

Using **plantR** to Manage Animal Taxonomy with the CTFB Backbone

Carlos Eduardo Pinto* Renato A. Ferreira de Lima†

17 outubro 2025

Contents

1	Installing plantR	1
2	A practical example	2
3	Preparing names using <code>fixSpecies()</code>	2
3.1	Internal functions	3
4	Validating taxon names using <code>prepSpecies()</code>	3
4.1	Internal functions	4
5	Validating family names using <code>prepFamily()</code>	5
6	Brief code summary	5
7	Citation	6

1 Installing **plantR**

Install the packages from GitHub if needed and load **plantR**.

```
if (!requireNamespace("remotes"))  
  install.packages("remotes")  
library(remotes)  
  
if (!requireNamespace("plantR"))  
  install_github("LimaRAF/plantR")  
  
if (!requireNamespace("plantRdata"))  
  install_github("LimaRAF/plantRdata")  
  
library(plantR)
```

*Departamento de Ciências Biológicas, ESALQ, Universidade de São Paulo

†Departamento de Ciências Biológicas, ESALQ, Universidade de São Paulo, <https://github.com/LimaRAF>

2 A practical example

We will start with a small list of **names of animal species** that includes common issues (misspellings, synonyms, wrong capitalization, invalid names). We will save it in an objects called **names**:

```
names <- c(
  "Apis mellifera Linnaeus, 1758",      # name accepted with author
  "Apis melifera",                     # misspelling
  "Apis cf. mellifera",                 # open nomenclature
  "Ancyloscelis armatus",               # synonym of Ancyloscelis apiformis (Fabricius, 1793)
  "Centris aenea",                      # name accepted without author
  "Centris rufa",                       # synonym of Centris aenea Lepeletier, 1841
  "Centris Rhodoprocta Moure & Seabra, 1960", # wrong capitalization
  "Lutjanus purpureus",                 # synonym of Lutjanus campechanus (Poey, 1860)
  "Parotocinclus amazonensis",          # invalid in CTFB; no synonym for this name
  "Panthera onca",                      # name accepted without author
  "Solenopsis bicolor (Emery, 1906)",    # name accepted with author
  "Eucopricus columbi MacLeay, 1819",    # synonym of Sulcophanaeus columbi (MacLeay, 1819)
  "Eucopricus sp.1"                     # incomplete identification
)
```

3 Preparing names using fixSpecies()

`fixSpecies()` formats and cleans names (notation, casing, authorship split, notation flags). It accepts either a character vector or a data frame (default `scientificName` column).

```
names_fixed <- fixSpecies(names)
names_fixed[, -c(2,4)]
```

```
#>      scientificName      scientificName.new
#> 1      Apis mellifera Linnaeus, 1758      Apis mellifera
#> 2      Apis melifera                      Apis melifera
#> 3      Apis cf. mellifera                  Apis mellifera
#> 4      Ancyloscelis armatus                Ancyloscelis armatus
#> 5      Centris aenea                      Centris aenea
#> 6      Centris rufa                      Centris rufa
#> 7      Centris Rhodoprocta Moure & Seabra, 1960      Centris rhodoprocta
#> 8      Lutjanus purpureus                  Lutjanus purpureus
#> 9      Parotocinclus amazonensis          Parotocinclus amazonensis
#> 10     Panthera onca                      Panthera onca
#> 11     Solenopsis bicolor (Emery, 1906)    Solenopsis bicolor
#> 12     Eucopricus columbi MacLeay, 1819    Eucopricus columbi
#> 13     Eucopricus sp.1                    Eucopricus sp.1
#>      scientificNameStatus
#> 1      name_w_authors
#> 2      possibly_ok
#> 3      conferre
#> 4      possibly_ok
#> 5      possibly_ok
#> 6      possibly_ok
#> 7      name_w_wrong_case|name_w_authors
```

```
#> 8           possibly_ok
#> 9           possibly_ok
#> 10          possibly_ok
#> 11          name_w_authors
#> 12          name_w_authors
#> 13           indet
```

3.1 Internal functions

For this specific list, some functions may not change anything (which is fine). The goal is to illustrate correct usage of the internal functions on the **same** input vector of names.

```
fixIndet(names)      # detects undetermined names (e.g., "sp.", "indet")
fixCase(names)       # fixes casing (e.g., "Centris Rhodoprocta")
fixAuthors(names)    # splits taxon and author names, if present
```

4 Validating taxon names using prepSpecies()

Next, validate names against the CTFB backbone (ctfbNames) from plantRdata by loading it into the Global Environment and passing it to db.

```
# load the CTFB backbone (ctfbNames) into the Global Environment
utils::data("ctfbNames", package = "plantRdata")

# validate against CTFB
names_valid <- prepSpecies(
  names_fixed,
  tax.names = c("scientificName.new", "scientificNameAuthorship.new"),
  db = ctfbNames)

names_valid[, -c(2,3,4,9,11)]
```

#>	scientificName	scientificNameStatus	suggestedFamily
#> 1	Apis mellifera Linnaeus, 1758	name_w_authors	Apidae
#> 2	Apis melifera	possibly_ok	Apidae
#> 3	Apis cf. mellifera	conferre	Apidae
#> 4	Ancyloscelis armatus	possibly_ok	Apidae
#> 5	Centris aenea	possibly_ok	Apidae
#> 6	Centris rufa	possibly_ok	Apidae
#> 7	Centris Rhodoprocta Moure & Seabra, 1960	name_w_wrong_case name_w_authors	Apidae
#> 8	Lutjanus purpureus	possibly_ok	Lutjanidae
#> 9	Parotocinclus amazonensis	possibly_ok	Loricariidae
#> 10	Panthera onca	possibly_ok	Felidae
#> 11	Solenopsis bicolor (Emery, 1906)	name_w_authors	Formicidae
#> 12	Eucopricus columbi MacLeay, 1819	name_w_authors	Scarabaeidae
#> 13	Eucopricus sp.1	indet	Scarabaeidae

#>	suggestedName	suggestedAuthorship	tax.notes
#> 1	Apis mellifera	Linnaeus, 1758	name accepted
#> 2	Apis mellifera	Linnaeus, 1758	name misspelled
#> 3	Apis mellifera	Linnaeus, 1758	name accepted

```

#> 4      Ancyloscelis apiformis      (Fabricius, 1793)      replaced synonym
#> 5              Centris aenea      Lepeletier, 1841      name accepted
#> 6              Centris aenea      Lepeletier, 1841      replaced synonym
#> 7      Centris rhodoprocta Moure & Seabra, 1960      name accepted
#> 8      Lutjanus campechanus      (Poey, 1860)      replaced synonym
#> 9 Parotocinclus amazonensis      Garavello, 1977 synonym not replaced
#> 10             Panthera onca      (Linnaeus, 1758)      name accepted
#> 11             Solenopsis bicolor      (Emery, 1906)      name accepted
#> 12      Sulcophanaeus columbi      (MacLeay, 1819)      replaced synonym
#> 13      Sulcophanaeus d'Olsoufieff, 1924      replaced synonym
#>
#>      scientificNameFull
#> 1      Apis mellifera Linnaeus, 1758
#> 2      Apis mellifera Linnaeus, 1758
#> 3      Apis mellifera Linnaeus, 1758
#> 4      Ancyloscelis apiformis (Fabricius, 1793)
#> 5              Centris aenea Lepeletier, 1841
#> 6              Centris aenea Lepeletier, 1841
#> 7      Centris rhodoprocta Moure & Seabra, 1960
#> 8      Lutjanus campechanus (Poey, 1860)
#> 9 Parotocinclus amazonensis Garavello, 1977
#> 10             Panthera onca (Linnaeus, 1758)
#> 11             Solenopsis bicolor (Emery, 1906)
#> 12      Sulcophanaeus columbi (MacLeay, 1819)
#> 13      Sulcophanaeus d'Olsoufieff, 1924

```

Tip 1: for large name lists, consider altering the argument `split.letters`, `parallel`, and `cores`. The minimal fuzzy similarity is controlled by `sug.dist`.

Tip 2: The maximum distance in fuzzy matching (defaults to 10%) is controlled by the argument `sug.dist`.

4.1 Internal functions

`nameMatching()` is the internal function used for exact and fuzzy matching. Below, we demonstrate it using the same names (reference names are the “accepted/standardized” targets):

```

input_names <- c(
  "Apis mellifera Linnaeus, 1758",
  "Apis mellifica",
  "Ancyloscelis apiformis",
  "Centris aenea",
  "Lutjanus purpureus",
  "Parotocinclus amazonensis",
  "Eucopricus columbi MacLeay, 1819"
)

ref_names <- c(
  "Apis mellifera Linnaeus, 1758",
  "Ancyloscelis apiformis (Fabricius, 1793)",
  "Centris aenea",
  "Centris rhodoprocta Moure & Seabra, 1960",
  "Lutjanus campechanus (Poey, 1860)",
  "Panthera onca",

```

```

"Coelonertus baridioides Solari & Solari, 1906",
"Solenopsis bicolor (Emery, 1906)",
"Sulcophanaeus columbi (MacLeay, 1819)"
)

nameMatching(input_names, ref_names)

#> [1] 1 1 2 3 5 NA 9

```

5 Validating family names using prepFamily()

plantR contains an internal dictionary of valid family names which can be used via the function `prepFamily()`. Currently, valid family names are available only for plants. But a similar procedure will be included for animals in the near future. So, for now, the function does not change the input family names.

```

names_valid <- prepFamily(names_valid,
  fam.name = "suggestedFamily",
  spp.name = "scientificName.new",
  kingdom = "animalia",
  db = ctfbNames)

```

6 Brief code summary

A compact two-step workflow (CTFB-only):

```

# 1) Standardize
names_fixed <- fixSpecies(names)

# 2) Validate against the CTFB backbone
utils::data("ctfbNames", package = "plantRdata")
names_valid <- prepSpecies(
  names_fixed,
  tax.names = c("scientificName.new", "scientificNameAuthorship.new"),
  db = ctfbNames
)

names_valid[, c("scientificName.new", "scientificNameFull", "tax.notes")]

```

#>	scientificName.new	scientificNameFull	tax.notes
#> 1	Apis mellifera	Apis mellifera Linnaeus, 1758	name accepted
#> 2	Apis melifera	Apis mellifera Linnaeus, 1758	name misspelled
#> 3	Apis mellifera	Apis mellifera Linnaeus, 1758	name accepted
#> 4	Ancyloscelis armatus	Ancyloscelis apiformis (Fabricius, 1793)	replaced synonym
#> 5	Centris aenea	Centris aenea Lepeletier, 1841	name accepted
#> 6	Centris rufa	Centris aenea Lepeletier, 1841	replaced synonym
#> 7	Centris rhodoprocta	Centris rhodoprocta Moure & Seabra, 1960	name accepted
#> 8	Lutjanus purpureus	Lutjanus campechanus (Poey, 1860)	replaced synonym
#> 9	Parotocinclus amazonensis	Parotocinclus amazonensis Garavello, 1977	synonym not replaced
#> 10	Panthera onca	Panthera onca (Linnaeus, 1758)	name accepted

#> 11	Solenopsis bicolor	Solenopsis bicolor (Emery, 1906)	name accepted
#> 12	Eucopricus columbi	Sulcophanaeus columbi (MacLeay, 1819)	replaced synonym
#> 13	Eucopricus sp.1	Sulcophanaeus d'Olsoufieff, 1924	replaced synonym

Or, even simpler, using the wrapper `formatTax()`:

```
names_df <- data.frame(scientificName = names)
names_df_valid <- formatTax(names_df,
                             db = ctfbNames,
                             kingdom = "animalia")
```

7 Citation

If you use **plantR**, please cite it as:

Lima, R.A.F., Sánchez-Tapia, A., Mortara, S.R., ter Steege, H., Siqueira, M.F. (2021). *plantR*: An R package and workflow for managing species records from biological collections. *Methods in Ecology and Evolution* 14(2): 332–339. <https://doi.org/10.1101/2021.04.06.437754>

And please also cite the taxonomic backbones that you used:

Boeger, W., & Valim, M. P. (2024). Brazilian Zoology Group 2023 (version 1.1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.10498290>