Package 'holobiont'

October 16, 2024

Type Package

Title Microbiome Analysis Tools

Version 0.1.2

Description We provide functions for identifying the core community phylogeny in any microbiome, drawing phylogenetic Venn diagrams, calculating the core Faith's PD for a set of communities, and calculating the core UniFrac distance between two sets of communities. All functions rely on construction of a core community phylogeny, which is a phylogeny where branches are defined based on their presence in multiple samples from a single type of habitat. Our package provides two options for constructing the core community phylogeny, a tip-based approach, where the core community phylogeny is identified based on incidence of leaf nodes and a branch-based approach, where the core community phylogeny is identified based on incidence of individual branches. We suggest use of the microViz package, which can be downloaded from the website provided under Additional repositories.

License GPL-2

Encoding UTF-8

biocViews Software, phyloseq

Imports ape, phyloseq, phytools, ggplot2, dplyr, utils, tibble

Suggests microViz

RoxygenNote 7.3.2

Additional_repositories https://david-barnett.r-universe.dev/builds

NeedsCompilation no

Author Sharon Bewick [aut, cre],
Benjamin Camper [aut],
National Science Foundation Division of Integrative Organismal Systems
(award #2104605) [fnd]

Maintainer Sharon Bewick <sbewick@clemson.edu>

Repository CRAN

Date/Publication 2024-10-16 17:40:02 UTC

Contents

---

coreFaithsPD *Faiths PD of the Core Microbiome*

Description

Calculates Faith's phylogenetic diversity (PD) of a core microbiome based on either the tip-based or the branch-based core community phylogeny.

Usage

```
coreFaithsPD(x, core_fraction, mode='branch', rooted=TRUE)
```

Arguments

x	(Required) Microbial community data. This must be in the form of a phyloseq object and must contain, at a minimum, an OTU abundance table and a phylogeny.
core_fraction	(Required) The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome.
mode	Whether to build a tip-based ('tip') or a branch-based ('branch') phylogeny. The default is 'branch'.
rooted	Whether to include the root of the phylogeny. The default is TRUE, meaning that the root is necessarily included in all phylogenies. This requires that the input tree be rooted.

Details

coreFaithsPD calculates Faith's PD (Faith, 1992) based on either the tip-based or the branch-based core community phylogeny. In both cases, Faith's PD is calculated as the sum of the branch lengths in the core community phylogeny. For more details, see Bewick and Camper (2025).

Value

This function returns the numeric value of Faith's PD for the core microbiome.

References

Bewick, S.A. and Benjamin T. Camper. "Phylogenetic Measures of the Core Microbiome" <doi:TBD>

Faith, Daniel P. "Conservation evaluation and phylogenetic diversity." Biological conservation 61.1 (1992): 1-10.

McMurdie, Paul J., and Susan Holmes. "phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data." PloS one 8.4 (2013): e61217.

McMurdie, Paul J., and Susan Holmes. "Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data." Biocomputing 2012. 2012. 235-246.

Examples

```
#Test with enterotype dataset
library(phyloseq)
library(ape)
library(phytools)
data(enterotype)

set.seed(1)

#Generate an example tree and label it with the names of the microbial taxa
enterotype_tree<-rtree(length(taxa_names(enterotype)))
enterotype_tree$tip.label<-taxa_names(enterotype)

#Create a phyloseq object with a tree
example_phyloseq<-phyloseq(otu_table(enterotype),phy_tree(as.phylo(enterotype_tree)))

coreFaithsPD(example_phyloseq,0.5)
```

---

coreJaccard   *Jaccard Distance Between Core Microbiomes from Two Habitats*

Description

Calculates the Jaccard distance between the core microbiomes from two different types of habitats based on either the tip-based or the branch-based core community phylogeny.

Usage

```
coreJaccard(x, grouping, core_fraction)
```

Arguments

x	(Required) Microbial community data. This must be in the form of a phyloseq object and must contain, at a minimum, an OTU abundance table.
grouping	(Required) A vector specifying which samples belong to which habitat type.
core_fraction	(Required) The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome.

Details

`coreJaccard` calculates the Jaccard distance (Jaccard, 1901 A and B) between the core micro-biomes from two different types of habitats. Briefly, the Jaccard distance is calculated as the number of unique taxa in the core communities from each of the two habitats, divided by the total number of taxa in the two habitats combined. For more details, see Bewick and Camper (2025).

Value

This function returns the numeric value of the Jaccard distance between two core microbiomes.

References

Bewick, S.A. and Benjamin T. Camper. "Phylogenetic Measures of the Core Microbiome" <doi:TBD>

Jaccard, P. (1901A) Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin de la Société Vaudoise des Sciences Naturelles 37, 547-579.

Jaccard, P. (1901B) Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. Bulletin de la Société Vaudoise des Sciences Naturelles 37, 241-272.

McMurdie, Paul J., and Susan Holmes. "phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data." PloS one 8.4 (2013): e61217.

McMurdie, Paul J., and Susan Holmes. "Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data." Biocomputing 2012. 2012. 235-246.

Examples

```
#Test with enterotype dataset
library(phyloseq)
library(ape)
library(phytools)
data(enterotype)

set.seed(1)

#Generate an example tree and label it with the names of the microbial taxa
enterotype_tree<-rtree(length(taxa_names(enterotype)))
enterotype_tree$tip.label<-taxa_names(enterotype)

#keep only those samples with gender identified
gendered<-which(!(is.na(sample_data(enterotype)$Gender)))
enterotypeMF<-prune_samples(sample_names(enterotype)[gendered],enterotype)

#Create a phyloseq object with a tree
example_phyloseq<-phyloseq(otu_table(enterotypeMF))

coreJaccard(example_phyloseq,grouping=sample_data(enterotypeMF)$Gender,0.5)
```

---

corePhyloVenn	*Phylogenetic Venn Diagram Comparing Core Microbiomes from Multiple Habitats*

Description

Generates a Venn diagram comparing and contrasting the shared and unique phylogenetic branch lengths of core microbiomes from two or more (up to seven) different types of habitats based on either the tip-based or the branch-based core community phylogeny.

Usage

```
corePhyloVenn(x, grouping, core_fraction,
mode = 'branch',rooted=TRUE, ordered_groups=NULL,show_percentage=TRUE,decimal=2,
fill_color=c('red','orange','yellow','green','blue','purple','black'),fill_alpha=0.5,
stroke_color='black',stroke_alpha = 1, stroke_size = 1,stroke_linetype = "solid",
set_name_color = "black", set_name_size = 6,text_color = "black",text_size = 4)
```

Arguments

x	(Required) Microbial community data. This must be in the form of a phyloseq object and must contain, at a minimum, an OTU abundance table and a phylogeny.
grouping	(Required) A vector specifying which samples belong to which habitat type.
core_fraction	(Required) The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome.
mode	Whether to build a tip-based ('tip') or a branch-based ('branch') phylogeny. The default is 'branch'.
rooted	Whether to include the root of the phylogeny. The default is TRUE, meaning that the root is necessarily included in all phylogenies. This requires that the input tree be rooted.
ordered_groups	When provided, specifies the order in which different habitats should be plotted. This order is matched to the color order specified in fill_color (see below).
show_percentage	
	If TRUE the Venn diagram is shown as percentages of the total branch length in each compartment. If FALSE the Venn diagram is shown as absolute branch lengths.
decimal	The number of decimal points to report in the Venn diagram compartments (either for percentages or absolute branch lengths).
fill_color	A vector specifying the colors to use for each of the different habitats in the Venn diagram. The default is to use the six colors of the rainbow, followed by black.
fill_alpha	The transparency of the fill colors in the Venn diagram.
stroke_color	The color of the outlines in the Venn diagram.

stroke_alpha	The transparency of the outlines in the Venn diagram.
stroke_size	The width of the outlines in the Venn diagram.
stroke_linetype	
	The style of the outlines in the Venn diagram.
set_name_color	The color of the font used to label each habitat.
set_name_size	The size of the font used to label each habitat
text_color	The color of the font used to report the value in each compartment of the Venn diagram
text_size	The size of the font used to report the value in each compartment of the Venn diagram

Details

corePhyloVenn generates a Venn diagram showing either the percentage of the phylogenetic branch length or the absolute phylogenetic branch length shared between the core community phylogenies from all possible combinations of up to seven different habitat types. For more details, see Bewick and Camper (2025).

Value

This function returns a plot of the phylogenetic Venn diagram.

References

Bewick, S.A. and Benjamin T. Camper. "Phylogenetic Measures of the Core Microbiome" <doi:TBD>

McMurdie, Paul J., and Susan Holmes. "phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data." PloS one 8.4 (2013): e61217.

McMurdie, Paul J., and Susan Holmes. "Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data." Biocomputing 2012. 2012. 235-246.

Examples

```
#Test with enterotype dataset
library(phyloseq)
library(ape)
library(phytools)
library(ggplot2)
data(enterotype)

set.seed(1)

#Generate an example tree and label it with the names of the microbial taxa
enterotype_tree<-rtree(length(taxa_names(enterotype)))
enterotype_tree$tip.label<-taxa_names(enterotype)

#keep only those samples with gender identified
gendered<-which(!(is.na(sample_data(enterotype)$Gender)))
enterotypeMF<-prune_samples(sample_names(enterotype)[gendered],enterotype)
```

```
#Create a phyloseq object with a tree
example_phyloseq<-phyloseq(otu_table(enterotypeMF),phy_tree(as.phylo(enterotype_tree)))

corePhyloVenn(example_phyloseq,grouping=sample_data(enterotypeMF)$Gender,0.5)
```

---

coreTree   *The Core Community Phylogeny*

Description

Identifies and plots the tip-based or the branch-based core community phylogeny based on the occurrence of abundance of different microbial lineages in a set of samples from a common habitat (e.g., type of host or environment).

Usage

```
coreTree(x, core_fraction,
mode='branch', NCcol = 'black', Ccol='red',rooted=TRUE,
branch.width=4,label.tips=FALSE, remove_zeros=TRUE,
plot.chronogram=FALSE)
```

Arguments

x	(Required) Microbial community data. This must be in the form of a phyloseq object and must contain, at a minimum, an OTU abundance table and a phylogeny.
core_fraction	(Required) The fraction of samples that a microbial taxon must be found in to be considered part of the core microbiome.
mode	Whether to build a tip-based (`tip`) or a branch-based (`branch`) phylogeny. The default is `branch`.
NCcol	The color to plot all branches of the phylogeny that are NOT part of the core community phylogeny. The default is black.
Ccol	The color to plot all branches of the phylogeny that are ARE part of the core community phylogeny. The default is red.
rooted	Whether to include the root of the phylogeny. The default is TRUE, meaning that the root is necessarily included in all phylogenies. This requires that the input tree be rooted.
branch.width	The width to use when plotting the branches of the phylogeny. The default is 4.
label.tips	Whether or not to label the tips of the phylogeny with the microbial taxon names. The default is FALSE.
remove_zeros	Whether or not to remove taxa that are missing from all samples prior to drawing the phylogeny. The default is TRUE.
plot.chronogram	
	Whether to plot a phylogeny or a chronogram. The default is FALSE.

Details

coreTree identifies either the tip-based or the branch-based core community phylogeny. For the tip-based core community phylogeny, individual microbial taxa are retained based on being present in a threshold number of samples or at a threshold abundance. Once core microbial taxa have been identified, they are used to reconstruct the core community phylogeny. For the branch-based core community phylogeny, the phylogenetic tree for the entire dataset is examined, branch-by-branch, to determine which branches should be retained based on being present in a threshold number of samples or at a threshold abundance. If rooted = TRUE, branches are counted based on individual sample phylogenies that include the root node. Likewise, the tip-based tree is forced to include the root. If rooted = FALSE, branches are counted based on individual sample phylogenies that span all taxa present in the sample. Similarly, the tip-based phylogeny is the tree that spans all core taxa, and may not include the root. For more details, see Bewick and Camper (2025).

Value

This function plots the phylogeny for the entire dataset in black and colors the branches that are part of the core community phylogeny in red. These colors can be altered using the NCcol and Ccol variables.

References

Bewick, S.A. and Benjamin T. Camper. "Phylogenetic Measures of the Core Microbiome" <doi:TBD>

McMurdie, Paul J., and Susan Holmes. "phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data." PloS one 8.4 (2013): e61217.

McMurdie, Paul J., and Susan Holmes. "Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data." Biocomputing 2012. 2012. 235-246.

Examples

```
#Test with enterotype dataset
library(phyloseq)
library(ape)
library(phytools)
data(enterotype)

set.seed(1)

#Generate an example tree and label it with the names of the microbial taxa
enterotype_tree<-rtree(length(taxa_names(enterotype)))
enterotype_tree$tip.label<-taxa_names(enterotype)

#Create a phyloseq object with a tree
example_phyloseq<-phyloseq(otu_table(enterotype),phy_tree(as.phylo(enterotype_tree)))

coreTree(example_phyloseq,0.5)
```

---

coreUniFrac | *UniFrac Distance Between Core Microbiomes from Two Habitats*

Description

Calculates the UniFrac distance between the core microbiomes from two different types of habitats based on either the tip-based or the branch-based core community phylogeny.

Usage

```
coreUniFrac(x, grouping, core_fraction, mode='branch', rooted=TRUE)
```

Arguments

x	(Required) Microbial community data. This must be in the form of a phyloseq object and must contain, at a minimum, an OTU abundance table and a phylogeny.
grouping	(Required) A vector specifying which samples belong to which habitat type.
core_fraction	(Required) The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome.
mode	Whether to build a tip-based ('tip') or a branch-based ('branch') phylogeny. The default is 'branch'.
rooted	Whether to include the root of the phylogeny. The default is TRUE, meaning that the root is necessarily included in all phylogenies. This requires that the input tree be rooted.

Details

coreUniFrac calculates the UniFrac distance (Lozupone and Knight, 2005) between the core microbiomes from two different types of habitats based on either their tip-based or their branch-based core community phylogenies. In both cases, the UniFrac distance is calculated as the sum of the unique branch lengths in the core community phylogenies from each of the two habitats, divided by the total branch length of all branches in the core community phylogenies from the two habitats combined. For more details, see Bewick and Camper (2025).

Value

This function returns the numeric value of the UniFrac distance between two core microbiomes.

References

Bewick, S.A. and Benjamin T. Camper. "Phylogenetic Measures of the Core Microbiome" <doi:TBD>

Lozupone, Catherine, and Rob Knight. "UniFrac: a new phylogenetic method for comparing microbial communities." Applied and environmental microbiology 71.12 (2005): 8228-8235.

McMurdie, Paul J., and Susan Holmes. "phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data." PloS one 8.4 (2013): e61217.

McMurdie, Paul J., and Susan Holmes. "Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data." Biocomputing 2012. 2012. 235-246.

Examples

```
#Test with enterotype dataset
library(phyloseq)
library(ape)
library(phytools)
data(enterotype)

set.seed(1)

#Generate an example tree and label it with the names of the microbial taxa
enterotype_tree<-rtree(length(taxa_names(enterotype)))
enterotype_tree$tip.label<-taxa_names(enterotype)

#keep only those samples with gender identified
gendered<-which(!(is.na(sample_data(enterotype)$Gender)))
enterotypeMF<-prune_samples(sample_names(enterotype)[gendered],enterotype)

#Create a phyloseq object with a tree
example_phyloseq<-phyloseq(otu_table(enterotypeMF),phy_tree(as.phylo(enterotype_tree)))

coreUniFrac(example_phyloseq,grouping=sample_data(enterotypeMF)$Gender,0.5)
```

---

coreVennTree   *Phylogeny with Branches Colored Based on Venn Diagram*

Description

Plots a phylogeny with branches colored based on the compartment that they are associated with in the Venn diagram generated by the `corePhyloVenn` function.

Usage

```
coreVennTree(x, grouping, core_fraction,
mode = 'branch',rooted=TRUE, ordered_groups=NULL,branch_color=NULL,
remove_zeros=TRUE,plot.chronogram=FALSE)
```

Arguments

x	(Required) Microbial community data. This must be in the form of a phyloseq object and must contain, at a minimum, an OTU abundance table and a phylogeny.
grouping	(Required) A vector specifying which samples belong to which habitat type.
core_fraction	(Required) The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome.

mode	Whether to build a tip-based ('tip') or a branch-based ('branch') phylogeny. The default is 'branch'.
rooted	Whether to include the root of the phylogeny. The default is TRUE, meaning that the root is necessarily included in all phylogenies. This requires that the input tree be rooted.
ordered_groups	When provided, specifies the order in which different habitats should be plotted. This order is matched to the color order specified in `fill_color` (see below).
branch_color	A vector specifying what colors to use for branches associated with each of the different habitat combinations in the Venn diagram. This vector must be as long as the number of possible habitat combinations (number of compartments in the Venn diagram plus one for branches not included in the core of any habitats). When no colors are specified or a vector of the wrong length is specified, the default is to use a range of colors from blue to red.
remove_zeros	Whether or not to remove taxa that are missing from all samples prior to drawing the phylogeny. The default is TRUE.
plot.chronogram	
	Whether to plot a phylogeny or a chronogram. The default is FALSE.

Details

`coreVennTree` generates a phylogeny with branches colored according to the compartments of an associated Venn diagram as generated using the `coreVenn` function. For more details, see Bewick and Camper (2025).

Value

This function returns a color coded plot of the phylogeny. When a vector of colors is specified, the color key is printed out in the console.

References

Bewick, S.A. and Benjamin T. Camper. "Phylogenetic Measures of the Core Microbiome" <doi:TBD>

McMurdie, Paul J., and Susan Holmes. "phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data." PloS one 8.4 (2013): e61217.

McMurdie, Paul J., and Susan Holmes. "Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data." Biocomputing 2012. 2012. 235-246.

Examples

```
#Test with enterotype dataset
library(phyloseq)
library(ape)
library(phytools)
library(ggplot2)
data(enterotype)

set.seed(1)
```

```
#Generate an example tree and label it with the names of the microbial taxa
enterotype_tree<-rtree(length(taxa_names(enterotype)))
enterotype_tree$tip.label<-taxa_names(enterotype)

#keep only those samples with gender identified
gendered<-which(!(is.na(sample_data(enterotype)$Gender)))
enterotypeMF<-prune_samples(sample_names(enterotype)[gendered],enterotype)

#Create a phyloseq object with a tree
example_phyloseq<-phyloseq(otu_table(enterotypeMF),phy_tree(as.phylo(enterotype_tree)))

#Define the groups
bygender<-sample_data(enterotypeMF)$Gender

#Define the colors for group combinations
clist<-c('black','red','orange','yellow')

#Plot the tree
coreVennTree(example_phyloseq,grouping=bygender,0.5,branch_color=clist)
```

---

ggvenn2   *A modified version of functions from the ggvenn package for plotting*
                                 phylogenetic venn diagrams.

Description

A modified version of functions from the ggvenn package for plotting phylogenetic venn diagrams.

Usage

```
ggvenn2(
  data,
  columns = NULL,
  show_elements = FALSE,
  show_percentage = TRUE,
  digits = 1,
  fill_color = c("blue", "yellow", "green", "red"),
  fill_alpha = 0.5,
  stroke_color = "black",
  stroke_alpha = 1,
  stroke_size = 1,
  stroke_linetype = "solid",
  set_name_color = "black",
  set_name_size = 6,
  text_color = "black",
  text_size = 4,
  label_sep = ",",
  count_column = NULL,
```

```
  show_outside = c("auto", "none", "always"),
  auto_scale = FALSE,
  comma_sep = FALSE
)
```

Arguments

data	A data.frame or a list as input data.
columns	A character vector use as index to select columns/elements.
show_elements	Show set elements instead of count/percentage.
show_percentage	
	Show percentage for each set.
digits	The desired number of digits after the decimal point
fill_color	Filling colors in circles.
fill_alpha	Transparency for filling circles.
stroke_color	Stroke color for drawing circles.
stroke_alpha	Transparency for drawing circles.
stroke_size	Stroke size for drawing circles.
stroke_linetype	
	Line type for drawing circles.
set_name_color	Text color for set names.
set_name_size	Text size for set names.
text_color	Text color for intersect contents.
text_size	Text size for intersect contents.
label_sep	Separator character for displaying elements.
count_column	Specify column for element repeat count.
show_outside	Show outside elements (not belongs to any set).
auto_scale	Allow automatically resizing circles according to element counts.
comma_sep	Whether to use comma as separator for displaying numbers.

Value

The ggplot object to print or save to file.

References

Yan, Linlin, and Maintainer Linlin Yan. "Package "ggvenn"." (2021).

See Also

geom_venn, ggvenn

Examples

```
# use list as input
a <- list(`Set 1` = c(1, 3, 5, 7),
          `Set 2` = c(1, 5, 9),
          `Set 3` = c(1, 2, 8),
          `Set 4` = c(6, 7))
ggvenn2(a, c("Set 1", "Set 2"))
ggvenn2(a, c("Set 1", "Set 2", "Set 3"))
ggvenn2(a)
```

Index