

Package ‘MicrobiomeCluster’

January 28, 2022

Type Package

Title Microbiome Dataset Manipulating and Clustering

Version 1.0

Date 2019-11-20

Author Yushu Shi, Robert Jenq

Maintainer Yushu Shi <shiyushu2006@gmail.com>

Description Basic functions calculating OTU and tree node abundance. In addition, it also provides distance and levels of the phylogenetic tree nodes and tips from the tree root. Researchers can use MicrobiomeCluster package to manipulate an existing tree by trimming or growing branches, or check if a Beta metric is suitable to be applied for clustering a dataset. We also include a new combined Beta metric that can give good clustering performance for a variety of datasets.

License GPL (>= 2)

Imports ape, adephylo

Suggests mclust, GUniFrac, cluster, vegan

NeedsCompilation no

Repository CRAN

Depends R (>= 2.10)

R topics documented:

calNodeAbun	2
combMetric	2
convCounts2Abun	3
cumAbunofHighAbun	4
dist2Root	4
growBranch	5
level2Root	6
Martinez	7
Schnorr	7
Smits	7
switch2Zero	8
trimBranch	9

Index	11
--------------	-----------

calNodeAbun	<i>Calculate average abundance for phylogenetic tree nodes</i>
-------------	--

Description

Calculate average abundance for phylogenetic tree nodes

Usage

```
calNodeAbun(tree, AvgAbun)
```

Arguments

tree	A tree object
AvgAbun	A vector of average OTU abundance. Its names are OTU names.

Details

Here we adopt the name "CROTU" for tree node following the convention of Peled et al.

Value

A vector of the average abundance for each Tree Node.

Source

Jonathan U. Peled, Sean M. Devlin, Anna Staffas, Melissa Lumish, Raya Khanin, Eric R. Littmann, Lilan Ling, Satyajit Kosuri, Molly Maloy, John B. Slingerland, Katya F. Ahr, Kori A. Porosnicu Rodriguez, Yusuke Shono, Ann E. Slingerland, Melissa D. Docampo, Anthony D. Sung, Daniela Weber, Amin M. Alousi, Boglarka Gyurkocza, Doris M. Ponce, Juliet N. Barker, Miguel-Angel Perales, Sergio A. Giralt, Ying Taur, Eric G. Pamer, Robert R. Jenq, and Marcel R.M. van den Brink (2017) Intestinal Microbiota and Relapse After Hematopoietic-Cell Transplantation, *Journal of Clinical Oncology*, vol 35, num 15, 1650-1659

Yushu Shi, Liangliang Zhang, Christine B. Peterson, Kim-Anh Do, Robert R. Jenq *Performance determinants of unsupervised clustering methods for microbiome data Accepted by Microbiome*

combMetric	<i>Calculate combined metrics of the Bray Curtis dissimilarity and the unweighted UniFrac dissimilarity</i>
------------	---

Description

Calculate combined metrics, which inherits complimentary insights from the Bray Curtis and the Unweighted UniFrac dissimilarities

Usage

```
combMetric(bray,uu,alpha)
```

Arguments

bray	Bray Curtis distance matrix, with row names and column names correspond to the sample ID.
uu	Unweighted UniFrac distance matrix, with row names and column names correspond to the sample ID.
alpha	Relative contribution of the unweighted Unifrac distance.

Details

$$d^{combined} = \alpha * d_{normalized}^{UU} + (1 - \alpha) * d_{normalized}^{BC}$$

Value

combined metric distance matrix

Source

Yushu Shi, Liangliang Zhang, Christine B. Peterson, Kim-Anh Do, Robert R. Jenq *Performance determinants of unsupervised clustering methods for microbiome data Accepted by Microbiome*

convCounts2Abun	<i>Convert an OTU count table to an OTU abundance table</i>
-----------------	---

Description

Convert an OTU count table to an OTU abundance table

Usage

```
convCounts2Abun(otucounttable)
```

Arguments

otucounttable	An OTU count table. Its rows correspond to OTUs, while its columns correspond to observations.
---------------	--

Value

a matrix of OTU abundance

Source

Yushu Shi,Liangliang Zhang,Christine B. Peterson, Kim-Anh Do, Robert R. Jenq *Performance determinants of unsupervised clustering methods for microbiome data Accepted by Microbiome*

cumAbunofHighAbun	<i>Cumulative Abundance of High Abundance OTUs.</i>
-------------------	---

Description

Cumulative Abundance of High Abundance OTUs. Here the threshold of the high abundance OTU is set to be 0.001. A low function output suggests Bray Curtis may perform poorly in recovering the true sample clusters.

Usage

```
cumAbunofHighAbun(otuabundance, threshold=0.001)
```

Arguments

otuabundance	An OTU abundance table
threshold	The threshold of the high abundance OTU. The default is 0.001.

Value

Sum of the abundances of OTUs with average abundance greater than the threshold.

Source

Yushu Shi, Liangliang Zhang, Christine Peterson, Kim-Anh Do, Robert Jenq *Combined Beta Metric for Unsupervised Clustering of Microbiome Data Manuscript submitted*

dist2Root	<i>Calculate distance from tree tips/nodes to root</i>
-----------	--

Description

Calculate distance from tree tips/nodes to root

Usage

```
dist2Root(tree)
```

Arguments

tree	A tree object.
------	----------------

Details

Here we adopt the name "CROTU" for tree node following the convention of Peled et al. Distance refers to the sum of branch lengths.

Value

otudistance	A vector of distance from OTUs to the root
crotudistance	A vector of distance from phylogenetic tree nodes (CROTUs) to the root

Source

Jonathan U. Peled, Sean M. Devlin, Anna Staffas, Melissa Lumish, Raya Khanin, Eric R. Littmann, Lilan Ling, Satyajit Kosuri, Molly Maloy, John B. Slingerland, Katya F. Ahr, Kori A. Porosnicu Rodriguez, Yusuke Shono, Ann E. Slingerland, Melissa D. Docampo, Anthony D. Sung, Daniela Weber, Amin M. Alousi, Boglarka Gyurkocza, Doris M. Ponce, Juliet N. Barker, Miguel-Angel Perales, Sergio A. Giralt, Ying Taur, Eric G. Pamer, Robert R. Jenq, and Marcel R.M. van den Brink (2017) Intestinal Microbiota and Relapse After Hematopoietic-Cell Transplantation, *Journal of Clinical Oncology*, 1650-1659

growBranch

Bifurcate each tree tip and make a new tree.

Description

Bifurcate each tree tip and make a new tree. Sequences of each OTU are randomly assigned to either of the two newly formed tips.

Usage

```
growBranch(tree, otutable)
```

Arguments

tree	A tree object.
otutable	An OTU table

Value

tree	A new tree created from bifurcating the old tree tips.
otutable	A new OTU table after assigning sequences to the new tree tips.

Source

Yushu Shi, Liangliang Zhang, Christine B. Peterson, Kim-Anh Do, Robert R. Jenq *Performance determinants of unsupervised clustering methods for microbiome data Accepted by Microbiome*

Examples

```
library(MicrobiomeCluster)
library(vegan)
library(ape)
library(mclust)
library(cluster)
data("Martinez")

# martinez1 and martinez2 are the OTU tables of the first and second generation descendants

originalbray <- vegdist(t(Martinez$otutable), method="bray")
martinez1 <- growBranch(Martinez$tree, Martinez$otutable)
bray1 <- vegdist(t(martinez1$otutable), method="bray")
martinez2 <- growBranch(martinez1$tree, martinez1$otutable)
bray2 <- vegdist(t(martinez2$otutable), method="bray")
```

```
# The performance of Bray Curtis deteriorates as tree tips bifurcate.

adjustedRandIndex(pam(originalbray, 2, diss=TRUE)$cluster, Martinez$sampleinfo)
adjustedRandIndex(pam(bray1, 2, diss=TRUE)$cluster, Martinez$sampleinfo)
adjustedRandIndex(pam(bray2, 2, diss=TRUE)$cluster, Martinez$sampleinfo)

# plot the PCOA plots of the original data,
# the first generation descendants and the second generation descendants
pcoaorig <- pcoa(originalbray)
pcoa1 <- pcoa(bray1)
pcoa2 <- pcoa(bray2)

col <- ifelse(Martinez$sampleinfo=="USA", "red", "blue")
par(mfrow=c(1, 3))
plot(pcoaorig$vectors[,1], pcoaorig$vectors[,2], col=col)
plot(pcoa1$vectors[,1], pcoa1$vectors[,2], col=col,
      xlim=c(-0.25,0.45), ylim=c(-0.4,0.35))
plot(pcoa2$vectors[,1], pcoa2$vectors[,2], col=col,
      xlim=c(-0.25,0.45), ylim=c(-0.4,0.35))
```

level2Root	<i>Calculate level of tree tips/nodes to the root.</i>
------------	--

Description

Calculate level of tree tips/nodes to the root.

Usage

```
level2Root(tree)
```

Arguments

tree A tree object.

Details

Here we adopt the name "CROTU" for tree node following the convention of Peled et al. The level refers to the number of branches from the root to the CROTU or OTU.

Value

otulevel	Level of each OTU
crotulevel	Level for each phylogenetic tree node (CROTU)

Source

Jonathan U. Peled, Sean M. Devlin, Anna Staffas, Melissa Lumish, Raya Khanin, Eric R. Littmann, Lilan Ling, Satyajit Kosuri, Molly Maloy, John B. Slingerland, Katya F. Ahr, Kori A. Porosnicu Rodriguez, Yusuke Shono, Ann E. Slingerland, Melissa D. Docampo, Anthony D. Sung, Daniela Weber, Amin M. Alousi, Boglarka Gyurkocza, Doris M. Ponce, Juliet N. Barker, Miguel-Angel Perales, Sergio A. Giralt, Ying Taur, Eric G. Pamer, Robert R. Jenq, and Marcel R.M. van den Brink (2017) Intestinal Microbiota and Relapse After Hematopoietic-Cell Transplantation, *Journal of Clinical Oncology*, 1650-1659

Martinez

*Martinez data***Description**

Martinez dataset includes OTU table, tree and sample geographic information.

Source

Martinez, I., Stegen, J. C., Maldonado-Gomez, M. X., Eren, A. M., Siba, P. M., Green-hill, A. R., and Walter, J. (2015). *The gut microbiota of rural Papua New Guineans: composition, diversity patterns, and ecological processes*. *Cell reports*, 11(4):527-538.

Schnorr

*Schnorr data***Description**

Schnorr dataset includes OTU table, tree and sample geographic information.

Source

Schnorr, S. L., Candela, M., Rampelli, S., Centanni, M., Consolandi, C., Basaglia, G., Turrone, S., Biagi, E., Peano, C., Severgnini, M., et al. (2014). *Gut microbiome of the Hadza hunter-gatherers*. *Nature communications*, 5:3654.

Smits

*Smits data***Description**

Smits dataset includes OTU table, tree and sample seasonal information. For the purpose of unsupervised clustering, we include only the most separated seasons, the early wet season and the late dry season.

Source

Smits, S. A., Leach, J., Sonnenburg, E. D., Gonzalez, C. G., Lichtman, J. S., Reid, G., Knight, R., Manjurano, A., Chantalucha, J., Elias, J. E., Dominguez-Bello, M. G., and Sonnenburg, J. L. (2017). *Seasonal cycling in the gut microbiome of the hadza hunter-gatherers of Tanzania*. *Science*, 357(6353):802-806.

switch2Zero	<i>Convert OTU counts below the threshold to 0.</i>
-------------	---

Description

Convert OTU counts below the threshold to 0. This helps recovering the true clustering for un-weighted Unifrac distance.

Usage

```
switch2Zero(otutable, threshold)
```

Arguments

otutable	An OTU table
threshold	Desired threshold.

Value

otutable	A new OTU table after coverting counts below threshold to 0
----------	---

Source

Yushu Shi, Liangliang Zhang, Christine B. Peterson, Kim-Anh Do, Robert R. Jenq *Performance determinants of unsupervised clustering methods for microbiome data* Accepted by Microbiome

Examples

```
library(MicrobiomeCluster)
library(ape)
library(GUniFrac)
library(mclust)
library(cluster)

data("Smits")

uuorig <- as.dist(GUniFrac(t(Smits$otutable), Smits$tree)$unifrac[ , , "d_UW"])
newsmits <- switch2Zero(Smits$otutable, 30)
uu1 <- as.dist(GUniFrac(t(newsmits), Smits$tree)$unifrac[ , , "d_UW"])

# The performance of unweighted Unifrac improves after we convert counts below 30 to 0

adjustedRandIndex(pam(uuorig, 2, diss=TRUE)$cluster, Smits$sampleinfo)
adjustedRandIndex(pam(uu1, 2, diss=TRUE)$cluster, Smits$sampleinfo)

# Plot the PCOA plot for the original data and the modified data.
pcoaorig <- pcoa(uuorig)
pcoa1 <- pcoa(uu1)

col <- ifelse(Smits$sampleinfo=="Early Wet", "red", "blue")
par(mfrow=c(1,2))
plot(pcoaorig$vectors[ ,1], pcoaorig$vectors[ ,2], col=col,
     xlim=c(-0.4,0.4), ylim=c(-0.35,0.35))
```



```
plot(pcoa1$vector[,1], pcoa1$vector[,2], col=col,
     xlim=c(-0.4,0.4), ylim=c(-0.35,0.35))
```

trimBranch	<i>Trim tree tips and make a new tree.</i>
------------	--

Description

Trim the tree tips to the given value. Sequences from trimmed OTUs are combined to the upper level.

Usage

```
trimBranch(level, tree, otutable)
```

Arguments

level	Desired maximum hierarchy level of the output tree.
tree	A tree object.
otutable	An OTU table.

Value

tree	A new tree created from trimming distant tips and nodes.
otutable	A new OTU table after combining sequences that are trimmed to the upper level.

Source

Yushu Shi, Liangliang Zhang, Christine B. Peterson, Kim-Anh Do, Robert R. Jenq *Performance determinants of unsupervised clustering methods for microbiome data* Accepted by Microbiome

Examples

```
library(MicrobiomeCluster)
library(mclust)
library(cluster)
library(vegan)
library(ape)

data("Schnorr")

#High abundance OTUs control the performance of the Bray Curtis distance.

SchnorrAbundance <- convCounts2Abun(Schnorr$otutable)
cumAbunofHighAbun(SchnorrAbundance)

originalbray <- vegdist(t(Schnorr$otutable), method="bray")
SchnorrTrimmed <- trimBranch(37,Schnorr$tree,Schnorr$otutable)

SchnorrTrimmedAbundance <- convCounts2Abun(SchnorrTrimmed$otutable)
cumAbunofHighAbun(SchnorrTrimmedAbundance)
```

```
trimmedbray <- vegdist(t(SchnorrTrimmed$otutable), method="bray")

#The performance of Bray Curtis improves after trimming distant branches.

adjustedRandIndex(pam(originalbray,2,diss=TRUE)$cluster, Schnorr$sampleinfo)
adjustedRandIndex(pam(trimmedbray,2,diss=TRUE)$cluster, Schnorr$sampleinfo)

#plot the PCOA plots of the original data and the modified data

pcoaorig <- pcoa(originalbray)
pcoatrimmed <- pcoa(trimmedbray)

col <- ifelse(Schnorr$sampleinfo=="Italy", "red", "blue")
par(mfrow=c(1,2))
plot(pcoaorig$vectors[,1],pcoaorig$vectors[,2], col=col,
      xlim=c(-0.4,0.45), ylim=c(-0.4,0.45))
plot(pcoatrimmed$vectors[,1],pcoatrimmed$vectors[,2], col=col,
      xlim=c(-0.4,0.45), ylim=c(-0.4,0.45))
```

Index

calNodeAbun, [2](#)
combMetric, [2](#)
convCounts2Abun, [3](#)
cumAbunofHighAbun, [4](#)

dist2Root, [4](#)

growBranch, [5](#)

level2Root, [6](#)

Martinez, [7](#)

Schnorr, [7](#)
Smits, [7](#)
switch2Zero, [8](#)

trimBranch, [9](#)