Package 'MicrobiomeCluster'

February 23, 2019

Type Package

Title Microbiome Dataset Manipulating and Clustering

Version 1.0

Date 2019-02-23

Author Yushu Shi, Robert Jenq

Maintainer Yushu Shi <shiyushu2006@gmail.com>

Description This package includes basic functions calculating OTU and CROTU abundance. It also provides distance and levels of tree nodes and tips from its root. Researchers can use MicrobiomeCluster package to manipulate an existing tree by trimming or growing branches, or check if a Beta metric is suitable for clustering a dataset.

License GPL (>= 2)

Imports ape, adephylo

Suggests mclust, GUniFrac, cluster, vegan

NeedsCompilation no

Repository CRAN

Depends R (>= 2.10)

R topics documented:

CalculateCROTUAbundance

Calculate average abundance for CROTUs

Description

Calculate average abundance for CROTUs

Usage

CalculateCROTUAbundance(tree, AverageAbundance)

Arguments

tree   A tree object

AverageAbundance
                   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 CROTU.

Source

Jonathan U. Peled,corresponding author 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 Peterson, Kim-Anh Do, Robert Jenq *Performance Determinants of Unsupervised Clustering Methods for Microbiome Data Manuscript submitted*

ConvertOTUCountsToAbundances

Convert an OTU count table to an OTU abundance table

Description

Convert an OTU count table to an OTU abundance table

Usage

ConvertOTUCountsToAbundances(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 Peterson, Kim-Anh Do, Robert Jenq *Performance Determinants of Unsupervised Clustering Methods for Microbiome Data Manuscript submitted*

CumulativeAbunofHighAbun

Cumulative Abundance of High Abundance OTUs.

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

```
CumulativeAbunofHighAbun(otuabundance)
```

Arguments

otuabundance   An OTU abundance table

Value

Sum of the abundances of OTUs with average abundance greater than 0.001.

Source

Yushu Shi,Liangliang Zhang,Christine Peterson, Kim-Anh Do, Robert Jenq *Performance Determinants of Unsupervised Clustering Methods for Microbiome Data Manuscript submitted*

`DistanceToRoot`	*Calculate distance from tree tips/nodes to root*

Description

Calculate distance from tree tips/nodes to root

Usage

```
DistanceToRoot(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.

Source

Jonathan U. Peled,corresponding author 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

\itemotudistanceA vector of distance from OTUs to the root \itemcrotudistanceA vector of distance from CROTUs to the root

---

`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 Peterson, Kim-Anh Do, Robert Jenq *Performance Determinants of Unsupervised Clustering Methods for Microbiome Data Manuscript submitted*

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 descendan
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))
```

LevelToRoot	*Clculate level of tree tips/nodes to the root.*

Description

Calculate level of tree tips/nodes to the root.

Usage

```
LevelToRoot(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 CROTU

Source

Jonathan U. Peled,corresponding author 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., Turroni, 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., Changalucha, 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.

SwitchTo0 *Convert OTU counts below the threshold to 0.*

Description

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

Usage

```
SwitchTo0(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 Peterson, Kim-Anh Do, Robert Jenq *Performance Determinants of Unsupervised Clustering Methods for Microbiome Data Manuscript under preparation*

Examples

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

data("Smits")

uuorig <- as.dist(GUniFrac(t(Smits$otutable), Smits$tree)$unifracs[ , ,"d_UW"])
newsmits <- SwitchTo0(Smits$otutable, 30)
uu1 <- as.dist(GUniFrac(t(newsmits), Smits$tree)$unifracs[ , ,"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$vectors[ ,2], col=col,
xlim=c(-0.4,0.4), ylim=c(-0.35,0.35))
plot(pcoa1$vectors[ ,1], pcoa1$vectors[ ,2], col=col,
xlim=c(-0.4,0.4), ylim=c(-0.35,0.35))
```

TrimBranch   *Trim tree tips and make a new tree.*

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 Peterson, Kim-Anh Do, Robert Jenq *Performance Determinants of Unsupervised Clustering Methods for Microbiome Data Manuscript submitted*

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 <- ConvertOTUCountsToAbundances(Schnorr$otutable)
CumulativeAbunofHighAbun(SchnorrAbundance)

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

SchnorrTrimmedAbundance <- ConvertOTUCountsToAbundances(SchnorrTrimmed$otutable)
CumulativeAbunofHighAbun(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