

Package ‘TBRDist’

July 7, 2020

Title Rearrangement Distances Between Unrooted Phylogenetic Trees

Version 1.0.1

License GPL (>= 3)

Copyright Underlying C++ code copyright 2018 Chris Whidden.

Description Fast calculation of the Subtree Prune and Regraft (SPR), Tree Bisection and Reconnection (TBR) and Replug distances between unrooted trees, using the algorithms of Whidden and Matsen (2017) <arxiv:1511.07529>.

URL <https://ms609.github.io/TBRDist/>,
<https://github.com/ms609/TBRDist/>,
<https://github.com/cwhidden/uspr/>

BugReports <https://github.com/ms609/TBRDist/issues/>

Depends R (>= 3.4.0)

Imports ape, TreeDist, TreeTools (>= 1.1.0)

Suggests knitr, Rdpack, rmarkdown, testthat

SystemRequirements C++11

RdMacros Rdpack

LazyData true

ByteCompile true

Encoding UTF-8

X-schema.org-keywords phylogenetics, tree-distance

Language en-GB

LinkingTo BH, Rcpp

VignetteBuilder knitr

RoxygenNote 7.1.1

NeedsCompilation yes

Author Martin R. Smith [aut, cre, cph]
(<<https://orcid.org/0000-0001-5660-1727>>),
Chris Whidden [cph]

Maintainer Martin R. Smith <martin.smith@durham.ac.uk>

Repository CRAN

Date/Publication 2020-07-07 16:00:09 UTC

R topics documented:

TreeRearrangementDistances	2
Index	6

TreeRearrangementDistances

Calculate SPR, TBR and Replug distances on unrooted trees

Description

Calculate SPR, TBR and Replug distances on unrooted trees, and the information content of the maximum agreement forest.

Usage

```
USPRDist(
  tree1,
  tree2 = NULL,
  allPairs = is.null(tree2),
  checks = TRUE,
  useTbrApproxEstimate = TRUE,
  useTbrEstimate = TRUE,
  useReplugEstimate = TRUE
)
```

```
ReplugDist(
  tree1,
  tree2 = NULL,
  allPairs = is.null(tree2),
  checks = TRUE,
  maf = FALSE
)
```

```
TBRDist(
  tree1,
  tree2 = NULL,
  allPairs = is.null(tree2),
  checks = TRUE,
  maf = FALSE,
  countMafs = FALSE,
  printMafs = FALSE,
```

```

    exact = maf,
    approximate = !exact,
    optimize = TRUE,
    protectB = TRUE
  )

MAFInfo(tree1, tree2 = tree1, exact = FALSE)

```

Arguments

tree1, tree2	Trees of class <code>phylo</code> , or lists thereof.
allPairs	Logical; if TRUE, compare each tree in tree1 with each tree in tree2; if FALSE, compare each tree in tree1 only with the tree at the corresponding index in tree2. If tree2 is not specified, each tree in tree1 will be compared with each other tree in tree1.
checks	Logical specifying whether to check that trees are properly formatted and labelled. Specify FALSE at your peril, as improper input is likely to crash R.
useTbrApproxEstimate, useTbrEstimate, useReplugEstimate	Logical specifying whether to use approximate TBR distance, TBR distance or Replug distance to help estimate the SPR distance.
maf	Logical specifying whether to report a maximum agreement forest corresponding to the optimal score.
countMafs	Logical specifying whether to count the number of maximum agreement forests found.
printMafs	Logical specifying whether to print maximum agreement forests to stdout whilst counting. Use <code>capture.output(TBRDist(tree1, tree2, printMafs = TRUE))</code> to access these in R.
exact	Logical specifying whether to calculate the exact TBR distance.
approximate	Logical specifying whether to calculate the approximate TBR distance. By default, is set to the opposite of exact; either approximate or exact should usually be set to TRUE if a distance is required.
optimize	Logical specifying whether to use the default optimizations.
protectB	Logical specifying whether to use the 'PROTECT_B' optimization. Overrides value inherited from optimize.

Details

Note that these distances are NP-hard to compute, so the running time of the algorithms used in this software scale exponentially with the distance computed. The version of 'uspr' linked in this package is aimed at trees with up to 50 leaves and uSPR distances up to 14.

If you are interested in comparing rooted trees in terms of SPR operations, you should use 'rspr' instead. 'rspr' is also much more efficient and can easily handle pairs of binary rooted trees with 200+ leaves and distances > 50. rspr is not yet incorporated in this R package; please [contact the maintainer](#) if this would be useful to you.

Value

USPRDist() returns a vector of SPR distances between each pair of unrooted trees.

ReplugDist() returns a vector of Replug distances between each pair of trees, or (if maf = TRUE) a named list whose second and third elements list a vector of maximum agreement forests for each pair of trees.

TBRDist() returns a named list, each element of which bears a vector corresponding to the requested value for each tree pair. If only the exact value is requested (exact = TRUE), an unnamed vector of distances is returned.

MAFInfo() returns the information content of the maximum agreement forest, in bits. This is defined as the sum of the phylogenetic information content of each constituent subtree, plus the entropy of the clusters implied by the division of the tree into subtrees. Note that as there is no guarantee that the most informative MAF will be encountered, this measure is approximate only. exact will only serve to guarantee that a MAF corresponding to the exact TBR distance is among those sampled.

Author(s)

Algorithms implemented by Chris Whidden (cwhidden@fredhutch.org)

R wrappers by Martin R. Smith (martin.smith@durham.ac.uk)

References

If you use these functions in your research, please cite:

- Chris Whidden and Frederick A. Matsen IV. Calculating the Unrooted Subtree-Prune-and-Regraft Distance. arXiv:[1511.07529](https://arxiv.org/abs/1511.07529).

Examples

```
tree1 <- TreeTools::BalancedTree(6)
tree2 <- TreeTools::PectinateTree(6)

# SPR distance
USPRDist(tree1, tree2)

# Replug distance
ReplugDist(tree1, tree2)
ReplugDist(tree1, tree2, maf = TRUE)

# TBR distance between two trees
TBRDist(tree1, tree2, exact = TRUE)

# Compare a list against one tree, using approximate distances
TBRDist(list(tree1, tree2), tree2, exact = FALSE)

# Compare all pairs in two lists
TBRDist(list(tree1, tree2), list(tree1, tree2, tree2), allPairs = TRUE,
        exact = FALSE)
```

```
# Compare each tree in a list against each other
TBRDist(list(one = tree1, two = tree2, twoAgain = tree2))

# Compare each pair in two lists
TBRDist(list(tree1, tree2, tree2),
        list(tree2, tree1, tree2),
        exact = TRUE, approximate = TRUE, countMafs = TRUE)

# Capture maximum agreement forests
mafs <- capture.output(TBRDist(tree1, tree2, approximate = FALSE,
                              printMafs = TRUE))
head(mafs)

MAFInfo(tree1, tree2)
MAFInfo(list(tree2, tree1), list(tree1, tree2))
```

Index

`capture.output`, [3](#)

`MAFInfo (TreeRearrangementDistances)`, [2](#)

`ReplugDist`
`(TreeRearrangementDistances)`, [2](#)

`TBRDist (TreeRearrangementDistances)`, [2](#)

`TreeRearrangementDistances`, [2](#)

`USPRDist (TreeRearrangementDistances)`, [2](#)