

Package ‘rtreeoflife’

August 4, 2026

Title Access Tree of Life Data Releases

Version 0.1.0

Description Search species and specimen records from the Royal Botanic Gardens Tree of Life Explorer and retrieve selected genomic sequence files. Provides a bundled species index, tools for matching scientific names, and selective download helpers for sequence files. Supports tidy data workflows with list-columns, summaries, exploratory visualisation with 'ggplot2', and illustrative tree construction from shared genes.

License MIT + file LICENSE

URL <https://github.com/PaulESantos/rtreeoflife>,
<https://treeoflife.kew.org/>,
https://sftp.kew.org/pub/treeoflife/current_release/

BugReports <https://github.com/PaulESantos/rtreeoflife/issues>

Depends R (>= 4.1.0)

Imports cli, dplyr, ggplot2, tibble, tools, utils

Suggests ape, testthat (>= 3.0.0)

Config/testthat/edition 3

Encoding UTF-8

LazyData true

LazyDataCompression xz

RoxygenNote 8.0.0

NeedsCompilation no

Author Paul E. Santos Andrade [aut, cre, cph] (ORCID:
<<https://orcid.org/0000-0002-6635-0375>>)

Maintainer Paul E. Santos Andrade <paulefrens@gmail.com>

Repository CRAN

Date/Publication 2026-08-04 09:40:02 UTC

Contents

tol_attach_fasta	2
tol_base_url	3
tol_build_gene_tree	3
tol_common_genes	4
tol_data_dir	4
tol_download	5
tol_download_bundle	6
tol_download_directory	7
tol_download_fasta	8
tol_download_release	9
tol_export_fasta	10
tol_fasta_long	10
tol_fasta_summary	11
tol_index	11
tol_known_bundle	12
tol_known_bundles	12
tol_manifest	12
tol_match_species	13
tol_plot_fasta_summary	14
tol_plot_gene_recovery	14
tol_plot_tree	15
tol_read_fasta	15
tol_release_url	16
tol_resolve_fasta	16
tol_save_fasta	17
tol_search_species	17
tol_set_base_url	18
tol_species	18
tol_species_index	19
Index	20

tol_attach_fasta	<i>Attach parsed FASTA data to a download plan</i>
------------------	--

Description

Attach parsed FASTA data to a download plan

Usage

```
tol_attach_fasta(  
  plan,  
  column = "fasta",  
  as = c("data.frame", "list", "text"),  
  missing = c("empty", "error")  
)
```

Arguments

plan	A data frame returned by <code>tol_download_fasta()</code> , <code>tol_resolve_fasta()</code> , or <code>tol_save_fasta()</code> .
column	Name of the list-column to create.
as	FASTA representation passed to <code>tol_read_fasta()</code> .
missing	What to store for missing files. "empty" stores an empty data frame/list/text value; "error" stops on the first missing file.

Value

A data frame with a FASTA list-column.

tol_base_url	<i>Kew Tree of Life base URL</i>
--------------	----------------------------------

Description

Kew Tree of Life base URL

Usage

```
tol_base_url()
```

Value

A character scalar with the default public Kew data URL.

tol_build_gene_tree	<i>Build an illustrative tree from one shared gene</i>
---------------------	--

Description

This helper is intended for exploration. It extracts one homologous gene from each selected FASTA, trims sequences to their common minimum length, computes raw DNA distances with ape, and returns an hclust tree plus the distance matrix. Rigorous phylogenetics should use explicit alignment and model selection outside this helper.

Usage

```
tol_build_gene_tree(plan, gene_id = NULL, min_records = 3)
```

Arguments

plan	A data frame returned by <code>tol_attach_fasta()</code> or a FASTA plan with local files.
gene_id	Optional gene id. If NULL, the most shared gene is selected.
min_records	Minimum records required to build the tree.

Value

A list with tree, distances, gene_id, sequences, and note.

tol_common_genes	<i>Choose a gene shared by multiple downloaded FASTA files</i>
------------------	--

Description

Choose a gene shared by multiple downloaded FASTA files

Usage

```
tol_common_genes(plan, min_records = 3)
```

Arguments

plan	A data frame returned by <code>tol_attach_fasta()</code> or a FASTA plan with local files.
min_records	Minimum number of records that must contain the gene.

Value

A tibble ranked by number of records and median width.

tol_data_dir	<i>Local data directory for downloaded files</i>
--------------	--

Description

The default can be overridden with option `rteoflife.data_dir` or environment variable `RTREEOFLIFE_DATA_DIR`.

Usage

```
tol_data_dir(create = TRUE)
```

Arguments

create	Create the directory if it does not exist.
--------	--

Value

A normalized path.

tol_download	<i>Download files from a Kew Tree of Life release</i>
--------------	---

Description

Download files from a Kew Tree of Life release

Usage

```
tol_download(  
  paths,  
  dest_dir = tol_data_dir(),  
  release = "current_release",  
  overwrite = FALSE,  
  preserve_dirs = TRUE,  
  quiet = FALSE,  
  base_url = tol_base_url(),  
  timeout = 600,  
  retries = 3,  
  retry_wait = 5  
)
```

Arguments

paths	Character vector of file paths inside the release directory.
dest_dir	Local destination directory.
release	Release directory. Defaults to Kew's current release link.
overwrite	Replace existing local files.
preserve_dirs	Preserve the remote directory structure below dest_dir.
quiet	Passed to <code>utils::download.file()</code> .
base_url	Base public URL.
timeout	Download timeout in seconds.
retries	Number of attempts for each file.
retry_wait	Seconds to wait between attempts.

Value

A character vector with local file paths.

tol_download_bundle	<i>Download a predefined Kew Tree of Life bundle</i>
---------------------	--

Description

Download a predefined Kew Tree of Life bundle

Usage

```
tol_download_bundle(
  bundle = "manifests",
  dest_dir = tol_data_dir(),
  release = "current_release",
  overwrite = FALSE,
  quiet = FALSE,
  base_url = tol_base_url(),
  timeout = 600,
  retries = 3,
  retry_wait = 5
)
```

Arguments

bundle	Bundle name. See tol_known_bundles() .
dest_dir	Local destination directory.
release	Release directory. Defaults to Kew's current release link.
overwrite	Replace existing local files.
quiet	Passed to utils::download.file() .
base_url	Base public URL.
timeout	Download timeout in seconds.
retries	Number of attempts for each file.
retry_wait	Seconds to wait between attempts.

Value

A character vector with local file paths.

tol_download_directory

Download and validate one directory from a Kew Tree of Life release

Description

Convenience wrapper around [tol_download_release\(\)](#) for subdirectories such as "fasta/by_gene", "fasta/alignments", or "tree/species".

Usage

```
tol_download_directory(
  path,
  dest_dir = file.path("raw-data", normalize_release(release)),
  release = "current_release",
  manifest_path = file.path("raw-data", paste0("kew_download_", gsub("[^A-Za-z0-9]+",
    "_", normalize_remote_path(path)), ".csv")),
  overwrite = FALSE,
  validate_only = FALSE,
  quiet = FALSE,
  base_url = tol_base_url(),
  timeout = 600,
  retries = 5,
  retry_wait = 10
)
```

Arguments

path	Directory path inside the release.
dest_dir	Local destination directory for the release contents.
release	Release directory. Defaults to Kew's current release link.
manifest_path	Local CSV manifest path.
overwrite	Replace existing local files.
validate_only	Only index and validate existing local files.
quiet	Passed to utils::download.file() .
base_url	Base public URL.
timeout	Download timeout in seconds.
retries	Number of attempts for each file.
retry_wait	Seconds to wait between attempts.

Value

A data frame with release id, remote paths, local paths, sizes, and validation status.

tol_download_fasta	<i>Download FASTA files for selected species records</i>
--------------------	--

Description

Download FASTA files for selected species records

Usage

```
tol_download_fasta(
  records = NULL,
  dest_dir = NULL,
  manifest_path = if (is.null(dest_dir)) NULL else file.path("raw-data",
    "kew_fasta_download_manifest.csv"),
  overwrite = FALSE,
  validate_only = FALSE,
  quiet = FALSE,
  timeout = 600,
  retries = 3,
  retry_wait = 5,
  ...
)
```

Arguments

records	Species records returned by tol_species_index() or tol_search_species() . If omitted, filters are applied to tol_species_index() .
dest_dir	Local directory where FASTA files should be stored. If NULL, files are downloaded to a session temporary directory.
manifest_path	Optional CSV manifest path. Use NULL to skip writing. If dest_dir = NULL, no manifest is written unless this is explicitly set.
overwrite	Replace existing local FASTA files.
validate_only	Report local status without downloading.
quiet	Passed to utils::download.file() .
timeout	Download timeout in seconds.
retries	Number of attempts for each file.
retry_wait	Seconds to wait between attempts.
...	Filters passed to tol_search_species() when records is omitted.

Value

A data frame with one row per requested FASTA.

tol_download_release *Download and validate a complete Kew Tree of Life release*

Description

This recursively indexes the remote release, downloads every file, validates local file sizes against the remote index, and writes a CSV manifest.

Usage

```
tol_download_release(
  dest_dir = file.path("raw-data", normalize_release(release)),
  path = "",
  release = "current_release",
  manifest_path = file.path("raw-data", "kew_download_manifest.csv"),
  overwrite = FALSE,
  validate_only = FALSE,
  quiet = FALSE,
  base_url = tol_base_url(),
  timeout = 600,
  retries = 3,
  retry_wait = 5
)
```

Arguments

dest_dir	Local destination directory for the release contents.
path	Directory path inside the release to download. Use "" for the whole release, or paths such as "fasta/by_gene".
release	Release directory. Defaults to Kew's current release link.
manifest_path	Local CSV manifest path.
overwrite	Replace existing local files.
validate_only	Only index and validate existing local files.
quiet	Passed to <code>utils::download.file()</code> .
base_url	Base public URL.
timeout	Download timeout in seconds.
retries	Number of attempts for each file.
retry_wait	Seconds to wait between attempts.

Value

A data frame with release id, remote paths, local paths, sizes, and validation status.

tol_export_fasta	<i>Export downloaded FASTA files and a manifest</i>
------------------	---

Description

Export downloaded FASTA files and a manifest

Usage

```
tol_export_fasta(
  plan,
  dest_dir = file.path("raw-data", "fasta", "by_recovery"),
  manifest_path = file.path(dest_dir, "fasta_export_manifest.csv"),
  overwrite = FALSE
)
```

Arguments

plan	A data frame returned by tol_download_fasta() or tol_resolve_fasta() .
dest_dir	Permanent destination directory for FASTA files.
manifest_path	Optional CSV manifest path. If NULL, no manifest is written.
overwrite	Replace files that already exist in dest_dir.

Value

A tibble with copied file paths and export status.

tol_fasta_long	<i>Convert attached FASTA data to a tidy long table</i>
----------------	---

Description

Convert attached FASTA data to a tidy long table

Usage

```
tol_fasta_long(plan, fasta_column = "fasta")
```

Arguments

plan	A data frame returned by tol_attach_fasta() or a FASTA plan with local files.
fasta_column	Name of the FASTA list-column. If it is not present, tol_attach_fasta() is called.

Value

A tibble with one row per sequence.

tol_fasta_summary	<i>Summarise FASTA content by species record</i>
-------------------	--

Description

Summarise FASTA content by species record

Usage

```
tol_fasta_summary(plan)
```

Arguments

plan	A data frame returned by tol_attach_fasta() or a FASTA plan with local files.
------	---

Value

A tibble with one row per species/specimen FASTA.

tol_index	<i>List files and directories in a Kew Tree of Life remote index</i>
-----------	--

Description

List files and directories in a Kew Tree of Life remote index

Usage

```
tol_index(path = "", release = "current_release", base_url = tol_base_url())
```

Arguments

path	Directory path inside the release.
release	Release directory. Defaults to Kew's current release link.
base_url	Base public URL.

Value

A data frame with name, path, url, and is_dir.

tol_known_bundle	<i>Known download bundles</i>
------------------	-------------------------------

Description

Alias for `tol_known_bundles()`.

Usage

```
tol_known_bundle()
```

Value

A data frame with bundle names and included remote paths.

tol_known_bundles	<i>Known download bundles</i>
-------------------	-------------------------------

Description

Known download bundles

Usage

```
tol_known_bundles()
```

Value

A data frame with bundle names and included remote paths.

tol_manifest	<i>Read a downloaded Kew Tree of Life manifest</i>
--------------	--

Description

Read a downloaded Kew Tree of Life manifest

Usage

```
tol_manifest(file, manifest = NULL, ...)
```

Arguments

file	Local path to a manifest file.
manifest	Manifest type. Used only to assign column names when known.
...	Additional arguments passed to <code>utils::read.delim()</code> .

Value

A data frame.

tol_match_species	<i>Match requested species names against the Kew Tree of Life index</i>
-------------------	---

Description

Match requested species names against the Kew Tree of Life index

Usage

```
tol_match_species(
  species,
  index = tol_species_index(),
  fuzzy = FALSE,
  max_distance = NULL,
  multiple = c("all", "best"),
  ignore_case = TRUE
)
```

Arguments

species	Character vector of requested scientific names.
index	A species index returned by <code>tol_species_index()</code> . If omitted, the built-in <code>tol_species</code> dataset is used.
fuzzy	If TRUE, return the closest available name when no exact match is found.
max_distance	Maximum edit distance for fuzzy matching. If NULL, a conservative threshold is computed from each requested name.
multiple	How to handle multiple records for the same matched species. "all" returns all records; "best" keeps the record with the highest <code>no_of_genes_recovered</code> and then <code>no_of_bp_recovered</code> .
ignore_case	Ignore case when matching names.

Value

A tibble with requested names, match status, matched records, and FASTA URLs where available.

`tol_plot_fasta_summary`*Plot FASTA summary with ggplot2*

Description

Plot FASTA summary with ggplot2

Usage

```
tol_plot_fasta_summary(summary)
```

Arguments

`summary` A summary returned by `tol_fasta_summary()`.

Value

A ggplot object.

`tol_plot_gene_recovery`*Plot recovered gene counts with ggplot2*

Description

Plot recovered gene counts with ggplot2

Usage

```
tol_plot_gene_recovery(records)
```

Arguments

`records` Species records returned by `tol_species_index()` or `tol_search_species()`.

Value

A ggplot object.

tol_plot_tree	<i>Plot an hclust tree with ggplot2</i>
---------------	---

Description

Plot an hclust tree with ggplot2

Usage

```
tol_plot_tree(tree_result, label_offset = NULL, label_size = 3)
```

Arguments

tree_result	A result returned by tol_build_gene_tree() or an hclust object.
label_offset	Numeric offset used to place species labels beyond the end of each terminal branch. If NULL, an offset is computed from the tree height.
label_size	Label text size.

Value

A ggplot object.

tol_read_fasta	<i>Read a FASTA file</i>
----------------	--------------------------

Description

Read a FASTA file

Usage

```
tol_read_fasta(file, as = c("data.frame", "list", "text"))
```

Arguments

file	Local FASTA file.
as	Output format. "data.frame" returns one row per sequence with header, sequence, and width; "list" returns a named character vector; "text" returns raw file lines.

Value

A data frame, named character vector, or character vector.

tol_release_url	<i>Build a URL for a Kew Tree of Life release</i>
-----------------	---

Description

Build a URL for a Kew Tree of Life release

Usage

```
tol_release_url(
  path = "",
  release = "current_release",
  base_url = tol_base_url()
)
```

Arguments

path	Optional path inside the release directory.
release	Release directory. Use "current_release" or "current" for Kew's current release link.
base_url	Base public URL.

Value

A character scalar URL.

tol_resolve_fasta	<i>Resolve FASTA download targets for selected species records</i>
-------------------	--

Description

Resolve FASTA download targets for selected species records

Usage

```
tol_resolve_fasta(records, dest_dir = NULL)
```

Arguments

records	Species records returned by tol_species_index() or tol_search_species() .
dest_dir	Local directory where FASTA files should be stored. If NULL, a session temporary directory is used.

Value

A data frame describing FASTA URLs and local paths.

tol_save_fasta	Save downloaded FASTA files to a permanent directory
----------------	--

Description

Save downloaded FASTA files to a permanent directory

Usage

```
tol_save_fasta(  
  plan,  
  dest_dir = file.path("raw-data", "fasta", "by_recovery"),  
  overwrite = FALSE  
)
```

Arguments

plan	A data frame returned by tol_download_fasta() or tol_resolve_fasta() .
dest_dir	Permanent destination directory.
overwrite	Replace files that already exist in dest_dir.

Value

A copy of plan with updated local_path and status.

tol_search_species	Search Kew Tree of Life species records
--------------------	---

Description

Search Kew Tree of Life species records

Usage

```
tol_search_species(  
  index = tol_species_index(),  
  sequence_id = NULL,  
  order = NULL,  
  family = NULL,  
  genus = NULL,  
  specific_epithet = NULL,  
  scientific_name = NULL,  
  query = NULL,  
  ignore_case = TRUE  
)
```

Arguments

index	A species index returned by <code>tol_species_index()</code> . If omitted, the built-in <code>tol_species</code> dataset is used.
sequence_id	Optional sequence identifier or vector of identifiers.
order, family, genus, specific_epithet, scientific_name	Optional taxonomic filters.
query	Optional free-text query.
ignore_case	Ignore case in text filters.

Value

A filtered data frame.

tol_set_base_url	<i>Set the Kew Tree of Life base URL for the current R session</i>
------------------	--

Description

Set the Kew Tree of Life base URL for the current R session

Usage

```
tol_set_base_url(base_url = "https://sftp.kew.org/pub/treeoflife")
```

Arguments

base_url	Base URL without the release directory.
----------	---

Value

The previous value of option `rtreeoflife.base_url`, invisibly.

tol_species	<i>Kew Tree of Life species index</i>
-------------	---------------------------------------

Description

A normalized species/specimen index exported from Kew Tree of Life Explorer. Each row represents one Tree of Life sequence record and includes taxonomic metadata, specimen metadata, recovery statistics, and the FASTA URL for the corresponding `fasta/by_recovery` file.

Usage

```
tol_species
```

Format

A tibble with 20485 rows and 16 columns:

sequence_id Kew Tree of Life sequence identifier.
data_source Project, dataset, or repository source.
order Taxonomic order.
family Taxonomic family.
genus Taxonomic genus.
specific_epithet Specific epithet or infraspecific name text.
specimen_reference Voucher or specimen reference.
specimen_barcode Specimen barcode where available.
collection_date Collection year where available.
country_of_origin Country of origin where available.
material_sampled Sample material type.
no_of_genes_recovered Number of recovered Angiosperms353 genes.
no_of_bp_recovered Number of recovered base pairs.
fasta_file_url Remote FASTA URL for this sequence recovery.
scientific_name Combined genus and specific epithet.
fasta_file_name Basename of fasta_file_url.

Source

Kew Tree of Life Explorer species list.

tol_species_index	<i>Read the Kew Tree of Life species index</i>
-------------------	--

Description

Read the Kew Tree of Life species index

Usage

```
tol_species_index(file = NULL, use_package_data = TRUE)
```

Arguments

file Local path to all_species_list.csv. If NULL, the package dataset [tol_species](#) is returned.
use_package_data Use the built-in [tol_species](#) dataset when file = NULL.

Value

A data frame with normalized column names and derived fields.

Index

* datasets

- tol_species, 18
- tol_attach_fasta, 2
- tol_attach_fasta(), 4, 10, 11
- tol_base_url, 3
- tol_build_gene_tree, 3
- tol_build_gene_tree(), 15
- tol_common_genes, 4
- tol_data_dir, 4
- tol_download, 5
- tol_download_bundle, 6
- tol_download_directory, 7
- tol_download_fasta, 8
- tol_download_fasta(), 3, 10, 17
- tol_download_release, 9
- tol_download_release(), 7
- tol_export_fasta, 10
- tol_fasta_long, 10
- tol_fasta_summary, 11
- tol_fasta_summary(), 14
- tol_index, 11
- tol_known_bundle, 12
- tol_known_bundles, 12
- tol_known_bundles(), 6, 12
- tol_manifest, 12
- tol_match_species, 13
- tol_plot_fasta_summary, 14
- tol_plot_gene_recovery, 14
- tol_plot_tree, 15
- tol_read_fasta, 15
- tol_read_fasta(), 3
- tol_release_url, 16
- tol_resolve_fasta, 16
- tol_resolve_fasta(), 3, 10, 17
- tol_save_fasta, 17
- tol_save_fasta(), 3
- tol_search_species, 17
- tol_search_species(), 8, 14, 16
- tol_set_base_url, 18

- tol_species, 13, 18, 18, 19
- tol_species_index, 19
- tol_species_index(), 8, 13, 14, 16, 18
- utils::download.file(), 5–9
- utils::read.delim(), 13