Package ‘Apoderoides’

January 14, 2024

Type Package

Title Prioritize and Delete Erroneous Taxa in a Large Phylogenetic Tree

Version 1.0.1

Date 2024-01-12

Description Finds, prioritizes and deletes erroneous taxa in a phylogenetic tree. This package calculates scores for taxa in a tree. Higher score means the taxon is more erroneous. If the score is zero for a taxon, the taxon is not erroneous. This package also can remove all erroneous taxa automatically by iterating score calculation and pruning taxa with the highest score.

License MIT + file LICENSE

Encoding UTF-8

LazyData true


Depends R (>= 3.5.0)
Imports ape, Rcpp, RcppProgress
LinkingTo Rcpp, RcppProgress
Suggests knitr, rmarkdown
VignetteBuilder knitr

NeedsCompilation yes

Author Satoshi Aoki [aut, cph, cre]

Maintainer Satoshi Aoki <aokis1ll11@gmail.com>

Repository CRAN

Date/Publication 2024-01-14 15:20:14 UTC
R topics documented:

- **Apoderoides-internal**
- **autoDeletion**
- **calc.Score**
- **deleteAnomaly**
- **get.upperRank**
- **getAllCentroids**
- **testRankList**
- **testTree**

Index

---

**Apoderoides-internal**  *Internal Apoderoides Functions*

---

**Description**

Internal Apoderoides functions

**Details**

These are not to be called by the user.

**Value**

Different values, depending on the function.

---

**autoDeletion**  *autoDeletion*

---

**Description**

Iterate calc.Score() and DeleteAnomaly() until all the tree tips have 0 score or the number of the tips becomes three or lower.

**Usage**

autoDeletion(tree, OTUrankData=NULL, show_progress=TRUE, num_threads=1)
calc.Score

Arguments

tree A phylogenetic tree to be checked. This is loaded by ape::read.tree() from a file.

OTUrankData A list composed of two character vectors. The first vector is tips of tree. The second vector is the upper rank of the tips. When this is NULL, the function assumes that all the tree tips are expressed as Genus_species like Homo_sapience, and calculates for genera. When this is not NULL, the function calculates based on the upper rank in this list.

show_progress If TRUE, calculation progress is shown on the R console.

num_threads A positive integer to specify the number of threads to calculate.

Value

A list of the length three. The first element is a list of phylogenetic tree from which erroneous taxa are deleted. The second is a character vector of deleted taxa. The third is a list of lists showing the transition of the score. See calc.Score about the contents of the third element.

Examples

data(testTree)
data(testRankList)
#calculate scores for the rank in the list, and delete all the erroneous tips
#this takes tens of seconds for calculation
result<-autoDeletion(testTree,testRankList)
#tree without erroneous tips
result[[1]]
#deleted tips
result[[2]]
#scores during iteration of score calculation and tip deletion
result[[3]]
Arguments

tree A phylogenetic tree to be checked. This is loaded by ape::read.tree() from a file.

OTUrankData A list composed of two character vectors. The first vector is tips of tree. The second vector is the upper rank of the tips. When this is NULL, the function assumes that all the tree tips are expressed as Genus_species like Homo_sapience, and calculate the score for genera. When this is not NULL, the function returns scores based on the upper rank in this list.

allRankNames This can be omitted. This is a unique character vector of the upper ranks of the tree tips. If given, the calculation will be a little faster.

allCentroids This can be omitted. This is a list of numeric vectors of the centroids of ranks. If given, the calculation will be a little faster.

dropIndex This can be omitted. A numeric vector of indices of tree tips. The tree tips indicated by this dropIndex will be removed from the score calculation.

sort If TRUE, the calculation result is sorted by descending order of the total score.

show_progress If TRUE, calculation progress is shown on the R console.

num_threads A positive integer to specify the number of threads to calculate the scores.

Value

A matrix of characters. The following explains the columns in the matrix.

<table>
<thead>
<tr>
<th>OTU</th>
</tr>
</thead>
<tbody>
<tr>
<td>The name of tree tip.</td>
</tr>
<tr>
<td>perCladeOTUScore</td>
</tr>
<tr>
<td>The final score calculated by &quot;sum&quot; divided by the number of OTUs with the same &quot;#clade&quot;.</td>
</tr>
<tr>
<td>sum</td>
</tr>
<tr>
<td>The sum of &quot;intruder&quot; and &quot;outlier&quot; for the OTU.</td>
</tr>
<tr>
<td>intruder</td>
</tr>
<tr>
<td>The intruder score showing how many ranks the OTU intruding into.</td>
</tr>
<tr>
<td>outlier</td>
</tr>
<tr>
<td>The outlier score showing how the OTU is far away from the core clade of the belonging rank.</td>
</tr>
<tr>
<td>#clade</td>
</tr>
<tr>
<td>The clade number. Monophyletic OTUs with the same rank has the same #clade.</td>
</tr>
</tbody>
</table>

Examples

data(testTree)
#calculate scores for genus
calc.Score(testTree)
data(testRankList)
#calculate scores for the rank in the list
calc.Score(testTree,testRankList)
deleteAnomaly
deleteAnomaly

Description
Delete tip(s) with the highest score from a tree.

Usage
deleteAnomaly(tree, score, OTUrankData = NULL, drop = FALSE)

Arguments
- **tree**: A phylogenetic tree to be checked. This is loaded by ape::read.tree() from a file.
- **score**: Score calculated by calc.Score function.
- **OTUrankData**: A list composed of two character vectors. The first vector is tips of tree. The second vector is the upper rank of the tips. When this is NULL, the function assumes that all the tree tips are expressed as Genus_species like Homo_sapiens and that the score is calculated based on genera. When this is not NULL, the function assumes the score is calculated based on the upper rank in this list.
- **drop**: Whether the dropped OTU(s) is included in the returned tree.

Value
A list of the length two. The first element is a vector of characters of deleted tip label(s). The second is a list of a phylogenetic tree without the deleted tip(s).

Examples
```r
data(testTree)
data(testRankList)
#calculate scores for the rank in the list
score<-calc.Score(testTree, testRankList)
#delete tip with the highest score from tree
deleteAnomaly(testTree, score, testRankList)
```

get.upperRank
get.upperRank

Description
Obtain upper rank of scientific names in data. When OTUrankData is not provided, this function returns genus names assuming the elements in data are scientific names connected by underlines like "Homo_sapiens". When OTUrankData is provided, this function searches data in OTUrankData[[1]] and returns OTUrankData[[2]] of the corresponding index.
getAllCentroids

Usage

get.upperRank(data, OTUrankData=NULL)

Arguments

data A vector of characters.

OTUrankData A list composed of two character vectors. The first vector is tips of tree. The second vector is the upper rank of the tips. When this is NULL, the function assumes that all the tree tips are expressed as Genus_species like Homo_sapience, and calculate the score for genera. When this is not NULL, the function returns scores based on the upper rank in this list.

Value

A vector of characters of upper rank.

Examples

#obtain genus name
generateRanks(c("Oxalis_nipponica", "Homo_sapiens"))
data(testTree)
data(testRankList)
#obtain higher rank names
generateRanks(testTree$tip[1:3], testRankList)

getAllCentroids

Description

Calculate all the centroids of ranks in the tree.

Usage

getAllCentroids(tree, OTUrankData=NULL, show_progress=FALSE, num_threads=1)

Arguments

tree A phylogenetic tree to be checked. This is loaded by ape::read.tree() from a file.

OTUrankData A list composed of two character vectors. The first vector is tips of tree. The second vector is the upper rank of the tips. When this is NULL, the function assumes that all the tree tips are expressed as Genus_species like Homo_sapience, and calculate the centroids for genera. When this is not NULL, the function returns centroids based on the upper rank in this list.

show_progress If TRUE, calculation progress is shown on the R console.

num_threads A positive integer to specify the number of threads to calculate the scores.
Value

A list containing vectors of integers of centroid node number(s).

Examples

data(testTree)
#calculate centroids for genus
getAllCentroids(testTree)
data(testRankList)
#calculate centroids for the rank in the list
getAllCentroids(testTree,testRankList)

description

Example data to test Apoderoides. testRankList is a list of two elements. The first element is the tip label of testTree, and the second element is corresponding family names of the tips.

Usage

data(testRankList)

description

Example data to test Apoderoides. testTree is a tree of land plants based on chlB gene.

Usage

data(testTree)
Index

* datasets
  testRankList, 7
  testTree, 7
* ts
  Apoderoides-internal, 2
  autoDeletion, 2
  calc.Score, 3, 3
  calcIntScore_C(Apoderoides-internal), 2
  calcOutScore_C(Apoderoides-internal), 2
  deleteAnomaly, 5
  extractOTUbyRankName_C(Apoderoides-internal), 2
  findSubTips_C(Apoderoides-internal), 2
  findUpperNode(Apoderoides-internal), 2
  get.upperRank, 5
  getAllCentroids, 6
  getAllCentroids_C(Apoderoides-internal), 2
  getAllRankNames(Apoderoides-internal), 2
  getRankCentroid_C(Apoderoides-internal), 2
  getTipNo_C(Apoderoides-internal), 2
  is.monophyleticByRank(Apoderoides-internal), 2
  testRankList, 7
  testTree, 7