Package 'rapidphylo'

July 23, 2025

Type Package

Title Rapidly Estimates Phylogeny from Large Allele Frequency Data Using Root Distances Method

Version 0.1.2

Author Arindam RoyChoudhury [aut, cre, cph],
Jing Peng [aut],
Ying Li [aut],
Laura Kubatko [aut, ths]

Maintainer Arindam RoyChoudhury <arr2014@med.cornell.edu>

Description Rapidly estimates tree-topology from large allele frequency data using Root Distances Method, under a Brownian Motion Model. See Peng et al. (2021) <doi:10.1016/j.ympev.2021.107142>.

License AGPL-3

URL https://github.com/ArindamRoyChoudhury/rapidphylo

BugReports https://github.com/ArindamRoyChoudhury/rapidphylo/issues

Depends R (>= 4.1.0)

Imports ape, phangorn, stats

Encoding UTF-8

LazyData true

LazyDataCompression xz

NeedsCompilation no

RoxygenNote 7.2.1

Repository CRAN

Date/Publication 2023-02-01 17:30:02 UTC

Contents

Human_Allele_Frequencies

Allele frequencies from 31,000 single nucleotide polymorphisms

Description

The dataset "Human_Allele_Frequencies" is a 5 × 31,000 matrix that contains allele frequencies from 31,000 single nucleotide polymorphisms in Chromosomes 1-10 in 5 human populations. The last population "San" is intended to be used as an outgroup. The allele frequencies have been compiled from ALFRED database at Yale University. The analysis from this dataset has been published in Peng et al 2021.

Usage

```
Human_Allele_Frequencies
```

Format

An object of class `matrix` (inherits from `array`) with 5 rows and 31000 columns.

RDM

Estimating tree-topology from allele frequency data

Description

`RDM()` estimates a tree-topology from allele frequencies.

Usage

```
RDM(
  mat_allele_freq,
  outgroup,
  use = c("complete.obs", "pairwise.complete.obs", "everything", "all.obs",
    "na.or.complete")
)
```

Arguments

`mat_allele_freq`	A $(P+1) \times L$ matrix containing the allele frequencies, where there are P taxa, plus one outgroup, and L loci.
`outgroup`	A variable that can be either the population name or a numerical row number of the outgroup data.
`use`	Specify which part of data is used to compute the covariance matrix. The options are "`complete.obs`", "`pairwise.complete.obs`", "`everything`", "`all.obs`", and "`na.or.complete`". See `stats::cov` for more details.

Details

The input matrix is the observed values of the frequencies at tips $1, 2, ..., P, P + 1$. A logit transformation is performed on the allele frequency data, so that the observed values are approximately normal. (The logit transformation of r refers to $\log \frac{r}{1-r}$.) The transformed matrix is converted into a data frame for further analyses.

Value

An estimated tree-topology in Newick format.

References

Peng J, Rajeevan H, Kubatko L, and RoyChoudhury A (2021) *A fast likelihood approach for estimation of large phylogenies from continuous trait data*. Molecular Phylogenetics and Evolution 161 107142.

Examples

```
# A dataset "Human_Allele_Frequencies" is loaded with the package;
# it has allele frequencies in 31,000 sites for
# 4 human populations and one outgroup human population.

# check data dimension
dim(Human_Allele_Frequencies)

# run RDM function
rd_tre <- RDM(Human_Allele_Frequencies, outgroup = "San", use = "pairwise.complete.obs")

# result visualization
plot(rd_tre, use.edge.length = FALSE, cex = 0.5)
```

Index