

Package ‘cloodl’

April 10, 2025

Title Fetch and Explore the Cornell Lab of Ornithology Open Tree of Life Avian Phylogeny

Version 0.1.1

Maintainer Eliot Miller <cloodlmaintainers@gmail.com>

URL <https://github.com/eliotmiller/cloodl>

BugReports <https://github.com/eliotmiller/cloodl/issues>

Depends R (>= 4.3.0), ape

Imports stats, dplyr, RCurl, jsonlite

LazyData true

LazyDataCompression xz

Description Fetches the Cornell Lab of Ornithology Open Tree of Life (cloodl) tree in a specified taxonomy. Optionally prune it to a given set of study taxa. Provide a recommended citation list for the studies that informed the extracted tree. Tree generated as described in McTavish et al. (2024) <[doi:10.1101/2024.05.20.595017](https://doi.org/10.1101/2024.05.20.595017)>.

License GPL-3

Encoding UTF-8

RoxygenNote 7.3.2

Suggests rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Eliot Miller [aut, cre],
Emily Jane McTavish [aut],
Luna L. Sanchez Reyes [ctb, aut]

Repository CRAN

Date/Publication 2025-04-10 14:40:02 UTC

Contents

cloodl_data	2
extractTree	3
getCitations	4
get_avesdata_repo	5
sampleTrees	6
set_avesdata_repo_path	7
taxonomyGet	8
treeGet	9

Index	10
--------------	-----------

cloodl_data	<i>A complex data store used in the package.</i>
-------------	--

Description

Taxonomy files and phylogenies.

Usage

cloodl_data

Format

List of csv files and Newick and Nexus phylogenies

Details

```

cloodl_data = list()
fullTree2021 <- treeGet("1.4", "2021", data_path = "~/projects/otapi/AvesData") fullTree2022 <- treeGet("1.4", "2022",
data_path = "~/projects/otapi/AvesData") fullTree2023 <- treeGet("1.4", "2023", data_path = "~/projects/otapi/AvesData")
cloodl_data$trees$Aves_1.4$summary.trees$year2021 <- fullTree2021 cloodl_data$trees$Aves_1.4$summary.trees$year2022
<- fullTree2022 cloodl_data$trees$Aves_1.4$summary.trees$year2023 <- fullTree2023
cloodl_data$versions <- c('1.2', '1.3', '1.4') tax2021 <- taxonomyGet(2021, data_path = "~/projects/otapi/AvesData")
tax2022 <- taxonomyGet(2022, data_path = "~/projects/otapi/AvesData") tax2023 <- taxonomyGet(2023,
data_path = "~/projects/otapi/AvesData")
cloodl_data$taxonomy.files$Year2021 <- tax2021 cloodl_data$taxonomy.files$Year2022 <- tax2022
cloodl_data$taxonomy.files$Year2023 <- tax2023
cloodl_data$tax_years <- c("2021", "2022", "2023") annot_filename <- "~/projects/otapi/AvesData/Tree_versions/Aves_1.4/O
all_nodes <- jsonlite::fromJSON(txt=annot_filename)
cloodl_data$trees$Aves_1.4$annotations <- all_nodes
This part pre-processes all citations for studies in the tree so we don't need to do any API calls later.
studies <- c() for (inputs in all_nodes$source_id_map) studies <- c(studies, inputs$study_id) studies
<- unique(studies) study_info <- cloodl:::api_studies_lookup(studies)
cloodl_data$study_info <- study_info save(cloodl_data, file = "~/projects/otapi/cloodl/data/cloodl_data.rda",
compress = "xz")

```

Source

<https://github.com/eliotmiller/clootl>

extractTree

Extract a tree from the complete Avian Phylogeny for a set of species

Description

Extract a tree from the complete Avian Phylogeny for a set of species

Usage

```
extractTree(
  species = "all_species",
  label_type = "scientific",
  taxonomy_year = 2023,
  version = "1.4",
  data_path = FALSE
)
```

Arguments

species	A character vector either of scientific names (directly as they come out of the eBird taxonomy, i.e. without underscores) or of six-letter eBird species codes. Any elements of the species vector that do not match a species-level taxon in the specified eBird taxonomy will result in an error. Default is set to "all_species".
label_type	Either "scientific" or "code". Default is set to "scientific".
taxonomy_year	The eBird taxonomy year the tree should be output in. Current options include 2021, 2022, and 2023. Both numeric and character inputs are acceptable here. Any value aside from these years will result in an error. Default is set 2023.
version	The desired version of the tree. Default to the most recent version of the tree. Other versions available are '1.2', '1.3', '1.4', and can be passed as a character string or as numeric.
data_path	Default to FALSE, it will look for a path containing the bird tree. If the tree has not been downloaded yet using <code>get_avesdata_repo()</code> , it will load the default tree using <code>utils::data()</code> as long as version and taxonomy_year are empty or match the default version. If the tree has been downloaded using <code>get_avesdata_repo()</code> , it will read the tree file corresponding to the version and taxonomy_year provided and load it as a phylo object.

Details

This function first ensures that the requested output species overlap with species-level taxa in the requested eBird taxonomy. If they do not, the function will error out. The onus is on the user to ensure the requested taxa are valid. This is critical to ensure no unexpected analysis hiccups later—you don't want to find out many steps later that your dataset doesn't match your phylogeny. The

eBird database is currently (as of Mar 2025) in 2024 taxonomy. Trees available in 2024 taxonomy will be available by June 2025. The 2025 taxonomy will be released to the public in October or November 2025. The intention is to release a tree in 2025 taxonomy concurrently with the publication of the taxonomy itself.

Value

A phylogeny of the specified taxa in the specified eBird taxonomy version and clootl tree version.

Author(s)

Eliot Miller, Luna Sanchez Reyes, Emily Jane McTavish

Examples

```
ex1 <- extractTree(species=c("amerob", "canwar", "reevir1", "yerwar", "gockin"),
  label_type="code")
ex2 <- extractTree(species=c("Turdus migratorius",
  "Setophaga dominica",
  "Setophaga ruticilla",
  "Sitta canadensis"),
  label_type="scientific",
  taxonomy_year="2021",
  version="1.4")
```

getCitations

Get the DOIs and quantify the contribution of published studies

Description

Standing on the shoulders of giants

Usage

```
getCitations(tree, version = "1.4", data_path = FALSE)
```

Arguments

tree	A phylogeny obtained from extractTree (see details).
version	The desired version of the tree. Default to the most recent version of the tree. Other versions available are '0.1','1.0','1.2','1.3','1.4' and can be passed as a character string or as numeric.
data_path	Default to FALSE, it will look for a path containing the bird tree. If the tree has not been downloaded yet using get_avesdata_repo() , it will load the default tree using utils::data() and version and taxonomy_year will be ignored?? If the tree has been downloaded using get_avesdata_repo() , it will read the tree file corresponding to the version and taxonomy_year provided and load it as a phylo object.

Details

Importantly: an internet connection is required for this function to work, as it relies on Open Tree of Life APIs. The function will determine what proportion of nodes in your phylogeny (possibly but not necessarily pruned to a set of study taxa) are supported by each study that goes into creating the final cloodl tree. In any resulting publication, you should always cite the cloodl tree, and you should also "always" cite all the trees/DOIs that contributed to your phylogeny. That said, we are well aware of citation and word count limits that plague modern publishing, and for this reason we quantify the contribution of each study; depending on your phylogeny, it is very possible that one or two studies contributed the majority of information. Currently, this function assumes your output tree matches the taxonomy of the corresponding tree on the OpenTree server. Since the function is actually using the named internal nodes for the API query, and these should not be lost between tree versions and taxonomies, this should not matter, but this has not yet been tested.

Value

A dataframe of the percent of internal nodes supported by a given study, as well as the DOI of that study. The proportion of taxa in the tree supported by taxonomic addition only is included in the dataframe.

Author(s)

Eliot Miller, Emily Jane McTavish

Examples

```
#pull the taxonomy file out
data(cloodl_data)
tax <- cloodl_data$taxonomy.files$Year2021
ls(tax)
#subset to species only
# TODO: this step seems no longer necessary, is it??
# tax <- tax[tax$CATEGORY=="species",]

#simulate extracting a tree for a particular family
temp <- tax[tax$FAMILY=="Rhinocryptidae (Tapaculos)",]
spp <- temp$SCI_NAME

#get your tree
prunedTree <- extractTree(species=spp, label_type="scientific",
  taxonomy_year=2021, version="1.4")

#get your citation DF
yourCitations <- getCitations(tree=prunedTree)
```

Description

Pull down full AvesData repository to a working directory

Usage

```
get_avesdata_repo(path, overwrite = FALSE)
```

Arguments

path	Path to download data zipfile to, and where it will be unpacked. To download into your working directory, use "."
overwrite	Default to FALSE. Will not redownload the data by default if path exists, unless overwrite=TRUE

Value

No return value, called to download the Aves Data repository.

sampleTrees	<i>Extract a cloud of trees from the complete Avian Phylogeny for a set of species</i>
-------------	--

Description

Extract a cloud of trees from the complete Avian Phylogeny for a set of species

Usage

```
sampleTrees(
  species = "all_species",
  label_type = "scientific",
  taxonomy_year = 2023,
  version = "1.4",
  count = 100,
  data_path = FALSE
)
```

Arguments

species	A character vector either of scientific names (directly as they come out of the eBird taxonomy, i.e. without underscores) or of six-letter eBird species codes. Any elements of the species vector that do not match a species-level taxon in the specified eBird taxonomy will result in an error. Default is set to "all_species".
label_type	Either "scientific" or "code". Default is set to "scientific".
taxonomy_year	The eBird taxonomy year the tree should be output in. Current options include 2021, 2022, and 2023. Both numeric and character inputs are acceptable here. Any value aside from these years will result in an error. Default is set 2023.

version	The desired version of the tree. Default to the most recent version of the tree. Other versions available are '1.2','1.3','1.4', and can be passed as a character string or as numeric.
count	Work in progress, can only sample 100 for now. Eventually: The desired number of sampled trees.
data_path	Default to FALSE, it will look for a path containing the bird tree. If the tree has been downloaded using <code>get_avesdata_repo()</code> , it will read the tree file corresponding to the version and taxonomy_year provided and load it as a phylo object.

Details

This function first ensures that the requested output species overlap with species-level taxa in the requested eBird taxonomy. If they do not, the function will error out. The onus is on the user to ensure the requested taxa are valid. This is critical to ensure no unexpected analysis hiccups later—you don't want to find out many steps later that your dataset doesn't match your phylogeny. The eBird database is currently (as of Mar 2025) in 2024 taxonomy. Trees available in 2024 taxonomy will be available by June 2025. The 2025 taxonomy will be released to the public in October or November 2025. The intention is to release a tree in 2025 taxonomy concurrently with the publication of the taxonomy itself.

Value

A set of phylogenies determined in count of the specified taxa in the specified eBird taxonomy version and clootl tree version.

Author(s)

Eliot Miller, Luna Sanchez Reyes, Emily Jane McTavish

Examples

```
if (Sys.getenv("AVESDATA_PATH") != "") {
  ex2 <- sampleTrees(species=c("Turdus migratorius",
                              "Setophaga dominica",
                              "Setophaga ruticilla",
                              "Sitta canadensis"))
}
```

set_avesdata_repo_path

*Set path to Aves Data folder already
somewhere on your computer Based on
[https://github.com/CornellLabofOrnithology/auk/blob/main/R/auk-
set-ebd-path.r](https://github.com/CornellLabofOrnithology/auk/blob/main/R/auk-set-ebd-path.r)*

Description

Set path to Aves Data folder already somewhere on your computer Based on <https://github.com/CornellLabofOrnithology/auk>
 set-ebd-path.r

Usage

```
set_avesdata_repo_path(path, overwrite = FALSE)
```

Arguments

path	A character vector with the path to the Aves Data folder.
overwrite	Boolean, default to FALSE, does not overwrite an existing Aves Data folder. Set to TRUE to overwrite.

Value

No return value, called to set the path to the Aves Data folder.

Examples

```
## Not run:
set_avesdata_repo_path("/home/ejmctavish/AvesData")

## End(Not run)
```

taxonomyGet	<i>Load a bird taxonomy into the R environment</i>
-------------	--

Description

taxonomyGet either reads a taxonomy file and loads it as a data frame, or loads the default taxonomy data object.

Usage

```
taxonomyGet(taxonomy_year, data_path = FALSE)
```

Arguments

taxonomy_year	The eBird taxonomy year the tree should be output in. Current options include 2021, 2022, and 2023. Both numeric and character inputs are acceptable here. Any value aside from these years will result in an error. Default is set 2023.
data_path	Default to FALSE, it will look for a path containing the bird taxonomy. If the taxonomy has not been downloaded yet using <code>get_avesdata_repo()</code> , it will load the default taxonomy using <code>utils::data()</code> as long as taxonomy_year as empty or matches the default version. If the taxonomy has been downloaded using <code>get_avesdata_repo()</code> , it will read the taxonomy file corresponding to the year given in taxonomy_year and load it as a data frame object.

Details

This will return a data object that has the taxonomy of the requested year.

Value

A `data.frame` with 17 columns of taxonomic information: order, species code, taxon concept, common name, scientific name, family, OpenTree Taxonomy data, etc.

treeGet

Load a bird tree into the R environment

Description

treeGet either reads a tree file and loads it as a phylo object, or loads the default tree data object.

Usage

```
treeGet(version, taxonomy_year, data_path = FALSE)
```

Arguments

version	The desired version of the tree. Default to the most recent version of the tree. Other versions available are '1.2','1.3','1.4', and can be passed as a character string or as numeric.
taxonomy_year	The eBird taxonomy year the tree should be output in. Current options include 2021, 2022, and 2023. Both numeric and character inputs are acceptable here. Any value aside from these years will result in an error. Default is set 2023.
data_path	Default to FALSE, it will look for a path containing the bird tree. If the tree has not been downloaded yet using get_avesdata_repo() , it will load the default tree using <code>utils::data()</code> and version and taxonomy_year will be ignored?? If the tree has been downloaded using get_avesdata_repo() , it will read the tree file corresponding to the version and taxonomy_year provided and load it as a phylo object.

Details

This will return a data object that has the requested tree.

Value

A phylo object with the requested version and taxonomy.

Index

* datasets

 cloodl_data, 2

cloodl_data, 2

extractTree, 3

get_avesdata_repo, 5

get_avesdata_repo(), 3, 4, 7–9

getCitations, 4

sampleTrees, 6

set_avesdata_repo_path, 7

taxonomyGet, 8

treeGet, 9

utils::data(), 3, 4, 8, 9