

# Package ‘rapidphylo’

December 19, 2022

**Type** Package

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

**Version** 0.1.1

**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  
Jing Peng et al. (2021) <[doi:10.1016/j.ympev.2021.107142](https://doi.org/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** 2022-12-19 13:10:05 UTC

## R topics documented:

Human_Allele_Frequencies . . . . .	2
RDM . . . . .	2
<b>Index</b>	<b>4</b>

---

Human\_Allele\_Frequencies

*Allele frequencies from 44,000 single nucleotide polymorphisms*

---

### Description

The dataset “Human\_Allele\_Frequencies” is a  $5 \times 44,000$  matrix that contains allele frequencies from 44,000 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 44000 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 <code>stats::cov</code> for more details.

**Details**

The input matrix is the observed values of the frequencies at tips  $1, 2, \dots, P, P + 1$ . 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**

Jing Peng, Haseena Rajeevan, Laura Kubatko, and Arindam RoyChoudhury (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 44,000 sites for
# 5 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

## \* **datasets**

Human\_Allele\_Frequencies, [2](#)

Human\_Allele\_Frequencies, [2](#)

RDM, [2](#)