

Package ‘aptg’

July 19, 2026

Type Package

Title Automatic Phylogenetic Tree Generator

Version 0.4.0

Description Generates phylogenetic trees and distance matrices from a list of taxon names, or from a higher taxon expanded down to a chosen lower rank. Trees are obtained as induced subtrees of the Open Tree of Life synthetic tree using the 'rotl' package (Michonneau, Brown and Winter, 2016, <[doi:10.1111/2041-210X.12593](https://doi.org/10.1111/2041-210X.12593)>). Expansion of a higher taxon to its descendants uses 'taxize' (Chamberlain and Szocs, 2013, <[doi:10.12688/f1000research.2-191.v2](https://doi.org/10.12688/f1000research.2-191.v2)>).

Imports ape, rotl, taxize, stats

Suggests fishtree, rgbif, geosphere, testthat (>= 3.0.0), knitr, rmarkdown

VignetteBuilder knitr

License MIT + file LICENSE

Encoding UTF-8

NeedsCompilation no

Config/testthat/edition 3

URL <https://github.com/christophebenjamin/aptg>

BugReports <https://github.com/christophebenjamin/aptg/issues>

Config/roxygen2/version 8.0.0

Author Christophe Benjamin [aut, cre]

Maintainer Christophe Benjamin <christophe.benjamin@protonmail.com>

Repository CRAN

Date/Publication 2026-07-19 07:20:08 UTC

Contents

downto.tree	2
region.tree	3
taxa.tree	5

Index

7

downto.tree	<i>Phylogenetic tree from a taxon down to a lower rank</i>
-------------	--

Description

Expands a higher taxon to all of its descendants at a chosen rank using [downstream](#), then builds the corresponding Open Tree of Life subtree via [taxa.tree](#).

Usage

```
downto.tree(
  taxon,
  downto,
  db = "ncbi",
  key = NULL,
  source = c("otl", "fish"),
  plot = TRUE,
  verbose = TRUE
)
```

Arguments

taxon	A single higher taxon name (e.g. a family or genus).
downto	Target lower rank to expand down to, e.g. "species" or "genus". Must be lower than taxon's rank.
db	Taxonomic database used for the downstream expansion. Default "ncbi". "gbif" and "itis" are also supported and need no API key; prefer them if you hit NCBI rate limits.
key	NCBI Entrez API key (relevant when db = "ncbi"). When supplied it is exported as the ENTREZ_KEY environment variable for the duration of the call, so both the get_uid and downstream requests use it; the previous value is restored on exit. Create a key with use_entrez . Alternatively, set ENTREZ_KEY yourself (e.g. in .Renvirom) and leave this NULL. Without a key NCBI throttles requests to 3 per second (10 with one).
source	Tree source passed to taxa.tree : "otl" (default) or "fish".
plot	Logical; plot the resulting tree(s). Default TRUE.
verbose	Logical; print progress messages. Default TRUE.

Details

Only the taxon-expansion step uses **taxize**; the tree itself comes from Open Tree of Life, so descendants that are absent from the Open Tree synthesis are dropped and reported in unmatched.

Value

Invisibly, the list returned by `taxa.tree` (trees, unmatched). Because descendants of a taxon may span more than one phylum, trees can contain more than one element.

See Also

`taxa.tree`

Examples

```
## Not run:
# All species in the deer family:
downto.tree("Cervidae", downto = "species")

## End(Not run)
```

region.tree

Phylogenetic tree(s) for a taxon within a geographic area

Description

Retrieves the species of a given clade recorded in an area from the Global Biodiversity Information Facility (GBIF, via **rgbif**), then builds the tree(s) with `taxa.tree` (inheriting the per-phylum split and the source option). The area is specified in one of two ways:

Usage

```
region.tree(
  taxon,
  lat = NULL,
  lon = NULL,
  radius_km = NULL,
  province = NULL,
  gadm = NULL,
  source = c("otl", "fish"),
  max_species = 2000L,
  plot = TRUE,
  verbose = TRUE
)
```

Arguments

taxon	Required. A single clade name to constrain the search (e.g. "Aves", "Mammalia", "Salmonidae"). Without it, a populated area returns an unusable number of species.
lat, lon, radius_km	Radius mode: latitude, longitude (decimal degrees) and radius in kilometres. Supply all three.

province	Admin mode: a Canadian province/territory name or postal code (e.g. "Quebec", "QC", "Nunavut", "BC").
gadm	Admin mode: a GADM GID (e.g. "CAN.11_1" for Quebec, or any GID world-wide). Takes precedence over province.
source	Passed to taxa.tree : "otl" (default) or "fish".
max_species	Cap on the number of distinct species retrieved from GBIF (facet limit). Default 2000.
plot, verbose	Passed to taxa.tree .

Details

- **Radius mode:** a circle of radius_km around (lat, lon).
- **Admin mode:** an administrative area given by gadm (a GADM GID, works worldwide) or province (a Canadian province or territory name/abbreviation; all 13 are supported).

Value

Invisibly, the [taxa.tree](#) result (`list(trees, unmatched)`) with an added species element: the distinct species names GBIF returned for the area.

Requirements

Needs the suggested **rgbif** package (both modes) and **geosphere** (radius mode only). Both hit the network.

Caveats

GBIF returns *occurrence records*, not a curated checklist. The result is presence-only and sampling-biased (some groups and places are far better recorded than others), so it reflects what has been observed and digitised, not a definitive species inventory. Only records with coordinates and without flagged geospatial issues are counted, but coordinate precision still varies. Treat the species list as a starting point, not ground truth.

Canadian areas

The province lookup covers Alberta, British Columbia, Manitoba, New Brunswick, Newfoundland and Labrador, Northwest Territories, Nova Scotia, Nunavut, Ontario, Prince Edward Island, Quebec, Saskatchewan and Yukon (and their postal codes). The GIDs follow GADM's alphabetical level-1 ordering; the resolved GID is printed when verbose = TRUE. If GBIF's GADM version ever disagrees, pass the GID directly via gadm (browse them at <https://api.gbif.org/v1/geocode/gadm/browse/CAN>).

See Also

[taxa.tree](#), [downto.tree](#)

Examples

```
## Not run:
# Birds within 50 km of Montreal:
region.tree("Aves", lat = 45.50, lon = -73.57, radius_km = 50)

# Mammals of Nunavut (admin mode, Canadian territory):
region.tree("Mammalia", province = "Nunavut")

# Freshwater fishes of Quebec, as a dated tree:
region.tree("Actinopterygii", province = "QC", source = "fish")

# Anywhere in the world via a raw GADM GID:
region.tree("Reptilia", gadm = "USA.5_1")

## End(Not run)
```

taxa.tree

Phylogenetic trees and distance matrices from a list of taxa

Description

Builds phylogenetic trees for a set of taxa. With the default Open Tree of Life backend, taxa are grouped by phylum and one induced subtree is built per phylum, so no tree ever spans above phylum (mammals, angiosperms, arthropods, etc. come back as separate trees). With source = "fish" a single dated tree is drawn from the Fish Tree of Life for the ray-finned fishes in the input.

Usage

```
taxa.tree(species, source = c("otl", "fish"), plot = TRUE, verbose = TRUE)
```

Arguments

species	Character vector of taxon names (typically species).
source	Tree source. "otl" (default) uses the Open Tree of Life synthetic tree via rotl (topology only, all of life). "fish" uses the Fish Tree of Life via fishtree (a dated chronogram; requires the suggested fishtree package and only covers ray-finned fishes, Actinopterygii).
plot	Logical; if TRUE (default) each tree is plotted, one panel per phylum.
verbose	Logical; print progress messages. Default TRUE.

Value

Invisibly, a list with:

trees a named list, one element per phylum (or "Actinopterygii" for source = "fish"). Each element is itself a list with tree (a [phylo](#)) and dist (a distance matrix).

unmatched names that could not be placed (unresolved, absent from the tree, or lacking a phylum in the Open Tree lineage).

Phylum safeguard

Different phyla have no shared, meaningfully calibrated deep phylogeny, and dated backends are clade-specific, so combining them into one tree is not sensible. To enforce this, the Open Tree backend looks up each taxon's phylum (via [taxonomy_taxon_info](#)) and builds a separate tree per phylum. A phylum with fewer than two placed taxa is skipped and reported.

Branch lengths

Open Tree induced subtrees are topology-only; unit edges are assigned when branch lengths are absent, so `dist` counts edges between tips. The "fish" backend returns a dated chronogram, so its `dist` is a patristic (time) distance.

See Also

[downto.tree](#)

Examples

```
## Not run:
# Mixed mammals + plants: returns two separate trees, one per phylum.
taxa.tree(c("Canis lupus", "Canis latrans", "Acer saccharum",
            "Acer rubrum", "Alces alces", "Betula alleghaniensis"))

# Dated fish tree (needs the fishtree package):
taxa.tree(c("Thunnus thynnus", "Gadus morhua", "Danio rerio",
            "Salmo salar"), source = "fish")

## End(Not run)
```

Index

downstream, [2](#)

downto.tree, [2](#), [4](#), [6](#)

phylo, [5](#)

region.tree, [3](#)

taxa.tree, [2-4](#), [5](#)

taxonomy_taxon_info, [6](#)

use_entrez, [2](#)