Package 'mpress'

November 7, 2018

Type Package

Title Microbiome Power Estimation with Sampling and Simulation

Version 0.9.2

Description Given a phyloseq representation of a microbiome complete with metadata, MPrESS returns the estimated power to detect differences between samples with different metadata values either by sampling or by simulation if there are insufficient samples.

Depends R(>= 3.4.0), ggplot2, phytools, phyloseq(>= 1.16), vegan, HMP, ape, DESeq2

License GPL-2

LazyData true

RoxygenNote 6.0.1

R topics documented:

`mpress-class` *An S4 class representing the MPrESS data and output*

Description

An S4 class representing the MPrESS data and output

Slots

`dist.mst` The distance matrix showing the distances between different genomes

`phy` The phylogenetic tree in newick format

`rank` The ranks of the respective genomes with lower getting higher priority in being called as a medoid

`cluster` A vector giving the numeric cluster ID for each genome

`h` The threshold variable used to make the clusters

`medoids` A vector giving the medoid for each cluster

`gmm` A data.frame containing all the gaussian distributions used to find the threshold when available

`gmm.orig` A data.frame containing all the gaussian distributions prior to cleaning. Used to recalculate the threshold when needed

`annotation` A data.frame containing the annotations foreach genome with the genomes as the rows and the annotations with the columns. The

---

`plot.mpress` *plot.mpress makes a ggplot with the p-value and power for all the samples investigated an mpress object*

Description

plot.mpress makes a ggplot with the p-value and power for all the samples investigated an mpress object

Usage

```
plot.mpress(x)
```

Examples

```
library(mpress);
#Loading in the  microbiome files
data(ChinaData);
data(SpainData);

#Use summary() to examine the data loaded

#Use plot() to see the plot of the power and p-value in the different sample numbers
plot(spain.ibs.power)
```

`power.est` *power.calc*

Description

power.calc() calculates the number of samples to dectect a difference between microbiomes with two different metadata values.

Usage

```
power.est(in.phyloseq, metadata.var, metadata.vals, start.n = 5,
  alpha = 0.05, beta = 0.95, test.type = "permanova",
  dist.metric = "bray", deseq.val = 0, n.rep = 100, burn.in = 10,
  verbose = T, binom = F, switch.val = 0, seed.val = 0)
```

Arguments

in.phyloseq	Required. The collection of microbiome samples to use. Required to be a class phyloseq data
metadata.var	Required. The column name in the metadata values that is being compared for the power.
metadata.vals	Required. Vector with the values to be compared.
alpha	Alpha value used to test the microbiome differences. Default is 0.05
beta	Beta value used to test proportion of estimated samples that contain significant differences. Default is 0.95
test.type	String giving the statistic to test the between the microbiome samples. Default is "permanova"
dist.metric	String giving the distance metric to use between the microbiome samples. Default "bray" (Bray-Curtis). Available options can be seen with distanceMethodList.
deseq.val	Value to use for the DESeq2 run to trim the taxa. if the value is 0,use all taxa (Default); if the value is between 0 and 1, all taxa with an FDR below the value will be selected; otherwise, if the value is >=1, that number of top taxa by FDR will be selected
n.rep	Maximum number of replicates to run before final check if above the beta threshold. Default = 100.
burn.in	Value to burning in for the binomial test. Default is 10
verbose	Flag to prints intermediate values to the screen. Default = F
binom	Flag to use the binomial test to stop low mapping values. Default = F
switch.val	sample number at which to automatically swap between sampling and simulation. Default = 0, or only swap once sample number exceeds the number of samples
seed.val	Value to set seed value for troubleshooting. Default = 0 (not set)
start.	The sample size at which to start the tests. Default is 5.

Value

Returns a class MPrESS variable with the sample number in addition to runtime information for use in verification

Examples

```
#The following data is from Chavda et al 2016 which phylotyped Enterobacter genomes

#Additional printing and plotting options are availible with plot() and print(). For more
#The following data is from Zhang et al 2015 using Microbiome data from multiple sites in
# Our example uses the data underpinning the tree shown in Figure 2

library(mpress);
#Loading in the Chinese metadata file
data(ChinaData);
data(SpainData);

#Use summary() to examine the data loaded
```

```
summary(china.power)

#Other microbiomes: china.trim from data(ChinaData): unnamed OTUs removed
#                  spain.ibs.trim from

#Use plot() to see the plot of the power and p-value in the different sample numbers
plot(china.power)
```

`print.mpress` *print.mpress makes a table with the runtime information for the samples investigated an mpress object*

Description

print.mpress makes a table with the runtime information for the samples investigated an mpress object

Usage

```
## S3 method for class 'mpress'
print(x)
```

Examples

```
library(mpress);
#Loading in the microbiome files
data(ChinaData);
data(SpainData);

#Use summary() to examine the data loaded

#Use print() to see the table of the run info in the different sample numbers
print(spain.ibs.power)
```

`summary.mpress` *summary.mpress writes the power value and the estimation type to the screen*

Description

summary.mpress writes the power value and the estimation type to the screen

Usage

```
## S3 method for class 'mpress'
summary(x)
```

Examples

```
library(mpress);
#Loading in the  microbiome files
data(ChinaData);
data(SpainData);

summary(china.trim.power)
```

Index