Package 'pldist'

August 24, 2018

Title Paired and Longitudinal Ecological Dissimilarities

Version 0.0.0.9000

Author Anna M. Plantinga [aut, cre], Jun Chen [aut]

Maintainer Anna M. Plantinga <amp9@williams.edu>

Description Calculates paired and longitudinal UniFrac, Bray-Curtis, Jaccard, Gower, and Kulczynski distances/dissimilarities. These metrics summarize changes in the microbiome over time and allow these changes to be compared across treatments, conditions, or levels of a covariate. For more information, please see Plantinga et al (2018+).

Depends R (>= 3.4.3)

Imports ape

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

NeedsCompilation no

Suggests knitr, rmarkdown, testthat

VignetteBuilder knitr

R topics documented:

braycurtis *Paired or longitudinal Bray-Curtis distances*

Description

The distances are calculated as follows, where d_k^X is the within-subject measure of change appropriate to the setting (paired/longitudinal and quantitative/qualitative), as described in the full package documentation and vignette. $D_{AB} = (1/m) * \sum_k |d_k^A - d_k^B|$

Usage

```
braycurtis(tsf.data, binary)
```

Arguments

tsf.data	Transformed OTU table and metadata (from function pl.transform)
binary	Logical indicating whether to use the binary version of the distance

Value

Returns an n x n distance matrix.

counts2props *counts2props*

Description

Converts OTU counts to OTU proportions/relative abundances.

Usage

```
counts2props(x)
```

Arguments

x	Matrix of OTU counts (rows are subjects, columns are taxa).

Value

n x p matrix of OTU proportions.

flexsign *flexsign*

Description

Sign function that considers 0 both positive and negative. Returns 1 if the two numbers are the same sign, 0 otherwise. Vectorized (compares vectors elementwise).

Usage

```
flexsign(v1, v2)
```

Arguments

v1	First vector
v2	Second vector

Value

Returns an n x n distance matrix.

gower *Paired or longitudinal Gower distances*

Description

The distances are calculated as follows, where d_k^X is the within-subject measure of change appropriate to the setting (paired/longitudinal and quantitative/qualitative), as described in the full package documentation and vignette.

$D_{AB} = (1/m) \sum_k (|d_k^A - d_k^B|)/(\max d_k - \min d_k)$

Usage

```
gower(tsf.data, binary)
```

Arguments

tsf.data	Transformed OTU table and metadata (from function pl.transform)
binary	Logical indicating whether to use the binary version of the distance

Value

Returns an n x n distance matrix.

jaccard *Paired or longitudinal Jaccard distances*

Description

The distances are calculated as follows, where d_k^X is the within-subject measure of change appropriate to the setting (paired/longitudinal and quantitative/qualitative), as described in the full package documentation and vignette. Paired, qualitative: $D_{AB} = 1 - \sum_k I(d_k^A = d_k^B) I(d_k^A \neq 0) / \sum_k [I(d_k^A \neq 0) + I(d$
Paired, quantitative: $D_{AB} = 1 - \sum_k \min(|d_k^A|, |d_k^B|)\, I(sgn(d_k^A) = sgn(d_k^B)) / \sum_k \max(|d_k^A|, |d_k^b|)$
Longitudinal: $D_{AB} = 1 - (\sum_k \min(d_k^A, d_k^B)) / (\sum_k \max(d_k^A, d_k^B))$

Usage

```
jaccard(tsf.data, paired, binary)
```

Arguments

tsf.data	Transformed OTU table and metadata (from function pl.transform)
paired	Logical indicating whether paired analysis is desired
binary	Logical indicating whether to use the binary version of the distance

Value

Returns an n x n distance matrix.

kulczynski *Paired or longitudinal Kulczynski distances*

Description

The distances are calculated as follows, where d_k^X is the within-subject measure of change appropriate to the setting (paired/longitudinal and quantitative/qualitative), as described in the full package documentation and vignette.

Paired, qualitative: $D_{AB} = 1 - (1/m) \sum_k I[d_k^A = d_k^B] I[d_k^A \neq 0]$ Paired, quantitative: $D_{AB} = 1 - (2/m) \sum_k \min(|d_k^A|, |d_k^B|) I[sgn(d_k^A) = sgn(d_k^B)]$ Longitudinal: $D_{AB} = 1 - (1/m) * \sum_k \min(d_k^A, d_k^B)$

Usage

```
kulczynski(tsf.data, paired, binary)
```

Arguments

tsf.data	Transformed OTU table and metadata (from function pl.transform)
paired	Logical indicating whether paired analysis is desired
binary	Logical indicating whether to use the binary version of the distance

Value

Returns an n x n distance matrix.

`LUniFrac` *LUniFrac*

Description

Longitudinal UniFrac distances for comparing changes in microbial communities across 2 time points.

Usage

```
LUniFrac(otu.tab, tree, gam = c(0, 0.5, 1), metadata)
```

Arguments

`otu.tab`	OTU count table, containing 2*n rows (samples) and q columns (OTUs)
`tree`	Rooted phylogenetic tree of R class "phylo"
`gam`	Parameter controlling weight on abundant lineages. The same weight is used within a subjects as between subjects.
`metadata`	Data frame with three columns: subject identifiers (n unique values), sample identifiers (must match row names of otu.tab), and time or group indicator (numeric variable, or factor with levels such that as.numeric returns the desired ordering). Column names should be subjID, sampID, time.

Details

Based in part on Jun Chen & Hongzhe Li (2012), GUniFrac.

Computes difference between time points and then calculates difference of these differences, resulting in a dissimilarity matrix that can be used in a variety of downstream distance-based analyses.

Value

Returns a (K+1) dimensional array containing the longitudinal UniFrac dissimilarities with the K specified gamma values plus the unweighted distance. The unweighted dissimilarity matrix may be accessed by result[,,"d_UW"], and the generalized dissimilarities by result[,,"d_G"] where G is the particular choice of gamma.

`pl.transform` *tsf_paired*

Description

OTU transformation for longitudinal data. Computes average within-subject change (in presence for qualitative metrics, abundance for quantitative metrics) during one unit of time for each taxon.

Usage

```
pl.transform(otus, metadata, paired)
```

Arguments

otus	Matrix of OTU counts or proportions. Notes: (1) Will be transformed to proportions if it's not already; (2) Row names must be sample identifiers (matching metadata), and column names must be OTU identifiers (enforced if using UniFrac distances).
metadata	Data frame with three columns: subject identifiers (n unique values, column name "subjID"), sample identifiers (must match row names of otu.tab, column name "sampID"), and time point or group identifier (if using longitudinal distances, this must be numeric or convertable to numeric).
paired	Logical indicating whether to use the paired version of the metric (TRUE) or the longitudinal version (FALSE). Paired analyis is only possible when there are exactly 2 unique time points/identifiers for each subject or pair.

Value

List with the following elements. Both data matrices have subject identifiers as row names and OTU identifiers as column names.

tsf.data	List with 3 elements: (1) dat.binary: n x p matrix of data after longitudinal, binary/qualitative transformation (2) dat.quant: n x p matrix of data after longitudinal, quantitative transformation (3) avg.prop: n x p matrix with overall average proportion of each taxon
type	Type of transformation that was used (paired, balanced longitudinal, unbalanced longitudinal) with a warning if unbalanced longitudinal.

---

pldist *pldist*

Description

Function that calculates paired and longitudinal ecological distance/dissimilarity matrices. Includes qualitative and quantitative versions of Bray-Curtis, Jaccard, Kulczynski, Gower, and unweighted and generalized UniFrac distances/dissimilarities. UniFrac-based metrics are based in part on GUniFrac (Jun Chen & Hongzhe Li (2012)).

Usage

```
pldist(otus, metadata, paired = FALSE, binary = FALSE, method,
  tree = NULL, gam = c(0, 0.5, 1))
```

Arguments

otus	OTU count or frequency table, containing one row per sample and one column per OTU.
metadata	Data frame with three columns: subject identifiers (n unique values, column name "subjID"), sample identifiers (must match row names of otu.tab, column name "sampID"), and time point or group identifier (if using longitudinal distances, this must be numeric or convertable to numeric).

paired	Logical indicating whether to use the paired version of the metric (TRUE) or the longitudinal version (FALSE). Paired analyis is only possible when there are exactly 2 unique time points/identifiers for each subject or pair.
binary	Logical indicating whether to use the qualitative (TRUE) or quantitative (FALSE) version of each metric. Qualitative analysis only incorporates changes in OTU presence or absence; quantitative analysis incorporates changes in abundance.
method	Desired distance metric. Choices are braycurtis, jaccard, kulczynski, gower, and unifrac, or any unambiguous abbreviation thereof.
tree	Rooted phylogenetic tree of R class "phylo". Default NULL; only needed for UniFrac family distances.
gam	Parameter controlling weight on abundant lineages for UniFrac family distances. The same weight is used within a subject as between subjects. Default (0, 0.5, 1).

Value

Returns a list with elements:

D	If any metric other than UniFrac is used, D is an n x n distance (or dissimilarity) matrix. For UniFrac-family dissimilarities, D is a (K+1) dimensional array containing the paired or longitudinal UniFrac dissimilarities with the K specified gamma values plus the unweighted distance. The unweighted distance matrix may be accessed by result[,,"d_UW"], and the generalized dissimilarities by result[,,"d_G"] where G is the particular choice of gamma.
type	String indicating what type of dissimilarity was requested.

`PUniFrac` *PUniFrac*

Description

Paired UniFrac distances for comparing changes in microbial communities across 2 groups or time points.

Usage

```
PUniFrac(otu.tab, tree, gam = c(0, 0.5, 1), metadata)
```

Arguments

otu.tab	OTU count table, containing 2*n rows (samples) and q columns (OTUs)
tree	Rooted phylogenetic tree of R class "phylo"
gam	Parameter controlling weight on abundant lineages. The same weight is used within a subject as between subjects.
metadata	Data frame with three columns: subject identifiers (n unique values, column name "subjID"), sample identifiers (must match row names of otu.tab, column name "sampID"), and time point or group identifier (variable with two unique levels, column name "time").

Details

Based in part on Jun Chen & Hongzhe Li (2012), GUniFrac.

Computes difference between time points and then calculates difference of these differences, resulting in a dissimilarity matrix that can be used in a variety of downstream distance-based analyses.

Value

Returns a (K+1) dimensional array containing the longitudinal UniFrac dissimilarities with the K specified gamma values plus the unweighted distance. The unweighted dissimilarity matrix may be accessed by result[,,"d_UW"], and the generalized dissimilarities by result[,,"d_G"] where G is the particular choice of gamma.

---

tsf_long *tsf_paired*

Description

OTU transformation for longitudinal data. Computes average within-subject change (in presence for qualitative metrics, abundance for quantitative metrics) during one unit of time for each taxon.

Usage

```
tsf_long(otus, metadata)
```

Arguments

otus	Matrix of OTU counts or proportions. Notes: (1) Will be transformed to proportions if it's not already; (2) Row names must be sample identifiers (matching metadata), and column names must be OTU identifiers (enforced if using UniFrac distances).
metadata	Data frame with three columns: subject identifiers (n unique values, column name "subjID"), sample identifiers (must match row names of otu.tab, column name "sampID"), and time point or group identifier (if using longitudinal distances, this must be numeric or convertable to numeric).

Value

List with the following elements. Both data matrices have subject identifiers as row names and OTU identifiers as column names.

dat.binary	n x p matrix of data after longitudinal, binary/qualitative transformation
dat.quant	n x p matrix of data after longitudinal, quantitative transformation
avg.prop	n x p matrix with overall average proportion of each taxon

tsf_paired *tsf_paired*

Description

OTU transformation for paired data. Computes within-subject change (in presence for qualitative metrics and abundance for quantitative metrics) between time points for each taxon.

Usage

```
tsf_paired(otus, metadata)
```

Arguments

otus	Matrix of OTU counts or proportions. Notes: (1) Will be transformed to proportions if it's not already; (2) Row names must be sample identifiers (matching metadata), and column names must be OTU identifiers (enforced if using UniFrac distances).
metadata	Data frame with three columns: subject identifiers (n unique values, column name "subjID"), sample identifiers (must match row names of otu.tab, column name "sampID"), and time point or group identifier (must have two unique values for paired transformation).

Value

List with the following elements. Both data matrices have subject identifiers as row names and OTU identifiers as column names.

dat.binary	n x p matrix of data after paired, binary/qualitative transformation
dat.quant	n x p matrix of data after paired, quantitative transformation
avg.prop	n x p matrix with overall average proportion of each taxon

Index