

Package: taxastand (via r-universe)

September 23, 2026

Title Taxonomic Name Standardization

Version 2.0.0.9000

Description Standardizes taxonomic names by matching species names to entries in a user-supplied reference taxonomy, then resolving synonyms to their accepted names. Fuzzy matching on parsed name components (genus, epithet, author, infraspecific rank, etc.) allows for orthographic differences between the query and reference names. This facilitates reproducible joining of datasets that use different taxonomic synonyms for the same species.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

URL <https://joelnitta.github.io/taxastand/>,
<https://github.com/joelnitta/taxastand>

BugReports <https://github.com/joelnitta/taxastand/issues>

Imports assertthat, digest, dplyr, fs, glue, magrittr, tibble, tidyr

Roxygen list(markdown = TRUE, roclets = c("`collate", ``namespace", ``rd", ``roxygenals::global_roclet"))

Depends R (>= 4.1.0)

Suggests covr, rmarkdown, knitr, roxygenals (>= 1.0.0), testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

Config/roxygen2/version 8.0.0

Config/pak/sysreqs cmake make libicu-dev libuv1-dev

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

Date/Publication 2026-07-25 04:26:18 UTC

RemoteUrl <https://github.com/joelnitta/taxastand>

RemoteRef HEAD

RemoteSha caf25f01a502debacc620cfbaa7caf25d8d34aa

Contents

filmy_taxonomy	2
ts_match_names	2
ts_parse_names	5
ts_resolve_names	7
Index	9

filmy_taxonomy	<i>Taxonomy of filmy ferns (family Hymenophyllaceae)</i>
----------------	--

Description

A dataset containing taxonomic names and associated metadata for the fern family Hymenophyllaceae. Downloaded from the [Catalog of Life](#), Version 1.5. All columns formatted according to [Darwin Core standard](#). Only includes taxa at the species or infraspecies level.

Usage

filmy_taxonomy

Format

A data frame with 2729 rows and 31 variables.

Source

<https://www.catalogueoflife.org/>

ts_match_names	<i>Match taxonomic names to a reference</i>
----------------	---

Description

Allows for orthographic differences between query and reference by using fuzzy matching on parsed taxonomic names.

Usage

```
ts_match_names(
  query,
  reference,
  manual_match = NULL,
  max_dist = 10,
  match_no_auth = FALSE,
  match_canon = FALSE,
  collapse_infra = FALSE,
  collapse_infra_exclude = NULL,
  simple = FALSE,
  tbl_out = getOption("ts_tbl_out", default = FALSE)
)
```

Arguments

query	Character vector or dataframe; taxonomic names to be queried. If a character vector, missing values not allowed and all values must be unique. If a dataframe, should be taxonomic names parsed with ts_parse_names() .
reference	Character vector or dataframe; taxonomic names to use as reference. If a character vector, missing values not allowed and all values must be unique. If a dataframe, should be taxonomic names parsed with ts_parse_names() .
manual_match	Optional. Dataframe of manually matched names that will override any results from the name-matching engine. Must include columns, query and match. Can only be used if query is a character vector.
max_dist	Max Levenshtein distance to allow during fuzzy matching (total insertions, deletions and substitutions). Default: 10.
match_no_auth	Logical; If no author is given in the query and the name (without author) occurs only once in the reference, accept the name in the reference as a match. Default: to not allow such a match (FALSE).
match_canon	Logical; Allow a "canonical name" match if only the genus, species epithet, and infraspecific epithet (if present) match exactly. Default: to not allow such a match (FALSE).
collapse_infra	Logical; if the specific epithet and infraspecific epithet are the same, drop the infraspecific rank and epithet from the query.
collapse_infra_exclude	Character vector; taxonomic names to exclude from collapsing with collapse_infra. Any names used must match those in query exactly, or they won't be excluded.
simple	Logical; return the output in a simplified format with only the query name, matched reference name, and match type. Default: FALSE.
tbl_out	Logical vector of length 1; should a tibble be returned? If FALSE (default), output will be a data.frame. This argument can be controlled via the option ts_tbl_out; see Examples.

Details

Names are matched in two steps:

1. Scientific names are parsed into their component parts (genus, species, variety, author, etc).
2. Names are fuzzily matched following taxonomic rules using the component parts.

For more information on rules used for matching, [see taxon-tools manual](#), which uses the same logic.

Parsing is fairly fast (much faster than matching) but can take some time if the number of names is very large. If multiple queries will be made (e.g., to the same large reference database), it is recommended to first parse the names using `ts_parse_names()`, and use the results as input to query and/or reference.

`collapse_infra` is useful in situations where the reference database does not use names that have the same specific epithet and infraspecific epithet. For example, reference name "Blechnum lunare" and query "Blechnum lunare var. lunare". In this case, if `collapse_infra` is TRUE, "Blechnum lunare" will be queried instead of "Blechnum lunare var. lunare". Note that the `match_type` will be "exact" even though the literal query and the matched name are different (see example below).

Value

Dataframe with the following columns (if `simple` is FALSE):

- `query`: Query name
- `reference`: Matched reference name
- `match_type`: Type of match (for a summary of match types, [see taxon-tools manual](#))
- `id_query`: Unique ID of query
- `id_ref`: Unique ID of reference
- `genus_hybrid_sign_query`: Genus hybrid sign in query
- `genus_name_query`: Genus name of query
- `species_hybrid_sign_query`: Species hybrid sign in query
- `specific_epithet_query`: Specific epithet of query
- `infraspecific_rank_query`: Infraspecific rank of query
- `infraspecific_epithet_query`: Infraspecific epithet of query
- `author_query`: Taxonomic author of query
- `genus_hybrid_sign_ref`: Genus hybrid sign in reference
- `genus_name_ref`: Genus name of reference
- `species_hybrid_sign_ref`: Species hybrid sign in reference
- `specific_epithet_ref`: Specific epithet of reference
- `infraspecific_rank_ref`: Infraspecific rank of reference
- `infraspecific_epithet_ref`: Infraspecific epithet of reference
- `author_ref`: Taxonomic author of reference

If `simple` is TRUE, only return the first three columns above.

Examples

```

ts_match_names(
  "Crepidomanes minutus",
  c("Crepidomanes minutum", "Hymenophyllum polyanthos"),
  simple = TRUE
)

# If names are too distant, they won't match
ts_match_names(
  query = "Crepidblah foo",
  reference = c("Crepidomanes minutum", "Hymenophyllum polyanthos"),
  simple = TRUE
)

# But we can force a match manually
ts_match_names(
  query = "Crepidblah foo",
  reference = c("Crepidomanes minutum", "Hymenophyllum polyanthos"),
  manual_match = data.frame(
    query = c("Crepidblah foo"),
    match = c("Crepidomanes minutum")
  ),
  simple = TRUE
)

# If you always want tibble output without specifying `tbl_out = TRUE`
# every time, set the option:
options(ts_tbl_out = TRUE)
ts_match_names(
  "Crepidomanes minutus",
  c("Crepidomanes minutum", "Hymenophyllum polyanthos")
)

# Example using collapse_infra argument
ts_match_names(
  c(
    "Crepidomanes minutus", "Blechnum lunare var. lunare",
    "Blechnum lunare", "Bar foo var. foo", "Bar foo"
  ),
  c(
    "Crepidomanes minutum", "Hymenophyllum polyanthos", "Blechnum lunare",
    "Bar foo"
  ),
  collapse_infra = TRUE,
  collapse_infra_exclude = "Bar foo var. foo",
  simple = TRUE
)

```

Description

Parses scientific names into their component parts (genus, species, variety, author, etc).

Usage

```
ts_parse_names(
  taxa,
  tbl_out = getOption("ts_tbl_out", default = FALSE),
  quiet = FALSE
)
```

Arguments

taxa	Character vector; taxon names to be parsed. Missing values not allowed. Must all be unique.
tbl_out	Logical vector of length 1; should a tibble be returned? If FALSE (default), output will be a data.frame. This argument can be controlled via the option <code>ts_tbl_out</code> ; see Examples.
quiet	Logical; if TRUE, suppress warning messages that would normally be issued

Value

A dataframe including the following columns.

- id: A unique ID number assigned to the input name
- name: The input name
- genus_hybrid_sign: Hybrid sign for genus
- genus_name: Genus name
- species_hybrid_sign: Hybrid sign for species
- specific_epithet: Specific epithet (name)
- infraspecific_rank: Infraspecific rank
- infraspecific_epithet: Infraspecific epithet (name)
- author: Name of taxon

Examples

```
ts_parse_names("Foogenus x barspecies var. foosubsp (L.) F. Bar")
ts_parse_names(
  "Foogenus x barspecies var. foosubsp (L.) F. Bar",
  tbl_out = TRUE
)

# If you always want tibble output without specifying `tbl_out = TRUE`
# every time, set the option:
options(ts_tbl_out = TRUE)
ts_parse_names("Foogenus x barspecies var. foosubsp (L.) F. Bar")
ts_parse_names("Crepidomanes minutum (Blume) K. Iwats.")
```

ts_resolve_names	<i>Resolve synonyms in taxonomic names</i>
------------------	--

Description

After matching taxonomic names to a reference, some may match synonyms. This function resolves synonyms to their accepted names.

Usage

```
ts_resolve_names(
  query,
  ref_taxonomy,
  max_dist = 10,
  match_no_auth = FALSE,
  match_canon = FALSE,
  collapse_infra = FALSE,
  collapse_infra_exclude = NULL,
  tbl_out = getOption("ts_tbl_out", default = FALSE)
)
```

Arguments

query	Character vector or dataframe; taxonomic names to be resolved. If a character vector, missing values not allowed and all values must be unique. If a dataframe, should be taxonomic names matched with ts_match_names() .
ref_taxonomy	Dataframe; reference taxonomic data adhering to the Darwin Core standard with the following columns: - taxonID: Unique identifier for each taxon. - acceptedNameUsageID: If the taxon is a synonym, the unique identifier for the accepted name - taxonomicStatus: The status of the use of the scientificName as a label for the taxon. - scientificName: The full scientific name of the taxon, with authorship and date information if known.
max_dist	Max Levenshtein distance to allow during fuzzy matching (total insertions, deletions and substitutions). Default: 10.
match_no_auth	Logical; If no author is given in the query and the name (without author) occurs only once in the reference, accept the name in the reference as a match. Default: to not allow such a match (FALSE).
match_canon	Logical; Allow a "canonical name" match if only the genus, species epithet, and infraspecific epithet (if present) match exactly. Default: to not allow such a match (FALSE).
collapse_infra	Logical; if the specific epithet and infraspecific epithet are the same, drop the infraspecific rank and epithet from the query. For more information, see ts_match_names() .

collapse_infra_exclude	Character vector; taxonomic names to exclude collapsing with collapse_infra. Any names used must match those in query exactly, or they won't be excluded.
tbl_out	Logical vector of length 1; should a tibble be returned? If FALSE (default), output will be a data.frame. This argument can be controlled via the option ts_tbl_out; see Examples.

Details

query can take as input either a character vector of taxonomic names, or the output of `ts_match_names()`. If the former, it will run `ts_match_names()` to match the query to ref_taxonomy, then resolve synonyms. If the latter, the scientific names in ref_taxonomy should be the same used as reference with `ts_match_names()` (this is not checked).

ref_taxonomy must be taxonomic data adhering to the **Darwin Core standard**. Darwin Core includes many terms, but only four (taxonID, acceptedNameUsageID, taxonomicStatus, and scientificName) are required for this function.

Value

Dataframe; results of resolving synonyms in matched taxonomic names. Includes the following columns:

- query: Query name
- resolved_name: Accepted name after resolving synonyms
- matched_name: Name matched to query
- resolved_status: Taxonomic status of the resolved name (same as taxonomicStatus in ref_taxonomy)
- matched_status: Taxonomic status of the matched name (same as taxonomicStatus in ref_taxonomy)
- match_type: Type of match (for a summary of match types, see [taxon-tools manual](#))

Names that could not be matched or resolve to multiple, different synonyms have NA for resolved_name.

Examples

```
# Load reference taxonomy in Darwin Core format
data(filmy_taxonomy)

ts_resolve_names("Gonocormus minutum", filmy_taxonomy)
# If you always want tibble output without specifying `tbl_out = TRUE`
# every time, set the option:
options(ts_tbl_out = TRUE)
ts_resolve_names("Gonocormus minutum", filmy_taxonomy)
```

Index

* **datasets**

filmy_taxonomy, [2](#)

filmy_taxonomy, [2](#)

ts_match_names, [2](#), [7](#), [8](#)

ts_parse_names, [3](#), [4](#), [5](#)

ts_resolve_names, [7](#)