Package 'TBRDist'

May 7, 2026

Title Rearrangement Distances Between Phylogenetic Trees

Version 2.0.0

License GPL (>= 3)

Copyright Underlying C++ code copyright 2009-2021 Chris Whidden.

Description Fast calculation of tree rearrangement distances.
For unrooted trees: Subtree Prune and Regraft (SPR), Tree Bisection and Reconnection (TBR), and Replug distances, using the algorithms of Whidden and Matsen (2017) <doi:10.48550/arXiv.1511.07529>.
For rooted trees: rooted SPR (rSPR) distance, using the fixed-parameter algorithms of Whidden, Beiko, and Zeh (2013) <doi:10.1137/110845045>.

URL `https://ms609.github.io/TBRDist/`,

`https://github.com/ms609/TBRDist/`,

`https://github.com/cwhidden/uspr/`,

`https://github.com/cwhidden/rspr/`

BugReports `https://github.com/ms609/TBRDist/issues/`

SystemRequirements C++17

Depends R (>= 3.6.0)

Imports ape, Rdpack, TreeDist, TreeTools (>= 2.2.0),

Suggests knitr, rmarkdown, testthat,

LinkingTo BH, Rcpp, TreeTools,

Config/Needs/check rcmdcheck, testthat

Config/Needs/coverage covr

Config/Needs/memcheck testthat

RdMacros Rdpack

VignetteBuilder knitr

ByteCompile true

Encoding UTF-8

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

Language en-GB

RoxygenNote 7.3.3

NeedsCompilation yes

Author Martin R. Smith [aut, cre, cph] (ORCID: <https://orcid.org/0000-0001-5660-1727>, GitHub: ms609), Chris Whidden [cph] (rspr, uspr, GitHub: cwhidden)

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

Repository CRAN

Date/Publication 2026-03-30 14:10:45 UTC

Contents

`RSPRDist` *Calculate rSPR distance between rooted trees*

Description

Calculate rooted Subtree Prune-and-Regraft (rSPR) distances between pairs of rooted binary trees using the algorithms of Whidden, Beiko & Zeh (2013).

Usage

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

Arguments

`tree1, tree2`	Trees of class `phylo`, or lists thereof. All trees must be rooted and bear identical sets of tip labels.
`allPairs`	Logical; if `TRUE`, compare each tree in `tree1` with each tree in `tree2`; if `FALSE`, compare corresponding pairs. Defaults to `TRUE` when `tree2` is not supplied.
`checks`	Logical; validate tree labels and dimensions before computation. Set `FALSE` at your peril—improper input is likely to crash R.

`approx`	Logical; if `TRUE` return the linear-time 3-approximation instead of the exact distance.
`maf`	Logical; if `TRUE` return the maximum agreement forest alongside the exact distance (implies `approx = FALSE`).

Details

This function wraps the **rspr** C++ library by Chris Whidden. It handles rooted trees and is substantially more efficient than `USPRDist()` for large trees: it can handle pairs with 200+ leaves and distances > 50.

Note that these distances are NP-hard to compute exactly; running time scales as O(2^k n) where k is the rSPR distance and n is the number of leaves. The built-in cluster decomposition (enabled by default) provides a large practical speedup. Use `approx = TRUE` for a guaranteed linear-time 3-approximation.

Input trees must be **rooted**. An error is raised if any tree is unrooted.

Value

By default, an integer vector of rSPR distances (one per tree pair), or a `dist` object when `allPairs = TRUE` and `tree2 = NULL`.

If `maf = TRUE`, a named list with elements:

`exact`	Integer vector of rSPR distances.
`maf_1`, `maf_2`	Character vectors giving the maximum agreement forest for each pair of trees, expressed as space-separated Newick components.

References

Whidden C, Beiko RG, Zeh N (2013). Fixed-Parameter Algorithms for Maximum Agreement Forests. *SIAM Journal on Computing* **42**:1431–1466. doi:10.1137/110845045

Whidden C, Zeh N, Beiko RG (2014). Supertrees based on the subtree prune-and-regraft distance. *Systematic Biology* **63**:566–581. doi:10.1093/sysbio/syu023

See Also

`USPRDist()` for unrooted trees.

Examples

```
library(ape)
set.seed(1)
tree1 <- rtree(8)
tree2 <- rtree(8)
tree1$tip.label <- tree2$tip.label <- paste0("t", 1:8)

RSPRDist(tree1, tree2)

# All pairwise distances among a list of trees
trees <- c(tree1, tree2)
```

```
RSPRDist(trees)

# Fast 3-approximation
RSPRDist(tree1, tree2, approx = TRUE)

# With maximum agreement forest
RSPRDist(tree1, tree2, maf = TRUE)
```

`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 `phylo`, 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 `capture.output(TBRDist(tree1, tree2, printMafs = TRUE))` 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.

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