

The ouch Package

February 16, 2008

Version 1.2-4

Date 2007-05-08

Title Ornstein-Uhlenbeck models for phylogenetic comparative hypotheses

Author Aaron A. King <kingaa@umich.edu> and Marguerite A. Butler <mabutler@utk.edu>

Maintainer Aaron A. King <kingaa@umich.edu>

Description Fit and compare Ornstein-Uhlenbeck models for evolution along a phylogenetic tree.

Depends R(>= 1.9.1)

License GPL version 2 or later

R topics documented:

anolis.ssd	1
bimac	3
brown.fit	5
hansen.fit	6
tree.plot	8
Index	10

anolis.ssd	<i>Greater Antillean anolis lizard sexual size dimorphism data.</i>
------------	---

Description

The dataset consists of sexual size-dimorphism data for 38 species of anoles from Cuba, Hispaniola, Jamaica, and Puerto Rico (Butler, Schoener, and Losos 2000). Each of these species belongs to one of six microhabitat types, or “ecomorphs” (sensu Williams, 1972): trunk-ground, grass-bush, trunk, trunk-crown, twig, and crown-giant. The data were used to demonstrate an evolutionary association between habitat type and degree of sexual size dimorphism.

Usage

```
data(anolis.ssd)
```

Format

A data frame with 38 observations on the following 6 variables.

node Labels for the nodes.

species Names of extant species.

log.SSD Log sexual size dimorphism of extant species.

ancestor Ancestor node.

time Time of node.

OU.1 a factor with levels `ns`

OU.7 a factor with levels corresponding to ecomorph (`tg tc gb cg tw tr anc`)

Details

Size dimorphism was calculated as the log-ratio of male snout-to-vent length to female snout-to-vent length. (males are larger).

In this example, we tested three models of evolution: Brownian motion, Ornstein-Uhlenbeck with one global optimum, and Ornstein-Uhlenbeck with 7 optima (one for each ecomorph type plus an additional one for an “unknown” type).

For the 7-optima model, we assigned each terminal branch to an optimum according to the ecomorph type of the extant species. Because we had no information to help guide hypotheses about internal branches, we assigned internal branches to the “unknown” selective regime. The phylogeny of these species is consistent with and adaptive radiation, with a burst of speciation events early in the evolutionary history of this clade (see phylogeny in Butler & King (2004) or execute the following commands:

```
data(anolis.ssd)
attach(anolis.ssd)
tree.plot(ancestor, time, species, OU.7)
```

Author(s)

Marguerite A. Butler <mabutler at utk dot edu> and Aaron A. King <kingaa at umich dot edu>

Source

Butler, M.A. and A.A. King (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *American Naturalist*, in press.

References

- Butler, M. A., T. W. Schoener, and J. B. Losos. 2000. The relationship between sexual size dimorphism and habitat use in Greater Antillean Anolis lizards. *Evolution*, 54:259-272.
- Williams, E. E. 1972. The origin of faunas. Evolution of lizard congeners in a complex island fauna: a trial analysis. *Evol. Biol.*, 6:47-89.

Examples

```

data(anolis.ssd)
attach(anolis.ssd)
tree.plot(node, ancestor, time, species, OU.7)
brown.fit(log.SSD, node, ancestor, time)
hansen.fit(log.SSD, node, ancestor, time, OU.1) # the root character state is estimated
hansen.fit(log.SSD, node, ancestor, time) # stationary state assumption
hansen.fit(log.SSD, node, ancestor, time, OU.7)
detach(anolis.ssd)

```

bimac

Anolis bimaculatus lizard size data.

Description

This is the *Anolis bimaculatus* dataset used in Butler & King 2004. It is used to test a hypothesis of character displacement using an interspecific dataset of body sizes and current data on sympatry/allopatry. The data frame consists of the following columns: `species` which are species names, `size` which is the phenotypic data, and the variables `ancestor` and `time` which specify the topology of the phylogeny and the location of the nodes in time, respectively. The columns `OU.1`, `OU.3`, `OU.4`, and `OU.LP` specify the hypothesized selective regimes (there are four alternatives in this example). Explanations of the data are given below. Mean species lengths.

Usage

```
data(bimac)
```

Format

A data frame with 45 observations on the following 8 variables.

node Labels for the nodes.

species Species names for extant species.

size Body size (head length in mm) of extant species.

ancestor Ancestral node.

time Time of node.

OU.1 a factor with levels `ns`

OU.3 a factor with levels `small medium large`

OU.4 a factor with levels `small medium large anc`

OU.LP a factor with levels `small medium large`

Details

Body size. We use the phenotypic data and phylogeny of Losos (1990), which employed the head lengths (of males) as a proxy for body size. In this group of lizards, head length correlates very strongly with snout-to-vent length and the cube root of mass, which are standard measures of body size. The data are head lengths in mm, note that we use the log of this value in analyses.

Tree topology The tree topology is encoded via two vectors: `ancestor` and `time`. Each node of the phylogenetic tree has a corresponding row in the data frame, numbered from 1 to 45. The columns `ancestor` and `time` specify the phylogeny. The `ancestor` variable specifies the topology: it is a list indicating the ancestor of each node. The root node has ancestor 0. The variable `time` specifies the temporal location of each node, with the root node being at time 0.

Specifications of selective regimes. (Columns `OU.1`, `OU.3`, `OU.4`, `OU.LP`). These columns are factors, the levels of which correspond to the “paintings” of the respective adaptive regime hypotheses onto the phylogeny. Each selective regime is named (small, medium, large, etc.). Put the corresponding name on each branch segment to indicate which selective regime it belongs to. Each column corresponds to a different painting of the selective regimes, and thus to a different hypothesis. In this example, there are 3 alternative models (see Butler & King 2004): `OU.4` is 4-regime model, `OU.3` is 3-regime model (all ancestors are medium), `OU.LP` is linear parsimony model.

Author(s)

Marguerite A. Butler <mabutler at utk dot edu> and Aaron A. King <kingaa at umich dot edu>

Source

Butler, M.A. and A.A. King (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *American Naturalist*, in press.

References

- Lazell, J. D. 1972. The anoles (Sauria: Iguanidae) of the Lesser Antilles. *Bull. Mus. Comp. Zool.*, 143:1-115.
- Losos, J. B. 1990. A phylogenetic analysis of character displacement in Caribbean Anolis lizards. *Evolution*, 44:558-569.

Examples

```
data(bimac)
attach(bimac)
tree.plot(node,ancestor,time,names=species,regimes=OU.LP)
brown.fit(log(size),node,ancestor,time/max(time))
hansen.fit(log(size),node,ancestor,time/max(time))
## a simple bootstrap:
sapply(hansen.dev(n=5,node,ancestor,time/max(time),alpha=0.19,sigma=0.22,theta=3.0),
       function(x)unlist(hansen.fit(x,node,ancestor,time/max(time))))
sapply(brown.dev(n=5,node,ancestor,time/max(time),sigma=0.22,theta=3),
       function(x)unlist(hansen.fit(x,node,ancestor,time/max(time))))
```

```

hansen.fit(log(size), node, ancestor, time/max(time), OU.3)
hansen.fit(log(size), node, ancestor, time/max(time), OU.4)
hansen.fit(log(size), node, ancestor, time/max(time), OU.LP)
detach(bimac)

```

brown.fit

Brownian-motion model of evolution along a phylogenetic tree

Description

These functions relate to the Brownian motion model for phylogenetic evolution.

`brown` fits the parameters σ and θ of this model to given data.

`brown.dev` generates simulated data sets.

Usage

```

brown.fit(data, node, ancestor, times)
brown.dev(n = 1, node, ancestor, times, sigma, theta)

```

Arguments

<code>data</code>	Phenotypic data for extant species, i.e., at the terminal ends of the phylogenetic tree.
<code>node</code>	Specification of the names of the nodes.
<code>ancestor</code>	Specification of the topology of the phylogenetic tree. This is in the form of a character vector of node names, one for each node in the tree. The <i>i</i> -th name is that of the ancestor of the <i>i</i> -th node. The root node is distinguished by having no ancestor (i.e., NA).
<code>times</code>	A vector of nonnegative numbers, one per node in the tree, specifying the time at which each node is located. The root node should be assigned time 0.
<code>n</code>	the number of simulated data sets to generate.
<code>sigma</code>	the value of σ to be used in the simulations.
<code>theta</code>	the value of θ to be used in the simulations.

Value

`brown` returns a list of the following elements:

<code>sigma</code>	Maximum likelihood estimate of σ .
<code>theta</code>	Maximum likelihood estimate of θ .
<code>loglik</code>	Log likelihood.
<code>deviance</code>	-2 loglik.
<code>aic</code>	Akaike information criterion.

sic Schwartz information criterion (=BIC)

df Number of parameters estimated (= 2).

brown.dev returns a list of n simulated data sets. Each data set corresponds exactly to the data used in the call to brown.fit.

Author(s)

Aaron A. King <kingaa at umich dot edu>

References

Butler, M.A. and A.A. King (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. American Naturalist, in press.

hansen.fit

Hansen model of evolution along a phylogenetic tree

Description

These functions relate to the Hansen model for phylogenetic evolution.

hansen.fit fits the parameters α , σ , and θ of this model.

hansen.prof computes the profile likelihood of the α parameter.

hansen.dev generates random deviates from the Hansen model.

Usage

```
hansen.fit(data, node, ancestor, times, regimes = NULL,
           interval = c(0, 100), tol = 1e-12)
hansen.prof(alpha, data, node, ancestor, times, regimes = NULL)
hansen.dev(n = 1, node, ancestor, times, regimes = NULL,
           alpha, sigma, theta)
```

Arguments

data	Phenotypic data for extant species, i.e., at the terminal ends of the phylogenetic tree.
node	Character vector of names of the nodes.
ancestor	Specification of the topology of the phylogenetic tree. This is in the form of a character vector of node names, one for each node in the tree. The i -th name is that of the ancestor of the i -th node. The root node is distinguished by having no ancestor (i.e., NA).
times	A vector of nonnegative numbers, one per node in the tree, specifying the time at which each node is located. The root node should be assigned time 0.

regimes	A vector of codes, one for each node in the tree, specifying the selective regimes hypothesized to have been operative. Corresponding to each node, enter the code of the regime hypothesized for the branch segment terminating in that node. For the root node, because it has no branch segment terminating on it, the regime specification is irrelevant.
interval	The interval which will be searched for the optimal α . By default, $0.001 < \alpha < 20$.
tol	Convergence tolerance.
n	Number of pseudorandom data sets to generate.
alpha	Value of α to use.
sigma	Value of σ to use.
theta	Value of θ to use.

Details

The vector `regimes` should be of class factor. If `regimes` is unspecified or NULL, all lineages are assumed to be evolving under a single, global OU process with a global optimum. In this case, rather than estimate the character state at the root node, the algorithm assumes that the character state at the root value follows the stationary distribution for the OU process. In general, it is impossible to identify both the root character state and the global optimum using contemporaneous data.

Value

`hansen.fit` returns a list containing the following elements:

alpha	Maximum likelihood estimate of α . Note that if α lies against one of the constraints (see <code>interval</code> above), then this is not a maximum-likelihood estimate.
sigma	Maximum likelihood estimate of σ .
theta	Maximum likelihood estimate of θ .
loglik	Log likelihood.
deviance	-2 loglik.
aic	Akaike information criterion.
sic	Schwartz information criterion (=BIC)
df	Number of parameters estimated (= 3 + number of regimes).
alpha	the specified <code>alpha</code>
loglik	the log likelihood
deviance	the deviance ($-2 \log L$)
aic	the Akaike information criterion value
sic	the Schwartz information criterion value

Note that when $\alpha = 0$ exactly, the computed log likelihood does not agree with the Brownian motion model.

`hansen.dev` returns a list of `n` simulated data sets. Each data set corresponds exactly to the data used in the call to `hansen.fit`.

Author(s)

Aaron A. King <kingaa at umich dot edu>

References

Butler, M.A. and A.A. King (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *American Naturalist* 164:683-695, 2004.

tree.plot

Validating and plotting phylogenetic trees.

Description

Validate or plot the phylogenetic tree in ouch format.

Usage

```
is.valid.ouch.tree(node, ancestor, times, regimes = NULL)
tree.plot(node, ancestor, times, names = NULL, regimes = NULL)
```

Arguments

node	A character vector giving the name of each node.
ancestor	Specification of the topology of the phylogenetic tree. This is in the form of a character vector naming the immediate ancestor of each node. In particular, the i-th name is that of the ancestor of the i-th node. The root node is distinguished by having no ancestor (i.e., NA).
times	A vector of nonnegative numbers, one per node in the tree, specifying the time at which each node is located. The root node should be assigned time 0.
names	Optional vector of species names.
regimes	A vector of codes, one for each node in the tree, specifying the selective regimes hypothesized to have been operative. Corresponding to each node, enter the code of the regime hypothesized for the branch segment terminating in that node. For the root node, because it has no branch segment terminating on it, the regime specification is irrelevant.

Details

tree.plot makes a simple plot of the phylogenetic tree. Labels and/or a selective-regime-based coloring scheme are optional.

is.valid.ouch.tree performs several checks to make sure that the tree is valid. These include checks to ensure that all the vector arguments are of the same length, that the nodes have unique names, that there is a unique root, that there is at least one terminal node, that every node's ancestor is in fact part of the tree, that there are no cycles in the 'tree', and that the tree is connected. It returns TRUE if the tree is valid, FALSE otherwise and gives diagnostic warnings.

Author(s)

Aaron A. King <kingaa at umich dot edu>

Examples

```
data(bimac)
attach(bimac)
is.valid.ouch.tree(node, ancestor, time, OU.LP)
tree.plot(node, ancestor, time, species, OU.LP)
```

Index

*Topic **hplot**

tree.plot, 8

*Topic **models**

anolis.ssd, 1

bimac, 3

brown.fit, 5

hansen.fit, 6

anolis.ssd, 1

bimac, 3

brown.dev(*brown.fit*), 5

brown.fit, 5

hansen.dev(*hansen.fit*), 6

hansen.fit, 6

hansen.prof(*hansen.fit*), 6

is.valid.ouch.tree(*tree.plot*), 8

tree.plot, 8