Package 'GUniFrac'

July 2, 2014

Type Package

Title Generalized UniFrac distances

Version 1.0

Date 2012-04-27

Author Jun Chen

Maintainer Jun Chen <chenjun@mail.med.upenn.edu>

Description Generalized UniFrac distance for comparing microbial communities. Permutational multivariate analysis of variance using multiple distance matrices.

License GPL-2

Depends vegan, ape

Suggests ade4

LazyLoad yes

Repository CRAN

Date/Publication 2012-04-28 16:47:08

NeedsCompilation no

R topics documented:

GUniFrac-package *Generalized UniFrac distance for comparing microbial communities.*

Description

A generalized version of commonly used UniFrac distance. Generalized UniFrac distance contains an extra parameter controlling the weight on abundant lineages so the distance is not dominated by highly abundant lineages. The unweighted and weighted UniFrac, and variance adjusted weighted UniFrac distances are also implemented. The package also implements a permutation-based multivariate analysis of variance using **MULTIPLE** distance matrices.

Details

Package:	GUniFrac
Type:	Package
Version:	1.0
Date:	2012-04-27
License:	GPL-2
LazyLoad:	yes

Author(s)

Jun Chen <chenjun@mail.med.upenn.edu>

References

Jun Chen and Hongzhe Li(2012). Associating microbiome composition with environmental covariates using generalized UniFrac distances. (Submitted)

Examples

```
data(throat.otu.tab)
data(throat.tree)
data(throat.meta)

groups <- throat.meta$SmokingStatus

# Rarefaction
otu.tab.rff <- Rarefy(throat.otu.tab)$otu.tab.rff

# Calculate the UniFracs
unifracs <- GUniFrac(otu.tab.rff, throat.tree, alpha=c(0, 0.5, 1))$unifracs

dw <- unifracs[, , "d_1"] # Weighted UniFrac
```

```
du <- unifracs[, , "d_UW"] # Unweighted UniFrac
dv <- unifracs[, , "d_VAW"] # Variance adjusted weighted UniFrac
d0 <- unifracs[, , "d_0"]      # GUniFrac with alpha 0
d5 <- unifracs[, , "d_0.5"]    # GUniFrac with alpha 0.5

# Permanova - Distance based multivariate analysis of variance
adonis(as.dist(d5) ~ groups)

# Combine d(0), d(0.5), d(1) for testing
PermanovaG(unifracs[, , c("d_0", "d_0.5", "d_1")] ~ groups)
```

GUniFrac *Generalized UniFrac distances for comparing microbial communities.*

Description

A generalized version of commonly used UniFrac distances. It is defined as:

$$d^{(\alpha)} = \frac{\sum_{i=1}^{m} b_i (p_i^A + p_i^B)^\alpha \left| \frac{p_i^A - p_i^B}{p_i^A + p_i^B} \right|}{\sum_{i=1}^{m} b_i (p_i^A + p_i^B)^\alpha},$$

where m is the number of branches, b_i is the length of ith branch, p_i^A, p_i^B are the branch proportion for community A and B.

Generalized UniFrac distance contains an extra parameter α controlling the weight on abundant lineages so the distance is not dominated by highly abundant lineages. $\alpha = 0.5$ has overall the best power.

The unweighted and weighted UniFrac, and variance adjusted weighted UniFrac distances are also implemented.

Usage

```
GUniFrac(otu.tab, tree, alpha = c(0, 0.5, 1))
```

Arguments

`otu.tab`	OTU count table, row - n sample, column - q OTU
`tree`	Rooted phylogenetic tree of R class "phylo"
`alpha`	Parameter controlling weight on abundant lineages

Value

Return a LIST containing

`unifracs`	A three dimensional array containing all the UniFrac distance matrices including weighted/unweighted UniFrac distances ("d_1", "d_UW") and Variance adjusted weighted UniFrac distance ("d_VAW"). Other UniFrac distances can be retrieved using `unifracs[, , paste("d", alpha, sep="_")]`

Note

The function only accepts rooted tree. To root a tree, you may consider using the R code from https://stat.ethz.ch/pipermail/r-sig-phylo/2010-September/000750.html

Author(s)

Jun Chen <chenjun@mail.med.upenn.edu>

References

Jun Chen and Hongzhe Li(2012). Associating microbiome composition with environmental covariates using generalized UniFrac distances. (Submitted)

See Also

Rarefy, PermanovaG

Examples

```
require(ade4)

data(throat.otu.tab)
data(throat.tree)
data(throat.meta)

groups <- throat.meta$SmokingStatus

# Rarefaction
otu.tab.rff <- Rarefy(throat.otu.tab)$otu.tab.rff

# Calculate the UniFracs
unifracs <- GUniFrac(otu.tab.rff, throat.tree, alpha=c(0, 0.5, 1))$unifracs

dw <- unifracs[, , "d_1"] # Weighted UniFrac
du <- unifracs[, , "d_UW"] # Unweighted UniFrac
dv <- unifracs[, , "d_VAW"] # Variance adjusted weighted UniFrac
d0 <- unifracs[, , "d_0"]      # GUniFrac with alpha 0
d5 <- unifracs[, , "d_0.5"]    # GUniFrac with alpha 0.5

# Permanova - Distance based multivariate analysis of variance
adonis(as.dist(d5) ~ groups)

# PCoA plot
s.class(cmdscale(d5, k=2), fac = groups)
```

PermanovaG — *Permutational multivariate analysis of variance using multiple distance matrices*

Description

In practice, we do not know a priori which type of change happens in the microbiome. Each distance measure is most powerful in detecting only a certain scenario. When multiple distance matrices are available, separate tests using each distance matrix will lead to loss of power due to multiple testing correction. Combing the distance matrices in a single test will improve power. PermanovaG combines multiple distance matrices by taking the maximum of pseudo-F statistics for each distance matrix. Significance is assessed by permutation.

Usage

```
PermanovaG(formula, dat = NULL, ...)
```

Arguments

formula	Left side of the formula (`Y ~ X`) is a three dimensional ARRAY containing the supplied distance matrices as produced by GUniFrac function.
dat	DATA.FRAME containing the covariates
...	Parameter passing to adonis function

Value

Return a LIST containing:

aov.tab	DATA.FRAME (columns - F.model, p.value, rows - covariates)

Author(s)

Jun Chen <chenjun@mail.med.upenn.edu>

References

Jun Chen and Hongzhe Li(2012). Associating microbiome composition with environmental covariates using generalized UniFrac distances. (Submitted)

See Also

Rarefy, GUniFrac

Examples

```
data(throat.otu.tab)
data(throat.tree)
data(throat.meta)

groups <- throat.meta$SmokingStatus

# Rarefaction
otu.tab.rff <- Rarefy(throat.otu.tab)$otu.tab.rff

# Calculate the UniFracs
unifracs <- GUniFrac(otu.tab.rff, throat.tree, alpha=c(0, 0.5, 1))$unifracs

# Combine unweighted and weighted UniFrac for testing
PermanovaG(unifracs[, , c("d_1", "d_UW")] ~ groups)
# Combine d(0), d(0.5), d(1) for testing
PermanovaG(unifracs[, , c("d_0", "d_0.5", "d_1")] ~ groups)
```

`Rarefy` *Rarefy the OTU table to an equal sequencing depth*

Description

GUniFrac is also sensitive to different sequencing depth. To compare microbiomes on an equal basis, rarefaction might be used.

Usage

```
Rarefy(otu.tab, depth = min(rowSums(otu.tab)))
```

Arguments

`otu.tab`	OTU count table, row - n sample, column - q OTU
`depth`	Required sequencing depth; If not specified, the lowest sequencing depth is used.

Value

Return a LIST containing:

`otu.tab.rff`	Rarefied OTU table
`discard`	IDs of samples that do not reach the specified sequencing depth

Author(s)

Jun Chen <chenjun@mail.med.upenn.edu>

References

Jun Chen and Hongzhe Li(2012). Associating microbiome composition with environmental covariates using generalized UniFrac distances. (Submitted)

See Also

GUniFrac, PermanovaG

Examples

```
data(throat.otu.tab)
# Rarefaction
otu.tab.rff <- Rarefy(throat.otu.tab, 1024)$otu.tab.rff
```

---

throat.meta *Meta data of the throat microbiome samples.*

Description

It is part of a microbiome data set for studying the effect of smoking on the upper respiratory tract microbiome. The original data set contains samples from both throat and nose microbiomes, and from both body sides. This data set comes from the throat microbiome of left body side. It contains 60 subjects consisting of 32 nonsmokers and 28 smokers.

Usage

```
data(throat.meta)
```

Format

The format is: chr "throat.meta"

Source

Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, et al. (2010) Disordered Microbial Communities in the Upper Respiratory Tract of Cigarette Smokers. PLoS ONE 5(12): e15216.

Examples

```
data(throat.meta)
## maybe str(throat.meta) ; plot(throat.meta) ...
```

throat.otu.tab *OTU count table from 16S sequencing of the throat microbiome samples.*

Description

It is part of a microbiome data set for studying the effect of smoking on the upper respiratory tract microbiome. The original data set contains samples from both throat and nose microbiomes, and from both body sides. This data set comes from the throat microbiome of left body side. It contains 60 subjects consisting of 32 nonsmokers and 28 smokers.

Usage

```
data(throat.otu.tab)
```

Format

The format is: chr "throat.otu.tab"

Details

The OTU table is produced by the QIIME software. Singleton OTUs have been discarded.

Source

Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, et al. (2010) Disordered Microbial Communities in the Upper Respiratory Tract of Cigarette Smokers. PLoS ONE 5(12): e15216.

Examples

```
data(throat.otu.tab)
## maybe str(throat.otu.tab) ; plot(throat.otu.tab) ...
```

throat.tree *UPGMA tree of the OTUs from 16S sequencing of the throat microbiome samples.*

Description

The OTU tree is constructed using UPGMA on the K80 distance matrice of the OTUs. It is a rooted tree of class "phylo".

Usage

```
data(throat.tree)
```

Format

The format is: chr "throat.tree"

Details

The OTUs are produced by the QIIME software. Singleton OTUs have been discarded.

Source

Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, et al. (2010) Disordered Microbial Communities in the Upper Respiratory Tract of Cigarette Smokers. PLoS ONE 5(12): e15216.

Examples

```
data(throat.tree)
## maybe str(throat.tree) ; plot(throat.tree) ...
```

Index