Package: ftolr (via r-universe)

September 10, 2024

Title Data for the Fern Tree of Life (FTOL)

Version 1.6.0

Description Provides the Fern Tree of Life (FTOL) phylogeny and
related data files.

License MIT + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.1

Imports ape, assertthat

Suggests stringr, dplyr, readr, tibble, testthat (>= 3.0.0),
tidyverse, patrick

Depends R (>= 2.10)

LazyData true

LazyDataCompression bzip2

Config/testthat/edition 3

Repository https://joelnitta.r-universe.dev

RemoteUrl https://github.com/fernphy/ftolr

RemoteRef HEAD

RemoteSha 0f48643b176ab6a759874e35ce623942fbba36a6

Contents

---

accessions_long   *GenBank accessions used in the Fern Tree of Life (FTOL), long format*

Description

One row per accession.

Usage

```
accessions_long
```

Format

A tibble (data frame) with 14965 rows and 8 columns.

species Species name; matches names of tips in tree

locus Name of locus (gene or intergenic spacer region)

accession GenBank accession number

seq_len Sequence length (bp), excluding any missing or ambiguous bases

sci_name Scientific name used in FTOL

ncbi_name Scientific name used in the NCBI taxonomic database

ncbi_taxid NCBI taxonomy database unique identifier

outgroup Logical; TRUE for outgroup taxa, FALSE for ingroup taxa (ferns)

Details

For details on methods used to infer the tree, see Nitta et al. 2022.

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously up-
dated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

| accessions_wide | *GenBank accessions used in the Fern Tree of Life (FTOL), wide format* |

Description

One row per species.

Usage

```
accessions_wide
```

Format

A tibble (data frame) with 5615 rows and 13 columns.

species Species name; matches names of tips in tree

atpA GenBank accession number for atpA

atpB GenBank accession number for atpB

matK GenBank accession number for matK

rbcL GenBank accession number for rbcL

rps4 GenBank accession number for rps4

rps4-trnS GenBank accession number for rps4-trnS

trnL-trnF GenBank accession number for trnL-trnF

plastome GenBank accession number for plastomes

join_by Method used to join loci

specimen_voucher Specimen voucher

publication Publication

outgroup Logical; TRUE for outgroup taxa, FALSE for ingroup taxa (ferns)

Details

join_by only provided for Sanger (non-plastome) sequences. specimen_voucher and publication only provided for Sanger (non-plastome) sequences, and only if these were the same across all loci. specimen_voucher has been converted to lower case, alpha-numeric characters only to account for orthographic variation in GenBank data.

For details on methods used to infer the tree, see Nitta et al. 2022.

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously updated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

---

backbone_tree   *Fern Tree of Life (FTOL) backbone phylogeny*

Description

Majority-rule extended consensus phylogenetic tree inferred from whole fern plastomes by maximum-likelihood, including maximal sampling of available species on GenBank.

Usage

```
backbone_tree
```

Format

List of class "phylo"

Details

Includes 504 tips and 503 internal nodes. Branch lengths are in units of estimated genetic change per site (not ultrametric). Node labels are bootstrap support (%).

For details on methods used to infer the tree, see Nitta et al. 2022.

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously updated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

---

con_fossils   *Fossil calibration points used for dating Fern Tree of Life (FTOL)*

Description

con_fossils was used for the consensus tree; ml_fossils was used for the ML tree.

Usage

```
con_fossils
```

```
ml_fossils
```

Format

A tibble (data frame) with 54 rows and 12 columns.

n_fos Unique ID number for fossil

minimum_age Minimum age to apply to fossil constraint

node_calibrated Node calibrated by fossil constraint. Combination of 'affinities' and 'affinities_group'

fossil_taxon Taxonomic name of fossil (without author)

affinities_group Type of group the fossil belongs to (crown or stem)

affinities Narrowest clade the fossil belongs to; the clade whose date is constrained by the fossil

monophyly Are the affinities monophyletic? 'Yes', 'No', or 'Monotypic'

number_tips Number of tips in the clade constrained by the fossil

mrca Node number of MRCA for the clade constrained by the fossil

stem_mrca Node number of the parent node of the MRCA for the clade constrained by the fossil

tip_1 Name of one taxon that defines the clade constrained by the fossil

tip_2 Name of another taxon that defines the clade constrained by the fossil

An object of class `spec_tbl_df` (inherits from `tbl_df`, `tbl`, `data.frame`) with 54 rows and 12 columns.

Details

The fossils in `con_fossils` and `ml_fossils` are generally the same, but the node each calibrates may differ between the trees because of differences in topology. Also, in some cases, a given fossil may only be applied to the ML or consensus tree (not both) because the topology of the tree makes that fossil redundant with another fossil.

The node corresponding to the fossil constraint is defined as the most recent common ancestor (MRCA, column 'mrca') of two tips columns ('tip_1' and 'tip_2') for crown affinities, or its parent node (column 'stem_mrca') for stem affinities. The two tips are identified automatically for monophyletic clades, or by hand for non-monophyletic clades. 'mrca' is not defined for monotypic groups (only 'stem_mrca').

Does not include the constraint on the root of the tree.

For further details on methods used for molecular dating, see Nitta et al. 2022.

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously updated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

---

ftol_con_dated_tree   *Fern Tree of Life (FTOL) dated consensus phylogeny*

Description

Majority-rule extended consensus phylogenetic tree of ferns inferred from seven loci by maximum-likelihood and dated using 60 fossil ferns as constraints, including maximal sampling of available species on GenBank.

Usage

```
ftol_con_dated_tree
```

Format

List of class "phylo"

Details

A backbone phylogeny based on full plastome sequences including 503 tips was first inferred and used as a constraint tree for the consensus tree.

Includes 5,610 tips and 5,609 internal nodes. Branch lengths are in units of time (Ma). Node labels are bootstrap support (BS) values (%).

For details on methods used to infer the tree, see Nitta et al. 2022.

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously updated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

---

ftol_con_tree   *Fern Tree of Life (FTOL) consensus phylogeny*

Description

Majority-rule extended consensus phylogenetic tree of ferns inferred from seven loci by maximum-likelihood, including maximal sampling of available species on GenBank.

Usage

```
ftol_con_tree
```

Format

List of class "phylo"

Details

A [backbone phylogeny](#) based on full plastome sequences including 503 tips was first inferred and used as a constraint tree for the consensus tree.

Includes 5,611 tips and 5,610 internal nodes. Branch lengths are in units of estimated genetic change per site (not ultrametric). Node labels are bootstrap support (BS) values (%).

For details on methods used to infer the tree, see Nitta et al. 2022.

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously updated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

---

ftol_ferns	*Fern Tree of Life (FTOL) fern species*

Description

Fern species names included in FTOL in alphabetical order. All species names are unique. Includes 5592 species.

Usage

```
ftol_ferns
```

Format

Character vector

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously updated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

---

ftol_match_results	*Fern Tree of Life (FTOL) taxonomic name matching results*

Description

Results of taxonomic name matching and resolution for species included in FTOL. Original names from the NCBI taxonomic database (https://www.ncbi.nlm.nih.gov/taxonomy) were matched to those in pteridocat (https://github.com/fernphy/pteridocat), and synonyms resolved to their accepted names in pteridocat. For compact printing, it is recommended to use the `tibble` library (`library(tibble)`).

Usage

```
ftol_match_results
```

Format

A tibble (data frame) with 6476 rows and 7 columns.

query Queried taxonomic name from NCBI

resolved_name Resolved name used in FTOL

matched_name Name matching query in pteridocat

resolved_status Taxonomic status of resolved name

matched_status Taxonomic status of matched name

match_type Type of match assigned by taxontools

taxid NCBI taxonomic ID

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously updated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

Pteridophyte Phylogeny Group I (2016) A community-derived classification for extant lycophytes and ferns. Journal of Systematics and Evolution 54:563–603. https://doi.org/10.1111/jse.12229

---

ftol_ml_dated_tree   *Fern Tree of Life (FTOL) dated maximum-likelihood phylogeny*

Description

A maximum-likelihood (ML) phylogenetic tree of ferns inferred from seven loci and dated using fossil ferns as constraints, including maximal sampling of available species on GenBank.

Usage

```
ftol_ml_dated_tree
```

Format

List of class "phylo"

Details

A backbone phylogeny based on full plastome sequences including 503 tips was first inferred and used as a constraint tree for the ML tree.

Includes 5,610 tips and 5,609 internal nodes. Branch lengths are in units of time (Ma). Node labels are bootstrap support (BS) values (%) separated by a forward slash. BS value before the slash is BS calculated from the full (Sanger) sampling; BS value after the slash is BS calculated from the backbone (plastome) sampling.

For details on methods used to infer the tree, see Nitta et al. 2022.

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously up-dated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

---

ftol_ml_tree	*Fern Tree of Life (FTOL) maximum-likelihood phylogeny*

Description

A maximum-likelihood (ML) phylogenetic tree of ferns inferred from seven loci, including maximal sampling of available species on GenBank.

Usage

```
ftol_ml_tree
```

Format

List of class "phylo"

Details

A backbone phylogeny based on full plastome sequences including 503 tips was first inferred and used as a constraint tree for the ML tree.

Includes 5,611 tips and 5,610 internal nodes. Branch lengths are in units of estimated genetic change per site (not ultrametric). Node labels are bootstrap support (BS) values (%) separated by a forward slash. BS value before the slash is BS calculated from the full (Sanger) sampling; BS value after the slash is BS calculated from the backbone (plastome) sampling.

For details on methods used to infer the tree, see Nitta et al. 2022.

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously up-dated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

---

ftol_taxonomy   *Fern Tree of Life (FTOL) taxonomic data*

Description

Taxonomic data for species included in FTOL. For compact printing, it is recommended to use the `tibble` library (`library(tibble)`).

Usage

```
ftol_taxonomy
```

Format

A tibble (data frame) with 5611 rows and 8 columns.

species Species name

genus Genus name

subfamily Subfamily name

family Family name

suborder Suborder name

order Order name

major_clade Informal higher level clade name, either order or suborder

outgroup Logical; TRUE for outgroup taxa, FALSE for ingroup taxa (ferns)

Details

Taxonomy follows Pteridophyte Phylogeny Group I (2016).

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously updated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

Pteridophyte Phylogeny Group I (2016) A community-derived classification for extant lycophytes and ferns. Journal of Systematics and Evolution 54:563–603. https://doi.org/10.1111/jse.12229

---

ft_data_ver | *Current Fern Tree of Life (FTOL) version and related metadata*

Description

Current Fern Tree of Life (FTOL) version and related metadata

Usage

```
ft_data_ver(what = "ftol")
```

Arguments

what Character vector of length 1; what type of data to return. Must choose from "ftol" (FTOL data version; default), "gb" (version of GenBank data used to build FTOL), or "cutoff" (cutoff date of GenBank data).

Details

FTOL is built from DNA sequence data obtained from the NCBI GenBank FTP server (https://ftp.ncbi.nlm.nih.gov/genbank/), which releases updated datasets periodically. For more information about the current GenBank data release, see the release notes.

Value

Character vector of length 1

Examples

```
ft_data_ver()
```

---

ft_seqs | *Load the DNA sequences used to build the Fern Tree of Life (FTOL)*

Description

For details about the data, see plastome_alignment for plastome sequences and sanger_alignment for Sanger sequences.

Usage

```
ft_seqs(
  loci = NULL,
  plastome = FALSE,
  aligned = TRUE,
  del_gaps = ifelse(is.null(loci), FALSE, TRUE),
  drop_og = FALSE
)
```

Arguments

loci	Character vector (optional); names of loci to include in the output. If NULL (default), all loci will be included. For a list of valid locus names, see [plastome_parts](#) for plastome loci and [sanger_parts](#) for Sanger loci.
plastome	Logical vector of length 1; If TRUE, the alignment will include loci and species with complete plastomes available; otherwise will include Sanger loci and all species. Default FALSE.
aligned	Logical vector of length 1; If TRUE, the output will be aligned sequences (matrix); otherwise will be unaligned (list). Default TRUE.
del_gaps	Logical vector of length 1; If TRUE, all gaps will be deleted from the alignment. Default is TRUE if loci is NULL, FALSE otherwise.
drop_og	Logical vector of length 1; If TRUE, the outgroup (non-ferns) will be excluded; otherwise the outgroup is included. Default FALSE.

Details

After subsetting loci, columns / rows consisting of only gaps will be deleted regardless of del_gaps argument (del_gaps deletes *all* gaps, typically resulting in unaligned sequences).

For details on methods used to assemble alignments, see Nitta et al. 2022.

Value

List or matrix of class "DNAbin"; DNA sequences.

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously updated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

Examples

```
# Default is aligned Sanger sequences
ft_seqs()
```

---

ft_tree   *Load the Fern Tree of Life (FTOL)*

Description

For details on methods used to infer the tree, see Nitta et al. 2022.

Usage

```
ft_tree(
  branch_len = "ultra",
  consensus = TRUE,
  rooted = TRUE,
  backbone = FALSE,
  drop_og = FALSE,
  label_ages = FALSE,
  decimals = NULL
)
```

Arguments

branch_len	Character vector of length 1; how to treat branch lengths. Must choose from "ultra" (ultrametric tree, branchlengths in units of time), "raw" (raw branch-lengths in units of genetic change), or "clado" (cladogram, no branchlengths). Default "ultra".
consensus	Logical vector of length 1; if TRUE, the majority-rule extended consensus phylogeny will be returned; otherwise returns the maximum-likelihood tree. Default TRUE.
rooted	Logical vector of length 1; if TRUE, the phylogeny will be rooted on bryophytes; otherwise the phylogeny is unrooted. Default TRUE.
backbone	Logical vector of length 1; if TRUE, the backbone phylogeny (only species with complete plastomes available) will be returned; otherwise the phylogeny will include all species. Default FALSE.
drop_og	Logical vector of length 1; if TRUE, the outgroup (non-ferns) will be excluded; otherwise the outgroup is included. Default FALSE.
label_ages	Logical vector of length 1; if TRUE, internal nodes will be labeled with ages. Only works if branch_len is "ultra". Default FALSE.
decimals	Numeric vector of length 1; number of decimals for rounding node labels if label_ages is TRUE; null (default) does no rounding.

Details

Not all combinations are possible. For example, branch_len = "ultra" is only available if backbone = FALSE.

Value

List of class "phylo"; a phylogenetic tree.

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously updated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

Examples

```
# Default is the consensus tree with branchlengths in units of time
ft_tree()
```

---

plastome_alignment	*Aligned plastome DNA sequences used to build Fern Tree of Life (FTOL)*

Description

Concatenation of 79 loci from whole fern plastomes, including maximal sampling of available species on GenBank.

Usage

```
plastome_alignment
```

Format

Matrix of class "DNAbin"

Details

Includes 504 species and 75,413 bases.

For details on methods used to assemble alignment, see Nitta et al. 2022.

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously updated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

---

plastome_parts	*Start and end positions of loci in plastome DNA sequence alignment*

Description

These data can be used to extract portions of the plastome DNA sequence alignment (plastome_alignment) corresponding to individual loci.

Usage

```
plastome_parts
```

Format

A tibble (data frame) with 79 rows and 3 columns.

locus Name of locus (gene or intergenic spacer region)

start Start position (column number) of locus in concatenated DNA alignment

end End position (column number) of locus in concatenated DNA alignment

Details

For details on methods used to assemble alignment, see Nitta et al. 2022.

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously updated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

---

sanger_alignment	*Aligned (mostly) Sanger DNA sequences used to build Fern Tree of Life (FTOL)*

Description

Concatenation of 7 loci primarily obtained from Sanger sequencing, including maximal sampling of available species on GenBank.

Usage

```
sanger_alignment
```

Format

Matrix of class "DNAbin"

Details

Includes 5,611 species and 12,648 bases.

For details on methods used to assemble alignment, see Nitta et al. 2022.

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously updated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

---

sanger_parts   *Start and end positions of loci in Sanger DNA sequence alignment*

Description

These data can be used to extract portions of the Sanger DNA sequence alignment (sanger_alignment) corresponding to individual loci.

Usage

```
sanger_parts
```

Format

A tibble (data frame) with 7 rows and 3 columns.

locus Name of locus (gene or intergenic spacer region)

start Start position (column number) of locus in concatenated DNA alignment

end End position (column number) of locus in concatenated DNA alignment

Details

For details on methods used to assemble alignment, see Nitta et al. 2022.

References

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. (2022) An open and continuously updated fern tree of life. https://doi.org/10.3389/fpls.2022.909768

Index