

Package ‘kdetrees’

February 20, 2015

Type Package

Title Nonparametric method for identifying discordant phylogenetic trees

Version 0.1.5

Date 2014-05-21

Author Grady Weyenberg and Ruriko Yoshida

Maintainer Grady Weyenberg <grady.weyenber@uky.edu>

Description A non-parametric method for identifying potential outlying observations in a collection of phylogenetic trees based on the methods of Owen and Provan (2011). Such discordant trees may indicate problems with sequence annotation or tree reconstruction, or they may represent interesting biological phenomena, such as horizontal gene transfers.

License GPL-2

Depends R (>= 2.15.1)

Imports ape, distory, ggplot2

URL <http://github.com/grady/kdetrees>

LazyData TRUE

NeedsCompilation no

Repository CRAN

Date/Publication 2014-05-30 01:09:11

R topics documented:

kdetrees-package	2
apicomplexa	2
as.data.frame.kdetrees	2
as.matrix.multiPhylo	3
bw.nn	4
dist.diss	4
estimate	5

hist.kdetrees	6
kdetrees	6
kdetrees.complete	7
normkern	8
plot.kdetrees	9
print.kdetrees	10

Index **11**

kdetrees-package	<i>kdetrees</i>
------------------	-----------------

Description

kdetrees

apicomplexa	<i>Apicomplexa gene trees sample data set.</i>
-------------	--

Description

Apicomplexa gene trees sample data set.

Format

a multiPhylo object with 268 phylogenetic trees

Examples

kdetrees(apicomplexa)

as.data.frame.kdetrees	<i>Convert kdetrees object to data.frame</i>
------------------------	--

Description

Converts a kdetrees object to a data.frame suitable for saving as output. It contains the density estimates for each tree, a Boolean value indicating if the tree was selected as an outlier, and optionally the newick string corresponding to the tree.

Usage

```
## S3 method for class 'kdetrees'
as.data.frame(x, row.names, optional, trees = NULL, ...)
```

Arguments

x	kdetrees object to be converted
row.names	ignored
optional	ignored
trees	If given the original list of trees, will convert to newick and add a column to the output
...	unused

Value

a data.frame

Author(s)

Grady Weyenberg

Examples

```
result <- kdetrees(apicomplexa)
as.data.frame(result)
```

as.matrix.multiPhylo *dissimilarity map tree vectorization*

Description

Dissimilarity maps convert trees to vectors using tip-to-tip path lengths. Branch length information may be optionally discarded (the default), resulting in vectors based solely on tree topology.

Usage

```
## S3 method for class 'multiPhylo'
as.matrix(x, ...)
```

Arguments

x	an ape::multiPhylo object.
...	additional options for ape::cophenetic.phylo

Value

a row matrix of tree vectors

Author(s)

Grady Weyenberg

Examples

```
as.matrix(apicomplexa[1:5])
```

bw.nn	<i>nearest-neighbor adaptive bandwidth selection</i>
-------	--

Description

For each row in pairwise distance matrix find the distance to the closest prop fraction of trees.

Usage

```
bw.nn(x, prop = 0.2, tol = 1e-06)
```

Arguments

x	pairwise distance matrix
prop	fraction of data to define the local neighborhood
tol	tolerance for zero-bandwidth check

Value

a vector of bandwidths for each tree (row) in x

Author(s)

Grady Weyenberg

Examples

```
dm <- as.matrix(dist.diss(apicomplexa[1:20]))
bw.nn(dm)
```

dist.diss	<i>Compute pairwise tree distances</i>
-----------	--

Description

Compute pairwise tree distances

Usage

```
dist.diss(x, ..., method = "euclidean", p = 2)
```

Arguments

x either a row matrix of tree vectors, or a multiPhylo object
 ... additional arguments passed to as.matrix.multiPhylo
 method option passed to dist
 p option passed to dist

Value

a dist object with tree-to-tree distances

Author(s)

Grady Weyenberg

See Also

dist

Examples

```
dist.diss(apicomplexa[1:5])
```

estimate	<i>estimate densities from kernel matrix</i>
----------	--

Description

estimate densities from kernel matrix

Usage

```
estimate(x, i = integer())
```

Arguments

x matrix of kernel contributions
 i vector of columns to exclude from calculation

Value

vector of density estimates for each tree

Author(s)

Grady Weyenberg

hist.kdtrees *Create a histogram of tree density estimates*

Description

Create a histogram of tree density estimates

Usage

```
## S3 method for class 'kdtrees'  
hist(x, ...)
```

Arguments

x kdtrees object to plot
... additional arguments passed to ggplot

Value

a ggplot object

Author(s)

Grady Weyenberg

Examples

```
result <- kdtrees(apicomplexa)  
hist(result)
```

kdtrees *Identify discordant trees in a sample*

Description

Analyze a set of phylogenetic trees and attempt to identify trees which are significantly discordant with other trees in the sample (outlier trees).

Usage

```
kdtrees(trees, k = 1.5, distance = c("geodesic", "dissimilarity"),  
      outgroup = NULL, topo.only = FALSE, bw = list(), greedy = FALSE, ...)
```

Arguments

trees	multiPhylo object
k	IQR multiplier for outlier detection
distance	Select "geodesic" or "dissimilarity" distance calculation method
outgroup	if a character, reroot all trees with this species as outgroup. The geodesic distance method requires rooted trees.
topo.only	set all branch lengths to 1 before analyzing?
bw	see Details
greedy	greedy outlier detection?
...	additional arguments for distance calculation function, see details

Details

If bw is a single number, it will be used as a single constant bandwidth. It can also be a vector, in which case it will be used as variable bandwidths for each tree, respectively. Finally, if it is a list (default), the list will be passed as arguments to the bw.nn adaptive bandwidth function.

... Is passed to either `distory::dist.multiPhylo` or `dist.diss`, as appropriate. See the help for these functions for more details.

Value

a kdtrees object; list(density,outliers)

Author(s)

Grady Weyenberg

Examples

```
kdeobj <- kdtrees(apicomplexa)
print(kdeobj)
kdeobj$outliers

kdtrees(apicomplexa, k=2.0, distance="dissimilarity",topo.only=FALSE)
```

kdtrees.complete

Complete kdtrees analysis convenience function

Description

Performs a complete kdtrees analysis, starting with reading trees from a newick file on disk, and writing result files to the working directory. Names and location of output files may be controlled by optional arguments.

Usage

```
kdtrees.complete(infile, ..., treeoutfile = "outliers.tre",
  csvfile = "results.csv", plotfile = "plot.png", histfile = "hist.png")
```

Arguments

infile	newick file with trees
...	additional parameters for kdtrees
treeoutfile	write outlier trees in newick format to this file
csvfile	write density results to this file
plotfile	print scatterplot of results to this file
histfile	print histogram of density estimates to this file

Value

result of kdtrees call

Author(s)

Grady Weyenberg

normkern

Generalized Gaussian kernel

Description

The un-normalized Gaussian kernel function: $\exp(-(\text{abs}(x/\text{bw}))^\Delta)/\text{bw}$

Usage

```
normkern(x, bw = 1, delta = 2L)
```

Arguments

x	places to evaluate kernel
bw	bandwidth values
delta	shape parameter for kernel

Details

The bandwidth parameter may be used in any way that makes sense in the above R expression. In particular, it may be a single value, for a constant bandwidth, or a vector, with each element corresponding the bandwidth of the kernel to be placed at each respective observation.

Value

an object of the same type as *x* with the kernel evaluations

Author(s)

Grady Weyenberg

plot.kdtrees *Plot the unnormalized density estimates for each tree.*

Description

Plot the unnormalized density estimates for each tree.

Usage

```
## S3 method for class 'kdtrees'  
plot(x, ...)
```

Arguments

x *kdtrees* object to be plotted
... additional arguments passed to *ggplot*

Value

a *ggplot* object

Author(s)

Grady Weyenberg

Examples

```
result <- kdtrees(apicomplexa)  
plot(result)
```

print.kdetrees	<i>Summarize a kdetrees object in human-readable form.</i>
----------------	--

Description

Pretty-prints the results of a kdetrees ananalysis to console.

Usage

```
## S3 method for class 'kdetrees'  
print(x, ...)
```

Arguments

x	object to be printed
...	unused, required for generic compatability

Value

invisible(x)

Author(s)

Grady Weyenberg

Index

`apicomplexa`, [2](#)
`as.data.frame.kdetrees`, [2](#)
`as.matrix.multiPhylo`, [3](#)

`bw.nn`, [4](#)

`dist.diss`, [4](#)

`estimate`, [5](#)

`hist.kdetrees`, [6](#)

`kdetrees`, [6](#)
`kdetrees-package`, [2](#)
`kdetrees.complete`, [7](#)

`normkern`, [8](#)

`plot.kdetrees`, [9](#)
`print.kdetrees`, [10](#)