

Package ‘cloodl’

October 29, 2025

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

Version 0.1.2

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 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 2025-10-29 06:00:02 UTC

Contents

cloodl_data	2
extractTree	4
getCitations	5
get_avesdata_repo	7
sampleTrees	8
set_avesdata_repo_path	9
taxonomyGet	10

Index	12
--------------	-----------

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

Description

A dataset containing taxonomy files, summary phylogenies, constituent study information, and other data needed for the package to function properly.

Usage

cloodl_data

Format

List of csv files, phylogenies, and other data components.

Details

The data object, `cloodl_data`, stores the most up-to-date stable version of the tree mapped to each of the different taxonomy years, the annotations of how each study contributed to the tree, the citation information for each study that contributed to the tree, the taxonomy crosswalks for different years, and some other variables.

The structure of the data store (a list) is as follows:

`cloodl_data$taxonomy.files` A list of data frames. Each element corresponds to a taxonomy year:

- Year2024
- Year2023
- Year2022
- Year2021

These originate as CSV files linking the Clements taxonomy for each of these years to OTT ids, Avibase ids, and other bird taxonomies (see README of <https://github.com/McTavishLab/AvesData>).

`cloodl_data$trees.summary.trees` Phylo objects of complete dated trees mapped to the Clements taxonomy year:

- year2024
- year2023
- year2022
- year2021

These are generated from `summary_dated_clements.nex` (see <https://github.com/McTavishLab/AvesData> README).

`annotations` Complete annotations of the OpenTree synthetic tree for this version, used to determine appropriate subtree citations.

`cloodl_data$study_info` A mapping of OpenTree study ids to full citations. Used with annotations to generate appropriate citations for trees and subtrees.

`cloodl_data$versions` A character vector of all possible tree versions. To access older versions, download the data repository using `get_avesdata_repo()`.

`cloodl_data$tax_years` A character vector of all available taxonomies. The current tree version is mapped to each of these taxonomies, along with crosswalks linking the Clements taxonomy for each year to other identifiers.

This data object is generated using the following code:

```
cloodl_data = list()

cloodl_data$versions <- c("1.2", "1.3", "1.4", "1.5")

fullTree2021 <- treeGet("1.5", "2021", data_path="~/projects/otapi/AvesData")
fullTree2022 <- treeGet("1.5", "2022", data_path="~/projects/otapi/AvesData")
fullTree2023 <- treeGet("1.5", "2023", data_path="~/projects/otapi/AvesData")
fullTree2024 <- treeGet("1.5", "2024", data_path="~/projects/otapi/AvesData")

tax2021 <- taxonomyGet(2021, data_path="~/projects/otapi/AvesData")
tax2022 <- taxonomyGet(2022, data_path="~/projects/otapi/AvesData")
tax2023 <- taxonomyGet(2023, data_path="~/projects/otapi/AvesData")
tax2024 <- taxonomyGet(2024, 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$taxonomy.files$Year2024 <- tax2024

cloodl_data$tax_years <- c("2021", "2022", "2023", "2024")

annot_filename <- "~/projects/otapi/AvesData/Tree_versions/Aves_1.5/OpenTreeSynth/annotated_supertree
all_nodes <- jsonlite::fromJSON(txt=annot_filename)
cloodl_data$trees$Aves_1.5$annotations <- all_nodes

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)
```

```
clootl_data$study_info <- study_info
save(clootl_data, file="~/projects/otapi/clootl/data/clootl_data.rda", compress="xz")
```

Source

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

extractTree	<i>Extract a complete or pre-pruned phylogeny from the clootl datastore</i>
-------------	---

Description

This function extracts one or more phylogenies in the desired taxonomy and tree version. It defaults to the pre-packaged summary trees, but can also be used to extract sets of phylogenies expressing uncertainty, once they have been downloaded from the online repository.

Usage

```
extractTree(
  species = "all_species",
  label_type = "scientific",
  taxonomy_year = 2024,
  version = "1.5",
  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. eBird taxonomy files can be accessed using <code>taxonomyGet()</code> . 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 are 2021-2024. Both numeric and character inputs are acceptable here. Any value aside from these years will result in an error. Default is most recent year.
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. If a summary, dated tree is desired, this is sufficient and does not need to be modified. However, if a user wishes to extract a set of complete dated trees, for example to iterate an analysis across a cloud of trees, or to use an older version of the tree than the current one packed in the data object, this function can also accept a path to the downloaded set of trees. If you have already

downloaded the AvesData repo available at <https://github.com/McTavishLab/AvesData> use `data_path=` the path to the download location. Alternately, you can download the full data repo using `get_avesdata_repo()`. This approach will download the data and set an environmental variable `AVESDATA_PATH`. When `AVESDATA_PATH` is set, the `data_path` will default to this value. To manually set `AVESDATA_PATH` to the location of your downloaded AvesData repo use `set_avesdata_repo_path()`

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. 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. Going forward, we will begin sunsetting older taxonomies, and intend to maintain the current year plus the two previous years.

Value

One or more phylogenies of the specified taxa in the specified eBird taxonomy version and clootree 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="1.5")
```

getCitations

Identify contributing studies

Description

Quantify the contribution of studies informing an extracted tree, and obtain DOI and citation information for those studies.

Usage

```
getCitations(tree, version = "1.5", data_path = FALSE)
```

Arguments

tree	A phylogeny obtained from extractTree (see details).
version	The version of the tree used in extract tree. Default to the most recent version of the tree. and can be passed as a character string or as numeric. If an alternate version was used to create the tree this function may fail or give incomplete or incorrect citation information.
data_path	Default to FALSE. If you are gathering citations for an older version of the tree than the current one packed in the data object, you will have already downloaded the data repo in order to generate that tree. The data is available at https://github.com/McTavishLab/AvesData . If you have manually downloaded the repo, use data_path= the path to the download location. Alternately, you can download the full data repo using get_avesdata_repo() . This approach will download the data and set an environmental variable AVESDATA_PATH. When AVESDATA_PATH is set, the data_path argument will default to this value. To manually set AVESDATA_PATH to the location of your downloaded AvesData repo use set_avesdata_repo_path()

Details

The function will determine what proportion of nodes in your phylogeny are supported by each study that goes into creating the final clootl tree. We use 'supported by' in the sense described in Redelings and Holder, PeerJ (2017) <https://peerj.com/articles/3058/>, and as shown in the tree.opentreeoflife.org tree viewer. We normalize these values to a percentage of internal nodes in the target tree supported by each study. In any resulting publication, please cite both the synthetic tree (McTavish et al. 2025), clootl (Miller et al. 2025) and "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. This function relies on the phylogenetic synthesis information directly, and is agnostic to taxonomy version.

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 also included in the dataframe.

Author(s)

Eliot Miller, Emily Jane McTavish

Examples

```
#pull the taxonomy file out
data(clootl_data)
tax <- clootl_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="1.5")

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

```

get_avesdata_repo *Download the AvesData full repository*

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. If path exists, will not re-download the data unless overwrite=TRUE.

Details

Will download full data repo from <https://github.com/McTavishLab/AvesData>. This data is required to use [sampleTrees\(\)](#) to sample from the distribution of dated trees, or to access earlier versions of the complete tree. This function will download the data and set an environmental variable AVESDATA_PATH to the location of the data download. When AVESDATA_PATH is set, the data_path in any clootl functions with a data_path argument will default to this value. To manually set AVESDATA_PATH to the location of your downloaded AvesData repo use [set_avesdata_repo_path\(\)](#)

Value

No return value. This function is used 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 = 2024,
  version = "1.5",
  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 '0.1', '1.0', '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. If a summary, dated tree is desired, this is sufficient and does not need to be modified. However, if a user wishes to extract a set of complete dated trees, for example to iterate an analysis across a cloud of trees, or to use an older version of the tree than the current one packed in the data object, this function can also accept a path to the downloaded set of trees. If you have already downloaded the AvesData repo available at https://github.com/McTavishLab/AvesData use data_path= the path to the download location. Alternately, you can download the full data repo using get_avesdata_repo() . This approach will download the data and set an environmental variable AVESDATA_PATH. When AVESDATA_PATH is set, the data_path will default to this value. To manually set AVESDATA_PATH to the location of your downloaded AvesData repo use set_avesdata_repo_path()

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

Description

Set path to Aves Data folder already somewhere on your computer

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.

Details

Based on <https://github.com/CornellLabofOrnithology/auk/blob/main/R/auk-set-ebd-path.r> Use this function to manually set or update location of a downloaded AvesData folder from <https://github.com/McTavishLab/AvesData>. When AVESDATA_PATH is set, the data_path in any clootl functions with a data_path argument will default to this value.

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

Load a bird taxonomy into the R environment

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 are 2021-2024. Both numeric and character inputs are acceptable here. Any value aside from these years will result in an error. Default is most recent year.
data_path	Default to FALSE. If a summary, dated tree is desired, this is sufficient and does not need to be modified. However, if a user wishes to extract a set of complete dated trees, for example to iterate an analysis across a cloud of trees, or to use an older version of the tree than the current one packed in the data object, this function can also accept a path to the downloaded set of trees. If you have already downloaded the AvesData repo available at https://github.com/McTavishLab/AvesData use data_path= the path to the download location. Alternately, you can download the full data repo using get_avesdata_repo() . This approach will download the data and set an environmental variable AVESDATA_PATH. When AVESDATA_PATH is set, the data_path will default to this value. To manually set AVESDATA_PATH to the location of your downloaded AvesData repo use set_avesdata_repo_path()

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.

Index

* datasets

cloutl_data, 2

cloutl_data, 2

extractTree, 4

get_avesdata_repo, 7

get_avesdata_repo(), 5, 6, 8, 10

getCitations, 5

sampleTrees, 8

sampleTrees(), 7

set_avesdata_repo_path, 9

set_avesdata_repo_path(), 5–8, 10

taxonomyGet, 10

taxonomyGet(), 4