

Package ‘holobiont’

May 8, 2026

Type Package

Title Microbiome Analysis Tools

Version 0.1.3

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.

License GPL-2

Encoding UTF-8

biocViews Software, phyloseq

Imports ape, phyloseq, phytools, ggplot2, dplyr, utils, tibble, castor, vegan, data.table

RoxygenNote 7.3.2

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 2025-07-21 22:50:02 UTC

Contents

| | |
|------------------------|---|
| basic_branch | 2 |
| basic_np | 3 |

| | |
|-------------------------|----|
| basic_tip | 4 |
| coreEdges | 5 |
| coreFaithsPD | 7 |
| coreJaccard | 9 |
| corePhyloVenn | 11 |
| coreRichness | 14 |
| coreTaxa | 15 |
| coreTree | 17 |
| coreUniFrac | 20 |
| coreVenn | 22 |
| coreVennTree | 24 |
| ggvenn2 | 27 |
| shade_branch | 29 |
| shade_np | 30 |
| shade_tip | 32 |

| | |
|--------------|-----------|
| Index | 34 |
|--------------|-----------|

| | |
|--------------|---|
| basic_branch | <i>Branch-based Core Community Phylogeny Using Thresholds</i> |
|--------------|---|

Description

Called internally. Identifies all edges and core edges of the core community phylogeny using a branch-based approach with core thresholds.

Usage

```
basic_branch(xv, nt, cf, abt1, abt2, abt3,rt)
```

Arguments

| | |
|------|---|
| xv | (Required) The phyloseq object passed from main functions and containing microbial community data. |
| nt | (Required) The microbial phylogeny passed from main functions. |
| cf | (Required) The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome. |
| abt1 | (Required) The threshold for mean abundance across samples. |
| abt2 | (Required) The threshold for the maximum abundance in any sample. |
| abt3 | (Required) The threshold for the minimum abundance across sample. |
| rt | 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

basic_branch is used internally in the holobiont package to identify core edges of a microbial phylogeny for the branch-based approach using thresholds.

Value

This function returns a list of all edges and core edges.

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)))

basic_branch(example_phyloseq, phy_tree(example_phyloseq), 0.5, 0, 0, 0,TRUE)
```

basic_np

Non-phylogenetic Core Community Phylogeny Using Thresholds

Description

Called internally. Identifies all edges and core edges of the core community phylogeny using a non-phylogenetic approach with core thresholds.

Usage

```
basic_np(xv, cf, abt1, abt2, abt3)
```

Arguments

| | |
|------|--|
| xv | (Required) The phyloseq object passed from main functions and containing microbial community data. |
| cf | (Required) The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome. |
| abt1 | (Required) The threshold for mean abundance across all samples. |
| abt2 | (Required) The threshold for maximum abundance in any sample. |
| abt3 | (Required) The threshold for the minimum abundance across all samples. |

Details

basic_np is used internally in the holobiont package to identify all taxa and all core taxa for the non-phylogenetic approach using thresholds.

Value

This function returns a list of all taxa and all core taxa.

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)))

basic_np(example_phyloseq, 0.5, 0, 0, 0)
```

basic_tip

Tip-based Core Community Phylogeny Using Thresholds

Description

Called internally. Identifies all edges and core edges of the core community phylogeny using a tip-based approach with core thresholds.

Usage

```
basic_tip(xv, nt, cf, abt1, abt2, abt3,rt)
```

Arguments

| | |
|------|--|
| xv | (Required) The phyloseq object passed from main functions and containing microbial community data. |
| nt | (Required) The microbial phylogeny passed from main functions. |
| cf | (Required) The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome. |
| abt1 | (Required) The threshold for mean abundance across all samples. |

| | |
|------|---|
| abt2 | (Required) The threshold for maximum abundance in any sample. |
| abt3 | (Required) The threshold for the minimum abundance across all samples. |
| rt | 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

basic_tip is used internally in the holobiont package to identify core edges of a microbial phylogeny for the tip-based approach using thresholds.

Value

This function returns a list of all edges and core edges, as well as all taxa and core taxa.

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)))

basic_tip(example_phyloseq, phy_tree(example_phyloseq), 0.5, 0, 0, 0,TRUE)
```

coreEdges

Edges of the Tree Included in the Core Community Phylogeny

Description

Finds the edges of a phylogenetic tree that are part of the core microbiome based on either the tip-based or the branch-based core community phylogeny.

Usage

```
coreEdges(x, core_fraction = 0.5, ab_threshold1 = 0,
ab_threshold2 = 0, ab_threshold3 = 0, mode='branch',
selection='basic', max_tax = NULL, increase_cutoff = 2,
initial_branches = NULL, rooted=TRUE, remove_zeros=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 | The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0.5. |
| ab_threshold1 | The threshold for mean relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold2 | The threshold for maximum relative abundance in any sample. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold3 | The threshold for minimum relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| mode | Whether to build a tip-based ('tip') or a branch-based ('branch') phylogeny. The default is 'branch'. |
| selection | Whether to use thresholds ('basic') or the Shade and Stopnisek method ('shade') to define the core community. The default is 'basic'. |
| max_tax | The maximum number of branches to add sequentially, as a percentage of the total branches when using the Shade and Stopnisek method. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| increase_cutoff | The threshold for the percent increase in beta diversity used to identify the taxon at which point adding more taxa yields diminishing returns in explanatory power. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| initial_branches | The number of branches to include prior to testing for increases in beta diversity. The default is to use all branches that are present in every sample and to begin testing branches that are missing from at least one sample. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| 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. |
| remove_zeros | Whether or not to remove taxa that are no reads associated across any of the samples. |

Details

coreEdges identifies the edges of the core community phylogeny based on either the tip-based or the branch-based core community phylogeny using either basic thresholds or a modification of the Shade and Stopnisek (2019) algorithm. For more details, see Bewick and Camper (2025).

Value

This function returns a list of the edges of the phylogenetic tree associated with the core microbiome.

References

Shade, Ashley, and Nejc Stopnisek. "Abundance-occupancy distributions to prioritize plant core microbiome membership." *Current opinion in microbiology* 49 (2019): 50-58.

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$tip.label<-taxa_names(enterotype)
enterotype_tree$node.label<-as.character(1:1:enterotype_tree$Nnode)

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

coreEdges(example_phyloseq)
```

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 = 0.5, ab_threshold1 = 0,
ab_threshold2 = 0, ab_threshold3 = 0, mode='branch',
selection='basic', max_tax = NULL, increase_cutoff = 2,
initial_branches = NULL, 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 | The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0.5. |
| ab_threshold1 | The threshold for mean relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold2 | The threshold for maximum relative abundance in any sample. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold3 | The threshold for minimum relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| mode | Whether to build a tip-based ('tip') or a branch-based ('branch') phylogeny. The default is 'branch'. |
| selection | Whether to use thresholds ('basic') or the Shade and Stopnisek method ('shade') to define the core community. The default is 'basic'. |
| max_tax | The maximum number of branches to add sequentially, as a percentage of the total branches when using the Shade and Stopnisek method. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| increase_cutoff | The threshold for the percent increase in beta diversity used to identify the taxon at which point adding more taxa yields diminishing returns in explanatory power. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| initial_branches | The number of branches to include prior to testing for increases in beta diversity. The default is to use all branches that are present in every sample and to begin testing branches that are missing from at least one sample. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| 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 using either basic thresholds or a modification of the Shade and Stopnisek (2019) algorithm. 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

Shade, Ashley, and Nejc Stopnisek. "Abundance-occupancy distributions to prioritize plant core microbiome membership." *Current opinion in microbiology* 49 (2019): 50-58.

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$tip.label<-taxa_names(enterotype)
enterotype_tree$node.label<-as.character(1:1:enterotype_tree$Nnode)

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

coreFaithsPD(example_phyloseq)
```

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 = 0.5, ab_threshold1 = 0,
ab_threshold2 = 0, ab_threshold3 = 0, selection='basic',
max_tax = NULL, increase_cutoff = 2)
```

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 | The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0.5. |

| | |
|-----------------|---|
| ab_threshold1 | The threshold for mean relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold2 | The threshold for maximum relative abundance in any sample. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold3 | The threshold for minimum relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| selection | Whether to use thresholds ('basic') or the Shade and Stopnisek method ('shade') to define the core community. The default is 'basic'. |
| max_tax | The maximum number of branches to add sequentially, as a percentage of the total branches when using the Shade and Stopnisek method. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| increase_cutoff | The threshold for the percent increase in beta diversity used to identify the taxon at which point adding more taxa yields diminishing returns in explanatory power. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |

Details

coreJaccard calculates the Jaccard distance (Jaccard, 1901 A and B) between the core microbiomes from two different types of habitats using either basic thresholds or a modification of the Shade and Stopnisek (2019) algorithm. 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.
- Shade, Ashley, and Nejc Stopnisek. "Abundance-occupancy distributions to prioritize plant core microbiome membership." *Current opinion in microbiology* 49 (2019): 50-58.

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)
```

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 = 0.5, ab_threshold1 = 0,
ab_threshold2 = 0, ab_threshold3 = 0, mode='branch', selection='basic', max_tax = NULL,
increase_cutoff = 2, initial_branches = NULL, 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 | The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0.5. |
| ab_threshold1 | The threshold for mean relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold2 | The threshold for maximum relative abundance in any sample. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold3 | The threshold for minimum relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| mode | Whether to build a tip-based ('tip') or a branch-based ('branch') phylogeny. The default is 'branch'. |
| selection | Whether to use thresholds ('basic') or the Shade and Stopnisek method ('shade') to define the core community. The default is 'basic'. |
| max_tax | The maximum number of branches to add sequentially, as a percentage of the total branches when using the Shade and Stopnisek method. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| increase_cutoff | The threshold for the percent increase in beta diversity used to identify the taxon at which point adding more taxa yields diminishing returns in explanatory power. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| initial_branches | The number of branches to include prior to testing for increases in beta diversity. The default is to use all branches that are present in every sample and to begin testing branches that are missing from at least one sample. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| 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.
- Shade, Ashley, and Nejc Stopnisek. "Abundance-occupancy distributions to prioritize plant core microbiome membership." *Current opinion in microbiology* 49 (2019): 50-58.

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)
```

```

enterotype_tree$node.label<-as.character(1:1:enterotype_tree$Nnode)

#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)

```

coreRichness

Richness of the Core Microbiome

Description

Calculates richness of a core microbiome using a non-phylogenetic approach.

Usage

```

coreRichness(x, core_fraction = 0.5, ab_threshold1 = 0, ab_threshold2 = 0,
ab_threshold3 = 0, selection='basic', max_tax = NULL, increase_cutoff = 2)

```

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 | The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0.5. |
| ab_threshold1 | The threshold for mean relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold2 | The threshold for maximum relative abundance in any sample. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold3 | The threshold for minimum relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| selection | Whether to use thresholds ('basic') or the Shade and Stopnisek method ('shade') to define the core community. The default is 'basic'. |
| max_tax | The maximum number of branches to add sequentially, as a percentage of the total branches when using the Shade and Stopnisek method. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |

increase_cutoff

The threshold for the percent increase in beta diversity used to identify the taxon at which point adding more taxa yields diminishing returns in explanatory power. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'.

Details

coreRichness calculates richness (a count of microbial taxa) using either basic thresholds or a modification of the Shade and Stopnisek (2019) algorithm.

Value

This function returns the numeric value of richness for the core microbiome.

References

Shade, Ashley, and Nejc Stopnisek. "Abundance-occupancy distributions to prioritize plant core microbiome membership." *Current opinion in microbiology* 49 (2019): 50-58.

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)))

coreRichness(example_phyloseq)
```

coreTaxa

Taxa Included in the Core Microbiome

Description

Finds the microbial taxa that are part of the core microbiome.

Usage

```
coreTaxa(x, core_fraction = 0.5, ab_threshold1 = 0, ab_threshold2 = 0,
ab_threshold3 = 0, selection='basic', max_tax = NULL,
increase_cutoff = 2, remove_zeros=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. |
| core_fraction | The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0.5. |
| ab_threshold1 | The threshold for mean relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold2 | The threshold for maximum relative abundance in any sample. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold3 | The threshold for minimum relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| selection | Whether to use thresholds ('basic') or the Shade and Stopnisek method ('shade') to define the core community. The default is 'basic'. |
| max_tax | The maximum number of taxa to add sequentially, as a percentage of the total taxa when using the Shade and Stopnisek method. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| increase_cutoff | The threshold for the percent increase in beta diversity used to identify the taxon at which point adding more taxa yields diminishing returns in explanatory power. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| remove_zeros | Whether or not to remove taxa that are no reads associated across any of the samples. |

Details

coreTaxa identifies the microbial taxa of the core microbiome based on either the tip-based or the branch-based core community phylogeny using either basic thresholds or a modification of the Shade and Stopnisek (2019) algorithm. For more details, see Bewick and Camper (2025).

Value

This function returns a list of the microbial taxa associated with the core microbiome.

References

Shade, Ashley, and Nejc Stopnisek. "Abundance-occupancy distributions to prioritize plant core microbiome membership." *Current opinion in microbiology* 49 (2019): 50-58.

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$tip.label<-taxa_names(enterotype)
enterotype_tree$node.label<-as.character(1:1:enterotype_tree$Nnode)

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

coreTaxa(example_phyloseq)
```

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 = 0.5, ab_threshold1 = 0,
ab_threshold2 = 0, ab_threshold3 = 0, mode='branch',
selection = 'basic', max_tax=NULL, increase_cutoff = 2,
initial_branches = NULL, NCcol = 'black', Ccol='red', rooted=TRUE,
branch.width=4, label.tips=FALSE, scaled = 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 | The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0.5. |

| | |
|------------------|---|
| ab_threshold1 | The threshold for mean relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold2 | The threshold for maximum relative abundance in any sample. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold3 | The threshold for minimum relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| mode | Whether to build a tip-based ('tip') or a branch-based ('branch') phylogeny. The default is 'branch'. |
| selection | Whether to use thresholds ('basic') or the Shade and Stopnisek method ('shade') to define the core community. The default is 'basic'. |
| max_tax | The maximum number of branches to add sequentially, as a percentage of the total branches when using the Shade and Stopnisek method. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| increase_cutoff | The threshold for the percent increase in beta diversity used to identify the taxon at which point adding more taxa yields diminishing returns in explanatory power. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| initial_branches | The number of branches to include prior to testing for increases in beta diversity. The default is to use all branches that are present in every sample and to begin testing branches that are missing from at least one sample. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| 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. |
| 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. |
| 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. |
| scaled | Whether or not to scale the branch lengths. 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." *Bioinformatics* 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 = 0.5, ab_threshold1 = 0,
ab_threshold2 = 0, ab_threshold3 = 0, mode='branch', selection='basic',
max_tax = NULL, increase_cutoff = 2, initial_branches = NULL, 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 | The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0.5. |
| ab_threshold1 | The threshold for mean relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold2 | The threshold for maximum relative abundance in any sample. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold3 | The threshold for minimum relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| mode | Whether to build a tip-based ('tip') or a branch-based ('branch') phylogeny. The default is 'branch'. |
| selection | Whether to use thresholds ('basic') or the Shade and Stopnisek method ('shade') to define the core community. The default is 'basic'. |
| max_tax | The maximum number of branches to add sequentially, as a percentage of the total branches when using the Shade and Stopnisek method. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| increase_cutoff | The threshold for the percent increase in beta diversity used to identify the taxon at which point adding more taxa yields diminishing returns in explanatory power. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |

| | |
|------------------|---|
| initial_branches | The number of branches to include prior to testing for increases in beta diversity. The default is to use all branches that are present in every sample and to begin testing branches that are missing from at least one sample. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| 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 and using either basic thresholds or a modification of the Shade and Stopnisek (2019) algorithm. 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.
- Shade, Ashley, and Nejc Stopnisek. "Abundance-occupancy distributions to prioritize plant core microbiome membership." *Current opinion in microbiology* 49 (2019): 50-58.

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$tip.label<-taxa_names(enterotype)
enterotype_tree$node.label<-as.character(1:1:enterotype_tree$Nnode)
```

```
#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)
```

| | |
|----------|--|
| coreVenn | <i>Venn Diagram Comparing (Non-Phylogenetic) Core Microbiomes from Multiple Habitats</i> |
|----------|--|

Description

Generates a Venn diagram comparing and contrasting the shared and unique taxa of core microbiomes from two or more (up to seven) different types of habitats using a non-phylogenetic approach.

Usage

```
coreVenn(x, grouping, core_fraction = 0.5, ab_threshold1 = 0,
ab_threshold2 = 0, ab_threshold3 = 0, selection='basic', max_tax = NULL,
increase_cutoff = 2,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 | The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0.5. |
| ab_threshold1 | The threshold for mean relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold2 | The threshold for maximum relative abundance in any sample. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |

| | |
|-----------------|---|
| ab_threshold3 | The threshold for minimum relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| selection | Whether to use thresholds ('basic') or the Shade and Stopnisek method ('shade') to define the core community. The default is 'basic'. |
| max_tax | The maximum number of branches to add sequentially, as a percentage of the total branches when using the Shade and Stopnisek method. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| increase_cutoff | The threshold for the percent increase in beta diversity used to identify the taxon at which point adding more taxa yields diminishing returns in explanatory power. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| 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

coreVenn generates a Venn diagram showing either the percentage of taxa or the taxon count shared between the core communities 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 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.
- Shade, Ashley, and Nejc Stopnisek. "Abundance-occupancy distributions to prioritize plant core microbiome membership." *Current opinion in microbiology* 49 (2019): 50-58.

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$tip.label<-taxa_names(enterotype)
enterotype_tree$node.label<-as.character(1:1:enterotype_tree$Nnode)

#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)))

coreVenn(example_phyloseq, grouping=sample_data(enterotypeMF)$Gender)
```

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 = 0.5, ab_threshold1 = 0,
ab_threshold2 = 0, ab_threshold3 = 0, mode='branch',
selection='basic', max_tax = NULL, increase_cutoff = 2, initial_branches = NULL,
rooted=TRUE, ordered_groups=NULL, branch_color=NULL, remove_zeros=TRUE,
plot.chronogram=FALSE, branch.width = 4, label.tips = FALSE, scaled = 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 | The fraction of samples that a microbial taxon must be found in to be considered part of the 'core' microbiome. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0.5. |
| ab_threshold1 | The threshold for mean relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold2 | The threshold for maximum relative abundance in any sample. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| ab_threshold3 | The threshold for minimum relative abundance across all samples. This variable is only used when selection = 'basic' and is ignored when selection = 'shade'. The default value is 0. |
| mode | Whether to build a tip-based ('tip') or a branch-based ('branch') phylogeny. The default is 'branch'. |
| selection | Whether to use thresholds ('basic') or the Shade and Stopnisek method ('shade') to define the core community. The default is 'basic'. |
| max_tax | The maximum number of branches to add sequentially, as a percentage of the total branches when using the Shade and Stopnisek method. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| increase_cutoff | The threshold for the percent increase in beta diversity used to identify the taxon at which point adding more taxa yields diminishing returns in explanatory power. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| initial_branches | The number of branches to include prior to testing for increases in beta diversity. The default is to use all branches that are present in every sample and to begin testing branches that are missing from at least one sample. This variable is only used when selection = 'shade' and is ignored when selection = 'basic'. |
| 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. |

| | |
|------------------------------|--|
| <code>ordered_groups</code> | When provided, specifies the order in which different habitats should be plotted. This order is matched to the color order specified in <code>fill_color</code> (see below). |
| <code>branch_color</code> | 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. |
| <code>branch.width</code> | The width to use when plotting the branches of the phylogeny. The default is 4. |
| <code>label.tips</code> | Whether or not to label the tips of the phylogeny with the microbial taxon names. The default is FALSE. |
| <code>scaled</code> | Whether or not to scale the branch lengths. The default is FALSE. |
| <code>remove_zeros</code> | Whether or not to remove taxa that are missing from all samples prior to drawing the phylogeny. The default is TRUE. |
| <code>plot.chronogram</code> | 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." *Bioinformatics* 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

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

shade_branch

Branch-based Core Community Phylogeny Using the Shade & Stopnisek (2019) algorithm.

Description

Called internally. Identifies all edges and core edges of the core community phylogeny using a branch-based approach with a modified Shade and Stopnisek (2019) algorithm.

Usage

```
shade_branch(xv, nt, mxtx, fi, st)
```

Arguments

| | |
|------|---|
| xv | (Required) The phyloseq object passed from main functions and containing microbial community data. |
| nt | (Required) The microbial phylogeny passed from main functions. |
| mxtx | (Required) The maximum number of branches to add sequentially, as a percentage of the total branches. |
| fi | (Required) The threshold for the percent increase in beta diversity used to identify the taxon at which point adding more taxa yields diminishing returns in explanatory power. |
| st | (Required) The number of branches to include prior to testing for increases in beta diversity. The default is to use all branches that are present in every sample and to begin testing for branches that are missing from at least one sample. |

Details

shade_branch is used internally in the holobiont package to identify core edges of a microbial phylogeny for the branch-based approach using a modified Shade and Stopnisek (2019) algorithm.

Value

This function returns a list of all edges and core edges.

References

Shade, Ashley, and Nejc Stopnisek. "Abundance-occupancy distributions to prioritize plant core microbiome membership." *Current opinion in microbiology* 49 (2019): 50-58. Briefly, branches are ranked according to abundance first and then occupancy (i.e., occupancy is more important). Next, starting from st branches with the highest occupancy/abundance, branches are added sequentially, up to a maximum number of branches. Each time a new branch is added, the mean beta-diversity (UniFrac) between each pair of samples is calculated and is then averaged across all samples. Next, total beta-diversity (full phylogeny) between pairs samples is calculated, and again averaged across all samples. Finally, the percent increase in total beta-diversity is calculated for each new branch added. A threshold is then selected based on the lowest ranked branch that yields a minimum percent increase in beta-diversity.

Examples

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

set.seed(1)

enterotype<-prune_taxa(taxa_names(enterotype)[2:21],enterotype)
enterotype<-prune_samples(sample_names(enterotype)[2:21],enterotype)

#Generate an example tree and label it with the names of the microbial taxa
enterotype_tree<-rtree(length(taxa_names(enterotype)))
enterotype_tree$tip.label<-taxa_names(enterotype)
enterotype_tree$node.label<-as.character(1:1:enterotype_tree$Nnode)

#Create a phyloseq object with a tree
example_phyloseq<-phyloseq(otu_table(enterotype),phy_tree(as.phylo(enterotype_tree)))
newtree<-bind.tip(phy_tree(example_phyloseq),tip.label='outgroup',edge.length=0.0001,position=0)

shade_branch(example_phyloseq, newtree, NULL, 2, NULL)
```

shade_np

Non-phylogenetic Core Community Using the Shade & Stopnisek (2019) algorithm.

Description

Called internally. Identifies all taxa and core taxa of the core community using a non-phylogenetic approach with a modified Shade and Stopnisek (2019) algorithm.

Usage

```
shade_np(xv, mxtx, fi)
```

Arguments

| | |
|-------------------|---|
| <code>xv</code> | (Required) The phyloseq object passed from main functions and containing microbial community data. |
| <code>mxtx</code> | (Required) The maximum number of branches to add sequentially, as a percentage of the total branches. |
| <code>fi</code> | (Required) The threshold for the percent increase in beta diversity used to identify the taxon at which point adding more taxa yields diminishing returns in explanatory power. |

Details

`shade_np` is used internally in the holobiont package to identify core edges of a microbial phylogeny for the non-phylogenetic approach using a modified Shade and Stopnisek (2019) algorithm. Briefly, taxa are ranked according to abundance first and then occupancy (i.e., occupancy is more important). Next, taxa are added sequentially, up to a maximum number of taxa. Each time a new taxon is added, the mean beta-diversity (Jaccard) between each pair of samples is calculated and is then averaged across all samples. Next, total beta-diversity (all taxa) between pairs samples is calculated, and again averaged across all samples. Finally, the percent increase in total beta-diversity is calculated for each new taxon added. A threshold is then selected based on the lowest ranked taxon that yields a minimum percent increase in beta-diversity.

Value

This function returns a list of all taxa and core taxa.

References

Shade, Ashley, and Nejc Stopnisek. "Abundance-occupancy distributions to prioritize plant core microbiome membership." *Current opinion in microbiology* 49 (2019): 50-58.

Examples

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

set.seed(1)

enterotype<-prune_taxa(taxa_names(enterotype)[2:21],enterotype)
enterotype<-prune_samples(sample_names(enterotype)[2:21],enterotype)

#Generate an example tree and label it with the names of the microbial taxa
enterotype_tree<-rtree(length(taxa_names(enterotype)))
enterotype_tree$tip.label<-taxa_names(enterotype)
enterotype_tree$node.label<-as.character(1:1:enterotype_tree$Nnode)

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

```
newtree<-bind.tip(phy_tree(example_phyloseq),tip.label='outgroup',edge.length=0.0001,position=0)
shade_np(example_phyloseq, NULL, 2)
```

| | |
|-----------|---|
| shade_tip | <i>Tip-based Core Community Phylogeny Using the Shade & Stopnisek (2019) algorithm.</i> |
|-----------|---|

Description

Called internally. Identifies all edges and core edges of the core community phylogeny using a tip-based approach with a modified Shade and Stopnisek (2019) algorithm.

Usage

```
shade_tip(xv, nt, mxtx, fi)
```

Arguments

| | |
|------|---|
| xv | (Required) The phyloseq object passed from main functions and containing microbial community data. |
| nt | (Required) The microbial phylogeny passed from main functions. |
| mxtx | (Required) The maximum number of branches to add sequentially, as a percentage of the total branches. |
| fi | (Required) The threshold for the percent increase in beta diversity used to identify the taxon at which point adding more taxa yields diminishing returns in explanatory power. |

Details

shade_tip is used internally in the holobiont package to identify core edges of a microbial phylogeny for the tip-based approach using a modified Shade and Stopnisek (2019) algorithm. Briefly, taxa are ranked according to abundance first and then occupancy (i.e., occupancy is more important). Next, taxa are added sequentially, up to a maximum number of taxa. Each time a new taxon is added, the mean beta-diversity (UniFrac) between each pair of samples is calculated and is then averaged across all samples. Next, total beta-diversity (all taxa) between pairs samples is calculated, and again averaged across all samples. Finally, the percent increase in total beta-diversity is calculated for each new taxon added. A threshold is then selected based on the lowest ranked taxon that yields a minimum percent increase in beta-diversity.

Value

This function returns a list of all edges and core edges, as well s all taxa and all core taxa.

References

Shade, Ashley, and Nejc Stopnisek. "Abundance-occupancy distributions to prioritize plant core microbiome membership." *Current opinion in microbiology* 49 (2019): 50-58.

Examples

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

set.seed(1)

enterotype<-prune_taxa(taxa_names(enterotype)[2:21],enterotype)
enterotype<-prune_samples(sample_names(enterotype)[2:21],enterotype)

#Generate an example tree and label it with the names of the microbial taxa
enterotype_tree<-rtree(length(taxa_names(enterotype)))
enterotype_tree$tip.label<-taxa_names(enterotype)
enterotype_tree$node.label<-as.character(1:1:enterotype_tree$Nnode)

#Create a phyloseq object with a tree
example_phyloseq<-phyloseq(otu_table(enterotype),phy_tree(as.phylo(enterotype_tree)))
newtree<-bind.tip(phy_tree(example_phyloseq),tip.label='outgroup',edge.length=0.0001,position=0)

shade_tip(example_phyloseq, newtree, NULL, 2)
```

Index

[basic_branch](#), [2](#)
[basic_np](#), [3](#)
[basic_tip](#), [4](#)

[coreEdges](#), [5](#)
[coreFaithsPD](#), [7](#)
[coreJaccard](#), [9](#)
[corePhyloVenn](#), [11](#)
[coreRichness](#), [14](#)
[coreTaxa](#), [15](#)
[coreTree](#), [17](#)
[coreUniFrac](#), [20](#)
[coreVenn](#), [22](#)
[coreVennTree](#), [24](#)

[ggvenn2](#), [27](#)

[shade_branch](#), [29](#)
[shade_np](#), [30](#)
[shade_tip](#), [32](#)