Package: TDbook (via r-universe)

October 12, 2024

Title Companion Package for the Book ``Data Integration, Manipulation and Visualization of Phylogenetic Trees'' by Guangchuang Yu (2022, ISBN:9781032233574, doi:10.1201/9781003279242)

Version 0.0.6

Description The companion package that provides all the datasets used in the book ``Data Integration, Manipulation and Visualization of Phylogenetic Trees'' by Guangchuang Yu (2022, ISBN:9781032233574, doi:10.1201/9781003279242).

Depends R (>= 3.5)

License Artistic-2.0

Encoding UTF-8

URL `https://www.amazon.com/Integration-Manipulation-Visualization-Phylogenetic-Computational-ebook/dp/B0B5NLZR1Z/`

BugReports `https://github.com/YuLab-SMU/TDbook/issues`

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.1

Repository https://yulab-smu.r-universe.dev

RemoteUrl https://github.com/yulab-smu/tdbook

RemoteRef HEAD

RemoteSha 60b2de7f8df45b440c93cc17513e868ee98c1496

Contents

`data-AA_sequence` *Data set containing multiple sequence alignment information*

Description

This data set contains multiple sequence alignment information to be attached to the tree. Chap.7.4

Format

AAbin

Examples

```
data(AA_sequence)
```

`data-df_alleles.csv` *Allele table*

Description

The allele table with original data to be processed to SNP data. Then attached to the tree after visualizing.
Chap.7.2

Format

A data frame contains 386 rows and 385 variables, with first row of tips labels.

Source

`https://raw.githubusercontent.com/katholt/plotTree/master/tree_example_april2015/alleles.csv`

Examples

```
data(df_alleles)
```

`data-df_alltax_info` *Data frame containing hierarchical relationship*

Description

This Data frame contains hierarchical relationship to be visualized with another data frame that contains taxa and factor information and pvalue, using `ggdiffcalde()` of The `MicrobiotaProcess` package.
It's the analysis result of `diff_analysis()` using public datasets (Kostic et al. 2012) with The `MicrobiotaProcess` package.
These data frame are used to demonstrates how to visualize the differential taxonomy tree with The `MicrobiotaProcess` package.
Chap.11.1

Format

A data frame to store taxonomy information, contains 1351 rows and 7 variables (classification levels).

References

Kostic, Aleksandar D, Dirk Gevers, Chandra Sekhar Pedamallu, Monia Michaud, Fujiko Duke, Ashlee M Earl, Akinyemi I Ojesina, et al. Genomic Analysis Identifies Association of Fusobacterium with Colorectal Carcinoma." Genome Research, Volume 22, Issue 2, 2012, Pages 292–98. doi:10.1101/gr.126573.111

Examples

```
data(df_alltax_info)
```

`data-df_barplot_attr` *Data set of the abundance of microbes at the body sites of greatest prevalence*

Description

The data set of abundance of microbes at the body sites of greatest prevalence to be visualized using barplot and attached to the tree.
Chap.10.3

Format

A data frame contains 332 rows and 3 variables, with a column of "node" or first column of taxa labels.

References

Asnicar, Francesco, George Weingart, Timothy L Tickle, Curtis Huttenhower, and Nicola Segata. Compact Graphical Representation of Phylogenetic Data and Metadata with GraPhlAn. PeerJ, 3: e1029, June 2015. doi:10.7717/peerj.1029

Examples

```
data(df_barplot_attr)
```

`data-df_bar_data` *Trait data*

Description

The trait data to be visualized with bar plot and attached to the tree.
Chap.7.2

Format

A data frame contains 386 rows and 2 variables, with a column of "node" or first column of taxa labels.

Source

`https://raw.githubusercontent.com/katholt/plotTree/master/tree_example_april2015/bar.csv`

Examples

```
data(df_bar_data)
```

`data-df_Candidaauris_data` *Population genetics data No.1*

Description

Population genetics data to be visualized using multiple graphs and attached to tree No.1 with `geo_fruit()` of `ggtreeExtra`.
Chap.10.4

Format

A data frame contains 305 rows and 20 variables, with a column of "node" or first column of taxa labels.

References

Chow, Nancy A., José F. Muñoz, Lalitha Gade, Elizabeth L. Berkow, Xiao Li, Rory M. Welsh, Kaitlin Forsberg, et al. Tracing the Evolutionary History and Global Expansion of Candida Auris Using Population Genomic Analyses. Edited by Geraldine Butler and Kirsten Nielsen. mBio, Volume 11, Issue 2, April 2020. doi:10.1128/mBio.0336419

Examples

```
data(df_Candidaauris_data)
```

`data-df_difftax` *Data frame containing taxa and factor information and pvalue*

Description

This Data frame contains taxa and factor information and pvalue to be visualized with another data frame that contains hierarchical relationship, using `ggdiffcalde()` of The `MicrobiotaProcess` package.
These data frame are used to demonstrates how to visualize the differential taxonomy tree with The `MicrobiotaProcess` package.
Chap.11.1

Format

A data frame contains 36 rows and 3 variables.

References

Kostic, Aleksandar D, Dirk Gevers, Chandra Sekhar Pedamallu, Monia Michaud, Fujiko Duke, Ashlee M Earl, Akinyemi I Ojesina, et al. Genomic Analysis Identifies Association of Fusobacterium with Colorectal Carcinoma." Genome Research, Volume 22, Issue 2, 2012, Pages 292–98. doi:10.1101/gr.126573.111

Examples

```
data(df_difftax)
```

`data-df_info` *Sampling information data set*

Description

The sampling information data set be attached to the tree tip.
Chap.7.2

Format

A data frame contains 386 rows and 6 variables, with a column of "node" or first column of taxa labels.

Source

`https://raw.githubusercontent.com/katholt/plotTree/master/tree_example_april2015/info.csv`

Examples

```
data(df_info)
```

`data-df_inode_data` *Nodedata to be mapped to tree*

Description

This data contains annotation information about nodes of "tree_boots.nwk", that is to be mapped to the tree using the `%<+%` operator.
Chap.7.1

Format

A data frame with 6 rows and 6 variables with a column of node or first column of taxa labels.

Source

https://raw.githubusercontent.com/TreeViz/metastyle/master/design/viz_targets_exercise/inode_data.csv

Examples

```
data(df_inode_data)
```

`data-df_NJIDqgsS` *Population genetics data No.2*

Description

Population genetics data to be visualized using multiple graphs and attached to tree No.2 with `geo_fruit()` of `ggtreeExtra`.
Chap.10.4

Format

A data frame contains 1351 rows and 7 variables, with a column of "node" or first column of taxa labels.

References

Wong, Vanessa K., Stephen Baker, Derek J. Pickard, Julian Parkhill, Andrew J. Page, Nicholas A. Feasey, Robert A. Kingsley, et al. Phylogeographical Analysis of the Dominant Multidrug-Resistant H58 Clade of Salmonella Typhi Identifies Inter- and Intracontinental Transmission Events. Nature Genetics, Volume 47, Issue 6, May 2015, Pages 632–39. doi:10.1038/ng.3281

Examples

```
data(df_NJIDqgsS)
```

`data-df_ring_heatmap` *Data set of the abundance and types of microbes*

Description

The data set of abundance of microbes at different body sites to be visualized using heatmap and attached to the tree.
Chap.10.3

Format

A data frame contains 2324 rows and 3 variables, with a column of "node" or first column of taxa labels.

References

Asnicar, Francesco, George Weingart, Timothy L Tickle, Curtis Huttenhower, and Nicola Segata. Compact Graphical Representation of Phylogenetic Data and Metadata with GraPhlAn. PeerJ, 3: e1029, June 2015. doi:10.7717/peerj.1029

Examples

```
data(df_ring_heatmap)
```

`data-df_svl` *Data frame containing slv information*

Description

Data frame containing slv information to scale the color of the tree using the `continuous` parameter of `ggtree` to plot continuous state transition in edges.
Chap.4.3.5. This file is provided in `http://www.phytools.org/eqg2015/asr.html`.

Format

A data frame contains 100 rows and 1 variables.

Source

`http://www.phytools.org/eqg2015/data/svl.csv`

Examples

```
data(df_svl)
```

`data-df_tippoint` *Data set of the abundance and types of microbes*

Description

The data set of abundance and types of microbes to be attached to the tree tips, provided by GraPhlAn (Asnicar et al. 2015).
Chap.10.3

Format

A data frame contains 332 rows and 4 variables,with a column of "node" or first column of taxa labels.

References

Asnicar, Francesco, George Weingart, Timothy L Tickle, Curtis Huttenhower, and Nicola Segata. Compact Graphical Representation of Phylogenetic Data and Metadata with GraPhlAn. PeerJ, 3: e1029, June 2015. doi:10.7717/peerj.1029

Examples

```
data(df_tippoint)
```

`data-df_tip_data` *Tipdata to be mapped to tree*

Description

This data contains annotation information about tips of `tree_boots`, that is to be mapped to the tree using the `%<+%` operator.
Chap.7.1

Format

A data frame contains 7 rows and 9 variables, with a column of "node" or first column of taxa labels.

Source

`https://raw.githubusercontent.com/TreeViz/metastyle/master/design/viz_targets_exercise/tip_data.csv`

Examples

```
data(df_tip_data)
```

`data-dna_HPV58_aln` *A DNAbin class to store the aligned sequnces of species of HPV58.tree*

Description

A DNAbin class to store the aligned sequnces of species of HPV58.tree

Format

S3 class:DNAbin

References

Chen, Zigui, Wendy C. S. Ho, Siaw Shi Boon, Priscilla T. Y. Law, Martin C. W. Chan, Rob DeSalle, Robert D. Burk, and Paul K. S. Chan. Ancient Evolution and Dispersion of Human Papillomavirus 58 Variants. Journal of Virology, Volume 91, Issue 21, November 2017, e01285–17. doi:10.1128/JVI.0128517

Examples

```
data(dna_HPV58_aln)
```

`data-text_RMI_tree` *Tree to display with symbolic points indicating partitioned bootstrap values*

Description

This tree is to display with symbolic points with different color or shape indicating the bootstrap values belong to which category.
Catagories are defined by cutting the bootstrap values into several interval to indicate whether the clade is high, moderate or low support.
Chap.13.2

Format

character

Examples

```
data(text_RMI_tree)
```

`data-tree_anole` *Tree to color with continuous state transition in edges*

Description

This tree is used to demonstrate the functionality of plotting continuous state transition in edges using the `continuous` parameter of `ggtree`.
Chap.4.3.5. This file is provided in `http://www.phytools.org/eqg2015/asr.html`.

Format

S3 class:phylo

Source

`http://www.phytools.org/eqg2015/data/anole.tre`

Examples

```
data(tree_anole)
```

`data-tree_boots` *Tree to be mapped to data*

Description

This tree is used to demonstrate the functionality of the `%<+%` operator to attach annotation data to a ggtree graphic object.
Chap.7.1

Format

S3 class:phylo

Source

`https://raw.githubusercontent.com/TreeViz/metastyle/master/design/viz_targets_exercise/tree_boots.nwk`

Examples

```
data(tree_boots)
```

`data-tree_Candidaauris` *Tree No.1 to display with visualized population genetics data*

Description

Tree No.1 to display with visualized population genetics data for demonstrating the utility of `ggtreeExtra`.
Reproduction of Fig.4 of (Chow et al. 2020).
Chap.10.4

Format

S3 class:phylo

References

Chow, Nancy A., José F. Muñoz, Lalitha Gade, Elizabeth L. Berkow, Xiao Li, Rory M. Welsh, Kaitlin Forsberg, et al. Tracing the Evolutionary History and Global Expansion of Candida Auris Using Population Genomic Analyses. Edited by Geraldine Butler and Kirsten Nielsen. mBio, Volume 11, Issue 2, April 2020. doi:10.1128/mBio.0336419

Examples

```
data(tree_Candidaauris)
```

`data-tree_hmptree` *Tree to display with multiple graphs for multi-dimensional data*

Description

This tree is to display with multiple graphs, and by reproducing Fig.2 of (Morgan, Segata, and Huttenhower 2013), demonstrates the feature of adding multiple layers to present different types of data by `ggtreeExtra`.
Chap.10.3

Format

S3 class:phylo

References

Morgan, Xochitl C., Nicola Segata, and Curtis Huttenhower. Biodiversity and Functional Genomics in the Human Microbiome. Trends in Genetics Volume 29, Issue 1, 2013 January, Pages 51–58. doi:10.1016/J.TIG.2012.09.005

Examples

```
data(tree_hmptree)
```

`data-tree_HPV58` *Tree to display with dot and line plots of pairwise nucleotide sequence distances*

Description

This tree is to provide accession numbers from its tip labels, which are used to calculates pairwise nucleotide sequence distances. The distance matrix then is visualized as dot and line plots and attached to the tree, demonstrating the abilities of `facet_plot()` of ggtree that can add multiple layers to a specific panel.
Reproduction fo Fig. 1 of (Chen et al. 2017).
Chap.13.1

Format

S3 class:phylo

References

Chen, Zigui, Wendy C. S. Ho, Siaw Shi Boon, Priscilla T. Y. Law, Martin C. W. Chan, Rob DeSalle, Robert D. Burk, and Paul K. S. Chan. Ancient Evolution and Dispersion of Human Papillomavirus 58 Variants. Journal of Virology, Volume 91, Issue 21, November 2017, e01285–17. doi:10.1128/JVI.0128517

Examples

```
data(tree_HPV58)
```

`data-tree_long_branch_example` *Tree to display with shrunk outlier long branch*

Description

This tree is to display with display with shrunk outlier long branch to demonstrate how to shrink outlier long branch with `ggtree` by modifying coordination of the out groups, or with the `ggbreak` package.
Appendix.A.5.3

Format

S3 class:Phylo

Examples

```
data(tree_long_branch_example)
```

`data-tree_nex` *Tree to display with silhouettes from Phylopic*

Description

This tree is to display with display with silhouettes from Phylopic, demonstrating the feature of `ggtree` that will automatically download phylopic figures by querying provided UID.
Chap.8.2

Format

S3 class:Phylo

Examples

```
data(tree_nex)
```

`data-tree_NJIDqgsS` *Tree No.2 to display with visualized population genetics data*

Description

Tree No.2 to display with visualized population genetics data for demonstrating the utility of `ggtreeExtra`.
Reproduction of Fig 1 of (Wong et al. 2015).
Chap.10.4

Format

S3 class:phylo

References

Wong, Vanessa K., Stephen Baker, Derek J. Pickard, Julian Parkhill, Andrew J. Page, Nicholas A. Feasey, Robert A. Kingsley, et al. Phylogeographical Analysis of the Dominant Multidrug-Resistant H58 Clade of Salmonella Typhi Identifies Inter- and Intracontinental Transmission Events. Nature Genetics, Volume 47, Issue 6, May 2015, Pages 632–39. doi:10.1038/ng.3281

Examples

```
data(tree_NJIDqgsS)
```

`data-tree_nwk` *Tree to display with sampling information, SNP and Trait data*

Description

The tree to display with sampling information, attached with dot chart of SNP and bar chart of Trait data.
Chap.7.2

Format

S3 class:phylo

Source

`https://raw.githubusercontent.com/katholt/plotTree/master/tree_example_april2015/tree.nwk`

Examples

```
data(tree_nwk)
```

`data-tree_seq_nwk` *Tree to display with multiple sequence alignment*

Description

This tree is used to display with multiple sequence alignment.
Chap.7.4

Format

S3 class:phylo

Examples

```
data(tree_seq_nwk)
```

`data-tree_treenwk_30.4.19` *Tree to group and highlight*

Description

This tree is used to reproduce Figure 1 of (Larsen et al. 2019) with grouping, highlighting and labelling.
Chap.13.3

Format

S3 class:Phylo

References

Larsen, Frederik T., Bertrand Bed'Hom, Bernt Guldbrandtsen, and Tina S. Dalgaard. Identification and Tissue-Expression Profiling of Novel Chicken c-Type Lectin-Like Domain Containing Proteins as Potential Targets for Carbohydrate-Based Vaccine Strategies. Molecular Immunology, Volume 114, Issue October, October 2019, pages 216–25. doi:10.1016/j.molimm.2019.07.022

Examples

```
data(tree_treenwk_30.4.19)
```

Index