.1439

See Also

[dd_ML](#), [dd_SR_loglik](#), [dd_KI_loglik](#)

Examples

```
dd_loglik(pars1 = c(0.5,0.1,100), pars2 = c(100,1,TRUE,1), brts = 1:10, missnumspec = 0)
```

dd_ML	<i>Maximization of the loglikelihood under a diversity-dependent diversification model</i>
-------	--

Description

This function computes the maximum likelihood estimates of the parameters of a diversity-dependent diversification model for a given set of phylogenetic branching times. It also outputs the corresponding loglikelihood that can be used in model comparisons.

Usage

```
dd_ML(brts, initparsopt = if (ddmodel < 5) { c(0.1 + bd(as.numeric(brts))$r1/(1 - bd(as.numeric(
  0.1, 2 * (length(brts) + missnumspec)))} else { c(0.1 + bd(as.numeric(brts))$r1/(1 - bd(as.num
  0.1, 2 * (length(brts) + missnumspec), 0.01)}, idparsopt = 1:length(initparsopt), idparsfix =
  5)))[-idparsopt], parsfix = (ddmodel < 5) * c(0.2,0.1, 2 * (length(brts) + missnumspec))[-idp
  c(0.2, 0.1, 2 * (length(brts) + missnumspec), 0)[-idparsopt], res = 10 * (1 + length(brts) +
  missnumspec = 0, cond = TRUE, btorph = 1)
```

Arguments

brts	A set of branching times of a phylogeny, all positive
initparsopt	The initial values of the parameters that must be optimized
idparsopt	The ids of the parameters that must be optimized, e.g. 1:3 for all parameters. The ids are defined as follows: id == 1 corresponds to lambda (speciation rate) id == 2 corresponds to mu (extinction rate) id == 3 corresponds to K (clade-level carrying capacity)
idparsfix	The ids of the parameters that should not be optimized, e.g. c(1,3) if lambda and K should not be optimized, but only mu. In that case idparsopt must be 2. The default is to fix all parameters not specified in idparsopt.
parsfix	The values of the parameters that should not be optimized
res	Sets the maximum number of species for which a probability must be computed, must be larger than 1 + length(brts)
ddmodel	Sets the model of diversity-dependence: ddmodel == 1 : linear dependence in speciation rate ddmodel == 2 : exponential dependence in speciation rate ddmodel == 3 : linear dependence in extinction rate ddmodel == 4 : exponential dependence in extinction rate
missnumspec	The number of species that are in the clade but missing in the phylogeny
cond	Conditioning on non-extinction of the phylogeny (TRUE or FALSE)
btorph	Sets whether the likelihood is for the branching times (0) or the phylogeny (1)

Details

The output is a dataframe where \$par and \$fvalues are the first elements. There are more elements that are all from the optimization algorithm, optimx. The computed loglikelihood contains the factor $q! m! / (q + m)!$ where q is the number of species in the phylogeny and m is the number of missing species, as explained in the supplementary material to Etienne et al. 2012.

Value

par	gives the list of parameter estimates
fvalues	gives the maximum loglikelihood

Author(s)

Rampal S. Etienne & Bart Haegeman

References

Etienne et al. 2011, Proc. Roy. Soc. B, doi: 10.1098/rspb.2011.1439

See Also

[dd_loglik](#), [dd_SR_ML](#), [dd_KI_ML](#), [optimx](#)

Examples

```
cat("Estimating parameters for a set of branching times brts with the default settings:")
brts = 1:20
dd_ML(brts)

cat("Estimating the intrinsic speciation rate lambda and the carrying capacity K")
cat("for a fixed extinction rate of 0.1:")
brts = 1:20
dd_ML(brts, initparsopt = c(0.2, 100), idparsopt = c(1, 3), parsfix = 0.1)

cat("This will estimate all parameters under the linear diversity-dependent")
cat("extinction model:")
brts = 1:20
dd_ML(brts, ddmodel = 3)

cat("This will estimate all parameters under the default model")
cat("when there are 2 missing species:")
brts = 1:20
dd_ML(brts, missnumspec = 2)
```

dd_SR_loglik	<i>Loglikelihood for diversity-dependent diversification models with a shift in the parameters at time $t = tshift$</i>
--------------	--

Description

This function computes loglikelihood of a diversity-dependent diversification model for a given set of branching times and parameter values where the parameters are allowed to shift at time $t = tshift$

Usage

```
dd_SR_loglik(pars1, pars2, brts, missnumspec)
```

Arguments

pars1	Vector of parameters: pars1[1] corresponds to lambda (speciation rate) before the shift pars1[2] corresponds to mu (extinction rate) before the shift pars1[3] corresponds to K (clade-level carrying capacity) before the shift pars1[4] corresponds to lambda (speciation rate) after the shift pars1[5] corresponds to mu (extinction rate) after the shift pars1[6] corresponds to K (clade-level carrying capacity) after the shift pars1[7] corresponds to tshift (the time of shift)
pars2	Vector of model settings: pars2[1] sets the maximum number of species for which a probability must be computed. This must be larger than $1 + \text{length}(\text{brts})$. pars2[2] sets the model of diversity-dependence: - pars2[2] == 1 linear dependence in speciation rate - pars2[2] == 2 exponential dependence in speciation rate - pars2[2] == 3 linear dependence in extinction rate - pars2[2] == 4 exponential dependence in extinction rate pars2[3] sets the conditioning on non-extinction of the phylogeny (TRUE or FALSE) pars2[4] sets whether the likelihood is for the branching times (0) or the phylogeny (1)
brts	A set of branching times of a phylogeny, all positive
missnumspec	The number of species that are in the clade but missing in the phylogeny

Value

The loglikelihood

Author(s)

Rampal S. Etienne & Bart Haegeman

References

- Etienne et al. 2012, Proc. Roy. Soc. B 279: 1300-1309, doi: 10.1098/rspb.2011.1439 - Etienne & Haegeman 2012. Am. Nat. In press.

See Also

[dd_SR_ML](#), [dd_loglik](#), [dd_KI_loglik](#)

Examples

```
dd_SR_loglik(pars1 = c(0.2,0.1,50,0.2,0.1,70,5), pars2 = c(100,1,TRUE,1), brts = 1:10, missnumspec = 0)
```

dd_SR_ML	<i>Maximization of the loglikelihood under a diversity-dependent diversification model with a shift in the parameters</i>
----------	---

Description

This function computes the maximum likelihood estimates of the parameters of a diversity-dependent diversification model with shifting parameters at time $t = tshift$ for a given set of phylogenetic branching times. It also outputs the corresponding loglikelihood that can be used in model comparisons.

Usage

```
dd_SR_ML(brts, initparsopt = c(0.5, 0.1, 2 * (1 + length(brts) + missnumspec),
  2 * (1 + length(brts) + missnumspec), max(brts)/2), parsfix = NULL, idparsopt = c(1:3, 6:7),
  idparsfix = NULL, idparsnoshift = (1:7)[c(-idparsopt, (-1)^(length(idparsfix) != 0) * idparsf
  res = 10 * (1 + length(brts) + missnumspec), ddmodel = 1, missnumspec = 0, cond = TRUE, btorp
  allbp = FALSE)
```

Arguments

brts	A set of branching times of a phylogeny, all positive
initparsopt	The initial values of the parameters that must be optimized
parsfix	The values of the parameters that should not be optimized
idparsopt	The ids of the parameters that must be optimized, e.g. 1:7 for all parameters. The ids are defined as follows: id == 1 corresponds to lambda (speciation rate) before the shift id == 2 corresponds to mu (extinction rate) before the shift id == 3 corresponds to K (clade-level carrying capacity) before the shift id == 4 corresponds to lambda (speciation rate) after the shift id == 5 corresponds to mu (extinction rate) after the shift id == 6 corresponds to K (clade-level carrying capacity) after the shift id == 7 corresponds to tshift (the time of shift)
idparsfix	The ids of the parameters that should not be optimized, e.g. c(1,3,4,6) if lambda and K should not be optimized, but only mu. In that case idparsopt must be c(2,5,7). The default is to fix all parameters not specified in idparsopt.

<code>idparsnoshift</code>	The ids of the parameters that should not shift; This can only apply to ids 4, 5 and 6, e.g. <code>idparsnoshift = c(4,5)</code> means that <code>lambda</code> and <code>mu</code> have the same values before and after <code>tshift</code>
<code>res</code>	sets the maximum number of species for which a probability must be computed, must be larger than $1 + \text{length}(\text{brts})$
<code>ddmodel</code>	sets the model of diversity-dependence: <code>ddmodel == 1</code> : linear dependence in speciation rate <code>ddmodel == 2</code> : exponential dependence in speciation rate <code>ddmodel == 3</code> : linear dependence in extinction rate <code>ddmodel == 4</code> : exponential dependence in extinction rate
<code>missnumspec</code>	The number of species that are in the clade but missing in the phylogeny
<code>cond</code>	Conditioning on non-extinction of the phylogeny (TRUE or FALSE)
<code>btorph</code>	Sets whether the likelihood is for the branching times (0) or the phylogeny (1)
<code>allbp</code>	Sets whether a search should be done with various initial conditions, with <code>tshift</code> at each of the branching points (TRUE/FALSE)

Details

The output is a dataframe where `$par` and `$fvalues` are the first elements. There are more elements that are all from the optimization algorithm, `optimx`. The computed loglikelihood contains the factor $q! m! / (q + m)!$ where q is the number of species in the phylogeny and m is the number of missing species, as explained in the supplementary material to Etienne et al. 2012.

Value

<code>pars</code>	gives the list of parameter estimates
<code>fvalues</code>	gives the maximum loglikelihood

Note

The optimization may get trapped in local optima. Try different starting values to search for the global optimum.

Author(s)

Rampal S. Etienne & Bart Haegeman

References

- Etienne & Haegeman 2012. In review.
- Etienne et al. 2012, Proc. Roy. Soc. B 279: 1300-1309, doi: 10.1098/rspb.2011.1439

See Also

[dd_SR_loglik](#), [dd_ML](#), [dd_KI_ML](#), [optimx](#)

Examples

```
cat("Estimating parameters for a set of branching times brts with the default settings:")  
brts = 1:20  
dd_SR_ML(brts)
```

Index

*Topic **models**

dd_KI_loglik, [3](#)

dd_KI_ML, [5](#)

dd_loglik, [7](#)

dd_ML, [8](#)

dd_SR_loglik, [10](#)

dd_SR_ML, [11](#)

DDD-package, [2](#)

dd_KI_loglik, [3](#), [6](#), [7](#), [11](#)

dd_KI_ML, [4](#), [5](#), [9](#), [12](#)

dd_loglik, [4](#), [7](#), [9](#), [11](#)

dd_ML, [6](#), [7](#), [8](#), [12](#)

dd_SR_loglik, [4](#), [7](#), [10](#), [12](#)

dd_SR_ML, [6](#), [9](#), [11](#), [11](#)

DDD (DDD-package), [2](#)

DDD-package, [2](#)

optimx, [6](#), [9](#), [12](#)