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

Carlos Eduardo Pinto* Renato A. Ferreira de Lima†

07 outubro 2025

Contents

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, 179
  "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             scientificNameSt
#> 1             Apis mellifera Linnaeus, 1758            Apis mellifera                   name_w_aut
#> 2                               Apis melifera             Apis melifera                      possibl
#> 3                          Apis cf. mellifera            Apis mellifera                         conf
#> 4                        Ancyloscelis armatus      Ancyloscelis armatus                      possibl
#> 5                               Centris aenea             Centris aenea                      possibl
#> 6                                Centris rufa              Centris rufa                      possibl
#> 7  Centris Rhodoprocta Moure & Seabra, 1960       Centris rhodoprocta name_w_wrong_case|name_w_aut
#> 8                            Lutjanus purpureus         Lutjanus purpureus                      possibl
#> 9                  Parotocinclus amazonensis Parotocinclus amazonensis                      possibl
#> 10                               Panthera onca             Panthera onca                      possibl
#> 11         Solenopsis bicolor (Emery, 1906)        Solenopsis bicolor                   name_w_aut
#> 12         Eucopricus columbi MacLeay, 1819        Eucopricus columbi                   name_w_aut
#> 13                             Eucopricus sp.1           Eucopricus sp.1                          i
```

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")
```

```
#>  [1] "Apis mellifera Linnaeus, 1758"            "Apis melifera"
#>  [3] "Apis cf. mellifera"                       "Ancyloscelis armatus"
#>  [5] "Centris aenea"                            "Centris rufa"
#>  [7] "Centris Rhodoprocta Moure & Seabra, 1960" "Lutjanus purpureus"
#>  [9] "Parotocinclus amazonensis"                "Panthera onca"
#> [11] "Solenopsis bicolor (Emery, 1906)"         "Eucopricus columbi MacLeay, 1819"
#> [13] "Eucopricus sp.1"
```

```
fixCase(names)        # fixes casing (e.g., "Centris Rhodoprocta")
```

```
#>              Apis mellifera Linnaeus, 1758                            Apis melifera
#>            "Apis mellifera Linnaeus, 1758"                          "Apis melifera"
#>                       Ancyloscelis armatus                             Centris aenea
#>                     "Ancyloscelis armatus"                           "Centris aenea"
#>   Centris Rhodoprocta Moure & Seabra, 1960                       Lutjanus purpureus
#> "Centris rhodoprocta Moure & Seabra, 1960"                     "Lutjanus purpureus"
#>                               Panthera onca          Solenopsis bicolor (Emery, 1906)           Eu
#>                             "Panthera onca"       "Solenopsis bicolor (Emery, 1906)"         "Euc
#>                             Eucopricus sp.1
#>                           "Eucopricus sp.1"
```

```
fixAuthors(names)     # splits taxon and author names, if present
```

```
#>                                  orig.name                tax.name                      tax.au
#> 1            Apis mellifera Linnaeus, 1758           Apis mellifera                  Linnaeus,
#> 2                           Apis melifera            Apis melifera
#> 3                      Apis cf. mellifera                 Apis cf.                        Melli
#> 4                     Ancyloscelis armatus     Ancyloscelis armatus
#> 5                           Centris aenea            Centris aenea
#> 6                            Centris rufa             Centris rufa
#> 7  Centris Rhodoprocta Moure & Seabra, 1960                  Centris Rhodoprocta Moure & Seabra,
#> 8                       Lutjanus purpureus       Lutjanus purpureus
#> 9                Parotocinclus amazonensis Parotocinclus amazonensis
#> 10                          Panthera onca            Panthera onca
#> 11         Solenopsis bicolor (Emery, 1906)       Solenopsis bicolor                   (Emery, 1
#> 12         Eucopricus columbi MacLeay, 1819       Eucopricus columbi                   MacLeay,
#> 13                          Eucopricus sp.1          Eucopricus sp.1
```

---

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    An
#> 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 Parot
#> 10                             Panthera onca                      possibly_ok         Felidae
#> 11         Solenopsis bicolor (Emery, 1906)                   name_w_authors      Formicidae
#> 12         Eucopricus columbi MacLeay, 1819                   name_w_authors    Scarabaeidae     S
#> 13                            Eucopricus sp.1                           indet    Scarabaeidae
#>     suggestedAuthorship            tax.notes                       scientificNameFull
#> 1         Linnaeus, 1758         name accepted             Apis mellifera Linnaeus, 1758
#> 2         Linnaeus, 1758       name misspelled             Apis mellifera Linnaeus, 1758
#> 3         Linnaeus, 1758         name accepted             Apis mellifera Linnaeus, 1758
#> 4      (Fabricius, 1793)      replaced synonym  Ancyloscelis apiformis (Fabricius, 1793)
#> 5       Lepeletier, 1841         name accepted            Centris aenea Lepeletier, 1841
#> 6       Lepeletier, 1841      replaced synonym            Centris aenea Lepeletier, 1841
#> 7  Moure & Seabra, 1960         name accepted  Centris rhodoprocta Moure & Seabra, 1960
#> 8           (Poey, 1860)      replaced synonym         Lutjanus campechanus (Poey, 1860)
#> 9        Garavello, 1977 synonym not replaced Parotocinclus amazonensis Garavello, 1977
#> 10      (Linnaeus, 1758)         name accepted            Panthera onca (Linnaeus, 1758)
#> 11         (Emery, 1906)         name accepted          Solenopsis bicolor (Emery, 1906)
#> 12       (MacLeay, 1819)      replaced synonym     Sulcophanaeus columbi (MacLeay, 1819)
#> 13    d'Olsoufieff, 1924      replaced synonym          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 armatus",
  "Centris aenea",
  "Centris rufa",
  "Centris Rhodoprocta Moure & Seabra, 1960",
  "Lutjanus purpureus",
```

```
  "Parotocinclus amazonensis",
  "Panthera onca",
  "Coelonertus baridioides Solari & Solari, 1906",
  "Solenopsis bicolor (Emery, 1906)",
  "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  3  4  5 NA  6  7  8  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 fuction does not change the input family names.

```
names_valid <- prepFamily(names_valid,
                          fam.name = "suggestedFamily",
                          spp.name = "scientificName.new",
                          kingdom = "animalia")
#> Warning: Synonyms for the input kingdom are currently not available. Returning the input family na
```

6 Brief code summary

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

```
# 1) Standardize
names_fixed <- fixSpecies(names)

# 2) Validate agaisnt 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)
#> The following family names were automatically replaced:
#>
#> |Genus          |Old fam.     |New fam.         |
#> |:-------------|:-----------|:---------------|
#> |Ancyloscelis  |Apidae       |Araceae          |
#> |Apis           |Apidae       |Fabaceae         |
#> |Centris        |Apidae       |Melastomataceae  |
#> |Lutjanus       |Lutjanidae   |Fabaceae         |
#> |Panthera       |Felidae      |Fabaceae         |
#> |Parotocinclus |Loricariidae |Cactaceae        |
#> |Solenopsis     |Formicidae   |Campanulaceae    |
#> |Sulcophanaeus |Scarabaeidae |Elaeocarpaceae   |
```

---

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