ParsGBIF Manual

Pablo Hendrigo Alves de Melo[[1]](#footnote-20)

Nadia Bystriakova[[2]](#footnote-22)

Alexandre Monro[[3]](#footnote-24)

2023-06-16

# ParsGBIF Manual

ParsGBIF package is designed to convert [Global Biodiversity Information Facility - GBIF](https://www.gbif.org/) species occurrence data to a more comprehensible format to be used for further analysis, e.g. spatial. The package provides tools for verifying and standardizing species scientific names and for selecting the most informative species records when duplicates are available. The Manual provides a brief introduction to ParsGBIF, with more information available from Help pages accessed via help(function\_name).

## Installation

You can install the development version of ParsGBIF from [GitHub](https://github.com/).

To install ParsGBIF, run

devtools::install\_github("pablopains/ParsGBIF")

Please site ParsGBIF as:

print(citation("ParsGBIF"), bibtex = FALSE)

## To cite package 'ParsGBIF' in publications use:  
##   
## Melo P, Bystriakova N, Monro A (2023). \_ParsGBIF: An R package for  
## parsing species occurrence records\_. R package version 0.0.1.

## Example

**ParsGBIF makes it easy to get species occurrence records based on GBIF.**

### 1. GBIF data preparation

#### 1.1. Obtaining occurrence data of the herbarium specimen from GBIF

1.1.1. Access a registered account in [GBIF](gbif.org)

1.1.2. Filter occurrences with the following parameters:

* Basis of record: *Preserved specimen*
* Occurrence status: *present*
* Scientific name: *Botanical family name* or **filter by other fields**

1.1.3. Request to download information in **DARWIN CORE ARCHIVE FORMAT**

1.1.4. Download compressed file and unzip downloaded file

1.1.5. Use the **occurrence.txt** file as input to the prepare\_gbif\_occurrence\_data(gbif\_occurrece\_file = ‘occurrence.txt’) function

#### 1.2. Preparing occurrence data downloaded from GBIF

To prepare occurrence data downloaded from GBIF to be used by ParsGBIF functions, run prepere\_gbif\_occurrence\_data.

library(ParsGBIF)  
  
 help(prepare\_gbif\_occurrence\_data)  
  
 occ\_file <- 'https://raw.githubusercontent.com/pablopains/ParsGBIF/main/dataGBIF/Achatocarpaceae/occurrence.txt'  
  
 occ <- prepare\_gbif\_occurrence\_data(gbif\_occurrece\_file = occ\_file,  
 columns = 'standard')  
  
 colnames(occ)

## [1] "Ctrl\_bibliographicCitation" "Ctrl\_language"   
## [3] "Ctrl\_institutionCode" "Ctrl\_collectionCode"   
## [5] "Ctrl\_datasetName" "Ctrl\_basisOfRecord"   
## [7] "Ctrl\_informationWithheld" "Ctrl\_dataGeneralizations"   
## [9] "Ctrl\_occurrenceID" "Ctrl\_catalogNumber"   
## [11] "Ctrl\_recordNumber" "Ctrl\_recordedBy"   
## [13] "Ctrl\_georeferenceVerificationStatus" "Ctrl\_occurrenceStatus"   
## [15] "Ctrl\_eventDate" "Ctrl\_year"   
## [17] "Ctrl\_month" "Ctrl\_day"   
## [19] "Ctrl\_habitat" "Ctrl\_fieldNotes"   
## [21] "Ctrl\_eventRemarks" "Ctrl\_locationID"   
## [23] "Ctrl\_higherGeography" "Ctrl\_islandGroup"   
## [25] "Ctrl\_island" "Ctrl\_countryCode"   
## [27] "Ctrl\_stateProvince" "Ctrl\_county"   
## [29] "Ctrl\_municipality" "Ctrl\_locality"   
## [31] "Ctrl\_verbatimLocality" "Ctrl\_locationRemarks"   
## [33] "Ctrl\_decimalLatitude" "Ctrl\_decimalLongitude"   
## [35] "Ctrl\_verbatimCoordinateSystem" "Ctrl\_verbatimIdentification"   
## [37] "Ctrl\_identificationQualifier" "Ctrl\_typeStatus"   
## [39] "Ctrl\_identifiedBy" "Ctrl\_dateIdentified"   
## [41] "Ctrl\_scientificName" "Ctrl\_family"   
## [43] "Ctrl\_taxonRank" "Ctrl\_nomenclaturalCode"   
## [45] "Ctrl\_taxonomicStatus" "Ctrl\_issue"   
## [47] "Ctrl\_mediaType" "Ctrl\_hasCoordinate"   
## [49] "Ctrl\_hasGeospatialIssues" "Ctrl\_verbatimScientificName"   
## [51] "Ctrl\_level0Name" "Ctrl\_level1Name"   
## [53] "Ctrl\_level2Name" "Ctrl\_level3Name"

head(occ)

## # A tibble: 6 × 54  
## Ctrl\_bibliographicCit…¹ Ctrl\_language Ctrl\_institutionCode Ctrl\_collectionCode  
## <chr> <chr> <chr> <chr>   
## 1 <NA> es Universidad de Anti… HUA   
## 2 <NA> es Universidad de Anti… HUA   
## 3 <NA> es Universidad de Anti… HUA   
## 4 <NA> es Universidad de Anti… HUA   
## 5 <NA> es Universidad de Anti… HUA   
## 6 <NA> es Universidad de Anti… HUA   
## # ℹ abbreviated name: ¹​Ctrl\_bibliographicCitation  
## # ℹ 50 more variables: Ctrl\_datasetName <chr>, Ctrl\_basisOfRecord <chr>,  
## # Ctrl\_informationWithheld <chr>, Ctrl\_dataGeneralizations <chr>,  
## # Ctrl\_occurrenceID <chr>, Ctrl\_catalogNumber <chr>, Ctrl\_recordNumber <chr>,  
## # Ctrl\_recordedBy <chr>, Ctrl\_georeferenceVerificationStatus <chr>,  
## # Ctrl\_occurrenceStatus <chr>, Ctrl\_eventDate <dttm>, Ctrl\_year <dbl>,  
## # Ctrl\_month <dbl>, Ctrl\_day <dbl>, Ctrl\_habitat <chr>, …

When parsing data, the user can choose between “standard” and “all” columns to be selected. The “standard” format has 54 data fields (columns), and the “all” format, 257 data fields (columns).

library(ParsGBIF)  
  
 help(select\_gbif\_fields)  
   
 col\_standard <- select\_gbif\_fields(columns = 'standard')  
   
 str(col\_standard)

## chr [1:54] "bibliographicCitation" "language" "institutionCode" ...

col\_all <- select\_gbif\_fields(columns = 'all')  
  
 str(col\_all)

## chr [1:257] "gbifID" "abstract" "accessRights" "accrualMethod" ...

### 2. Extracting GBIF issue to rank the quality of geographic coordinates

library(ParsGBIF)  
  
 occ\_file <- 'https://raw.githubusercontent.com/pablopains/ParsGBIF/main/dataGBIF/Achatocarpaceae/occurrence.txt'  
  
 occ <- prepare\_gbif\_occurrence\_data(gbif\_occurrece\_file = occ\_file,  
 columns = 'standard')  
   
   
 occ\_gbif\_issue <- extract\_gbif\_issue(occ = occ)  
  
 names(occ\_gbif\_issue)

## [1] "occ\_gbif\_issue" "summary"

head(occ\_gbif\_issue$summary)

## issue n\_occ  
## 1 INSTITUTION\_MATCH\_FUZZY 1919  
## 2 GEODETIC\_DATUM\_ASSUMED\_WGS84 1883  
## 3 OCCURRENCE\_STATUS\_INFERRED\_FROM\_INDIVIDUAL\_COUNT 1336  
## 4 CONTINENT\_DERIVED\_FROM\_COORDINATES 1038  
## 5 COORDINATE\_ROUNDED 978  
## 6 TYPE\_STATUS\_INVALID 858

colnames(occ\_gbif\_issue$occ\_gbif\_issue)

## [1] "COORDINATE\_UNCERTAINTY\_METERS\_INVALID"   
## [2] "CONTINENT\_COORDINATE\_MISMATCH"   
## [3] "CONTINENT\_COUNTRY\_MISMATCH"   
## [4] "CONTINENT\_DERIVED\_FROM\_COORDINATES"   
## [5] "CONTINENT\_DERIVED\_FROM\_COUNTRY"   
## [6] "COORDINATE\_ACCURACY\_INVALID"   
## [7] "COORDINATE\_PRECISION\_INVALID"   
## [8] "COORDINATE\_PRECISION\_UNCERTAINTY\_MISMATCH"   
## [9] "COORDINATE\_REPROJECTED"   
## [10] "ELEVATION\_NON\_NUMERIC"   
## [11] "ELEVATION\_NOT\_METRIC"   
## [12] "ELEVATION\_UNLIKELY"   
## [13] "CONTINENT\_INVALID"   
## [14] "COUNTRY\_DERIVED\_FROM\_COORDINATES"   
## [15] "COUNTRY\_INVALID"   
## [16] "ELEVATION\_MIN\_MAX\_SWAPPED"   
## [17] "COORDINATE\_ROUNDED"   
## [18] "DEPTH\_MIN\_MAX\_SWAPPED"   
## [19] "DEPTH\_NON\_NUMERIC"   
## [20] "DEPTH\_NOT\_METRIC"   
## [21] "DEPTH\_UNLIKELY"   
## [22] "COUNTRY\_MISMATCH"   
## [23] "COORDINATE\_REPROJECTION\_FAILED"   
## [24] "COORDINATE\_REPROJECTION\_SUSPICIOUS"   
## [25] "GEODETIC\_DATUM\_INVALID"   
## [26] "PRESUMED\_NEGATED\_LATITUDE"   
## [27] "PRESUMED\_NEGATED\_LONGITUDE"   
## [28] "PRESUMED\_SWAPPED\_COORDINATE"   
## [29] "GEODETIC\_DATUM\_ASSUMED\_WGS84"   
## [30] "COORDINATE\_INVALID"   
## [31] "COORDINATE\_OUT\_OF\_RANGE"   
## [32] "COUNTRY\_COORDINATE\_MISMATCH"   
## [33] "ZERO\_COORDINATE"   
## [34] "AMBIGUOUS\_COLLECTION"   
## [35] "AMBIGUOUS\_INSTITUTION"   
## [36] "BASIS\_OF\_RECORD\_INVALID"   
## [37] "COLLECTION\_MATCH\_FUZZY"   
## [38] "COLLECTION\_MATCH\_NONE"   
## [39] "DIFFERENT\_OWNER\_INSTITUTION"   
## [40] "FOOTPRINT\_SRS\_INVALID"   
## [41] "FOOTPRINT\_WKT\_INVALID"   
## [42] "FOOTPRINT\_WKT\_MISMATCH"   
## [43] "GEOREFERENCED\_DATE\_INVALID"   
## [44] "GEOREFERENCED\_DATE\_UNLIKELY"   
## [45] "IDENTIFIED\_DATE\_INVALID"   
## [46] "IDENTIFIED\_DATE\_UNLIKELY"   
## [47] "INDIVIDUAL\_COUNT\_CONFLICTS\_WITH\_OCCURRENCE\_STATUS"  
## [48] "INDIVIDUAL\_COUNT\_INVALID"   
## [49] "INSTITUTION\_COLLECTION\_MISMATCH"   
## [50] "INSTITUTION\_MATCH\_FUZZY"   
## [51] "INSTITUTION\_MATCH\_NONE"   
## [52] "INTERPRETATION\_ERROR"   
## [53] "MODIFIED\_DATE\_INVALID"   
## [54] "MODIFIED\_DATE\_UNLIKELY"   
## [55] "MULTIMEDIA\_DATE\_INVALID"   
## [56] "MULTIMEDIA\_URI\_INVALID"   
## [57] "OCCURRENCE\_STATUS\_INFERRED\_FROM\_BASIS\_OF\_RECORD"   
## [58] "OCCURRENCE\_STATUS\_INFERRED\_FROM\_INDIVIDUAL\_COUNT"   
## [59] "OCCURRENCE\_STATUS\_UNPARSABLE"   
## [60] "POSSIBLY\_ON\_LOAN"   
## [61] "RECORDED\_DATE\_INVALID"   
## [62] "RECORDED\_DATE\_MISMATCH"   
## [63] "RECORDED\_DATE\_UNLIKELY"   
## [64] "REFERENCES\_URI\_INVALID"   
## [65] "TAXON\_MATCH\_AGGREGATE"   
## [66] "TAXON\_MATCH\_FUZZY"   
## [67] "TAXON\_MATCH\_HIGHERRANK"   
## [68] "TAXON\_MATCH\_NONE"   
## [69] "TYPE\_STATUS\_INVALID"

head(occ\_gbif\_issue$occ\_gbif\_issue)

## COORDINATE\_UNCERTAINTY\_METERS\_INVALID CONTINENT\_COORDINATE\_MISMATCH  
## 1 FALSE FALSE  
## 2 FALSE FALSE  
## 3 FALSE FALSE  
## 4 FALSE FALSE  
## 5 FALSE FALSE  
## 6 FALSE FALSE  
## CONTINENT\_COUNTRY\_MISMATCH CONTINENT\_DERIVED\_FROM\_COORDINATES  
## 1 FALSE FALSE  
## 2 FALSE FALSE  
## 3 FALSE FALSE  
## 4 FALSE FALSE  
## 5 FALSE FALSE  
## 6 FALSE FALSE  
## CONTINENT\_DERIVED\_FROM\_COUNTRY COORDINATE\_ACCURACY\_INVALID  
## 1 FALSE FALSE  
## 2 FALSE FALSE  
## 3 FALSE FALSE  
## 4 FALSE FALSE  
## 5 FALSE FALSE  
## 6 FALSE FALSE  
## COORDINATE\_PRECISION\_INVALID COORDINATE\_PRECISION\_UNCERTAINTY\_MISMATCH  
## 1 FALSE FALSE  
## 2 FALSE FALSE  
## 3 FALSE FALSE  
## 4 FALSE FALSE  
## 5 FALSE FALSE  
## 6 FALSE FALSE  
## COORDINATE\_REPROJECTED ELEVATION\_NON\_NUMERIC ELEVATION\_NOT\_METRIC  
## 1 FALSE FALSE FALSE  
## 2 FALSE FALSE FALSE  
## 3 FALSE FALSE FALSE  
## 4 FALSE FALSE FALSE  
## 5 FALSE FALSE FALSE  
## 6 FALSE FALSE FALSE  
## ELEVATION\_UNLIKELY CONTINENT\_INVALID COUNTRY\_DERIVED\_FROM\_COORDINATES  
## 1 FALSE FALSE FALSE  
## 2 FALSE FALSE FALSE  
## 3 FALSE FALSE FALSE  
## 4 FALSE FALSE FALSE  
## 5 FALSE FALSE FALSE  
## 6 FALSE FALSE FALSE  
## COUNTRY\_INVALID ELEVATION\_MIN\_MAX\_SWAPPED COORDINATE\_ROUNDED  
## 1 FALSE FALSE FALSE  
## 2 FALSE FALSE TRUE  
## 3 FALSE FALSE TRUE  
## 4 FALSE FALSE TRUE  
## 5 FALSE FALSE TRUE  
## 6 FALSE FALSE FALSE  
## DEPTH\_MIN\_MAX\_SWAPPED DEPTH\_NON\_NUMERIC DEPTH\_NOT\_METRIC DEPTH\_UNLIKELY  
## 1 FALSE FALSE FALSE FALSE  
## 2 FALSE FALSE FALSE FALSE  
## 3 FALSE FALSE FALSE FALSE  
## 4 FALSE FALSE FALSE FALSE  
## 5 FALSE FALSE FALSE FALSE  
## 6 FALSE FALSE FALSE FALSE  
## COUNTRY\_MISMATCH COORDINATE\_REPROJECTION\_FAILED  
## 1 FALSE FALSE  
## 2 FALSE FALSE  
## 3 FALSE FALSE  
## 4 FALSE FALSE  
## 5 FALSE FALSE  
## 6 FALSE FALSE  
## COORDINATE\_REPROJECTION\_SUSPICIOUS GEODETIC\_DATUM\_INVALID  
## 1 FALSE FALSE  
## 2 FALSE FALSE  
## 3 FALSE FALSE  
## 4 FALSE FALSE  
## 5 FALSE FALSE  
## 6 FALSE FALSE  
## PRESUMED\_NEGATED\_LATITUDE PRESUMED\_NEGATED\_LONGITUDE  
## 1 FALSE FALSE  
## 2 FALSE FALSE  
## 3 FALSE FALSE  
## 4 FALSE FALSE  
## 5 FALSE FALSE  
## 6 FALSE FALSE  
## PRESUMED\_SWAPPED\_COORDINATE GEODETIC\_DATUM\_ASSUMED\_WGS84 COORDINATE\_INVALID  
## 1 FALSE FALSE FALSE  
## 2 FALSE FALSE FALSE  
## 3 FALSE FALSE FALSE  
## 4 FALSE FALSE FALSE  
## 5 FALSE FALSE FALSE  
## 6 FALSE FALSE FALSE  
## COORDINATE\_OUT\_OF\_RANGE COUNTRY\_COORDINATE\_MISMATCH ZERO\_COORDINATE  
## 1 FALSE FALSE FALSE  
## 2 FALSE FALSE FALSE  
## 3 FALSE FALSE FALSE  
## 4 FALSE FALSE FALSE  
## 5 FALSE FALSE FALSE  
## 6 FALSE FALSE FALSE  
## AMBIGUOUS\_COLLECTION AMBIGUOUS\_INSTITUTION BASIS\_OF\_RECORD\_INVALID  
## 1 FALSE FALSE FALSE  
## 2 FALSE FALSE FALSE  
## 3 FALSE FALSE FALSE  
## 4 FALSE FALSE FALSE  
## 5 FALSE FALSE FALSE  
## 6 FALSE FALSE FALSE  
## COLLECTION\_MATCH\_FUZZY COLLECTION\_MATCH\_NONE DIFFERENT\_OWNER\_INSTITUTION  
## 1 TRUE FALSE FALSE  
## 2 TRUE FALSE FALSE  
## 3 TRUE FALSE FALSE  
## 4 TRUE FALSE FALSE  
## 5 TRUE FALSE FALSE  
## 6 TRUE FALSE FALSE  
## FOOTPRINT\_SRS\_INVALID FOOTPRINT\_WKT\_INVALID FOOTPRINT\_WKT\_MISMATCH  
## 1 FALSE FALSE FALSE  
## 2 FALSE FALSE FALSE  
## 3 FALSE FALSE FALSE  
## 4 FALSE FALSE FALSE  
## 5 FALSE FALSE FALSE  
## 6 FALSE FALSE FALSE  
## GEOREFERENCED\_DATE\_INVALID GEOREFERENCED\_DATE\_UNLIKELY  
## 1 FALSE FALSE  
## 2 FALSE FALSE  
## 3 FALSE FALSE  
## 4 FALSE FALSE  
## 5 FALSE FALSE  
## 6 FALSE FALSE  
## IDENTIFIED\_DATE\_INVALID IDENTIFIED\_DATE\_UNLIKELY  
## 1 FALSE FALSE  
## 2 FALSE FALSE  
## 3 FALSE FALSE  
## 4 FALSE FALSE  
## 5 FALSE FALSE  
## 6 FALSE FALSE  
## INDIVIDUAL\_COUNT\_CONFLICTS\_WITH\_OCCURRENCE\_STATUS INDIVIDUAL\_COUNT\_INVALID  
## 1 FALSE FALSE  
## 2 FALSE FALSE  
## 3 FALSE FALSE  
## 4 FALSE FALSE  
## 5 FALSE FALSE  
## 6 FALSE FALSE  
## INSTITUTION\_COLLECTION\_MISMATCH INSTITUTION\_MATCH\_FUZZY  
## 1 FALSE TRUE  
## 2 FALSE TRUE  
## 3 FALSE TRUE  
## 4 FALSE TRUE  
## 5 FALSE TRUE  
## 6 FALSE TRUE  
## INSTITUTION\_MATCH\_NONE INTERPRETATION\_ERROR MODIFIED\_DATE\_INVALID  
## 1 FALSE FALSE FALSE  
## 2 FALSE FALSE FALSE  
## 3 FALSE FALSE FALSE  
## 4 FALSE FALSE FALSE  
## 5 FALSE FALSE FALSE  
## 6 FALSE FALSE FALSE  
## MODIFIED\_DATE\_UNLIKELY MULTIMEDIA\_DATE\_INVALID MULTIMEDIA\_URI\_INVALID  
## 1 FALSE FALSE FALSE  
## 2 FALSE FALSE FALSE  
## 3 FALSE FALSE FALSE  
## 4 FALSE FALSE FALSE  
## 5 FALSE FALSE FALSE  
## 6 FALSE FALSE FALSE  
## OCCURRENCE\_STATUS\_INFERRED\_FROM\_BASIS\_OF\_RECORD  
## 1 FALSE  
## 2 FALSE  
## 3 FALSE  
## 4 FALSE  
## 5 FALSE  
## 6 FALSE  
## OCCURRENCE\_STATUS\_INFERRED\_FROM\_INDIVIDUAL\_COUNT OCCURRENCE\_STATUS\_UNPARSABLE  
## 1 FALSE FALSE  
## 2 FALSE FALSE  
## 3 FALSE FALSE  
## 4 FALSE FALSE  
## 5 FALSE FALSE  
## 6 FALSE FALSE  
## POSSIBLY\_ON\_LOAN RECORDED\_DATE\_INVALID RECORDED\_DATE\_MISMATCH  
## 1 FALSE FALSE FALSE  
## 2 FALSE FALSE FALSE  
## 3 FALSE FALSE FALSE  
## 4 FALSE FALSE FALSE  
## 5 FALSE FALSE FALSE  
## 6 FALSE FALSE FALSE  
## RECORDED\_DATE\_UNLIKELY REFERENCES\_URI\_INVALID TAXON\_MATCH\_AGGREGATE  
## 1 FALSE FALSE FALSE  
## 2 FALSE FALSE FALSE  
## 3 FALSE FALSE FALSE  
## 4 FALSE FALSE FALSE  
## 5 FALSE FALSE FALSE  
## 6 FALSE FALSE FALSE  
## TAXON\_MATCH\_FUZZY TAXON\_MATCH\_HIGHERRANK TAXON\_MATCH\_NONE TYPE\_STATUS\_INVALID  
## 1 FALSE FALSE FALSE FALSE  
## 2 FALSE FALSE FALSE FALSE  
## 3 FALSE FALSE FALSE FALSE  
## 4 FALSE FALSE FALSE FALSE  
## 5 FALSE FALSE FALSE FALSE  
## 6 FALSE FALSE FALSE FALSE

#### 2.1. GBIF Issue is a database…

### 3. Check species names against WCVP database

library(ParsGBIF)  
  
help(batch\_checkName\_wcvp)  
  
  
# wcvp\_names <- get\_wcvp(read\_only\_to\_memory = TRUE)$wcvp\_names  
  
# wcvp\_names\_Achatocarpaceae is a subset of data from the wcvp names database used only to demonstrate the use of the ParsGBIF package. See below alternatives of loading wcvp name database.  
  
data(wcvp\_names\_Achatocarpaceae)  
wcvp\_names <- wcvp\_names\_Achatocarpaceae  
  
occ\_file <- 'https://raw.githubusercontent.com/pablopains/ParsGBIF/main/dataGBIF/Achatocarpaceae/occurrence.txt'  
  
occ <- prepare\_gbif\_occurrence\_data(gbif\_occurrece\_file = occ\_file,  
 columns = 'standard')  
  
res\_batch\_checkName\_wcvp <- batch\_checkName\_wcvp(occ = occ,  
 wcvp\_names = wcvp\_names,  
 if\_author\_fails\_try\_without\_combinations = TRUE,  
 wcvp\_selected\_fields = 'standard')

## [1] "1-22 Achatocarpus nigricans Triana"  
## [1] "2-22 Achatocarpus obovatus Schinz & Autran"  
## [1] "3-22 Achatocarpus balansae Schinz & Autran"  
## [1] "4-22 Achatocarpus pubescens C.H.Wright"  
## [1] "5-22 Achatocarpus gracilis H.Walter"  
## [1] "6-22 Achatocarpus brevipedicellatus H.Walter"  
## [1] "7-22 Achatocarpus bicornutus Schinz & Autran"  
## [1] "8-22 Phaulothamnus spinescens A.Gray"  
## [1] "9-22 Achatocarpus microcarpus Schinz & Autran"  
## [1] "10-22 Achatocarpus spinulosus Griseb."  
## [1] "11-22 Achatocarpus praecox Griseb."  
## [1] "12-22 Achatocarpus praecox var. bicornutus (Schinz & Autran) Botta"  
## [1] "13-22 Achatocarpus praecox f. praecox"  
## [1] "14-22 Achatocarpus praecox f. obovatus (Schinz & Autran) Hauman"  
## [1] "15-22 Achatocarpus mexicanus H.Walter"  
## [1] "16-22 Achatocarpus hasslerianus Heimerl"  
## [1] "17-22 Achatocarpus praecox f. spinulosus (Schinz & Autran) Hauman"  
## [1] "18-22 Achatocarpus oaxacanus Standl."  
## [1] "19-22 Achatocarpus mollis H.Walter"  
## [1] "20-22 Achatocarpus brasiliensis H.Walter"  
## [1] "21-22 Achatocarpus praecox var. praecox"  
## [1] "22-22 Ampelocera hondurensis Donn.Sm."

names(res\_batch\_checkName\_wcvp)

## [1] "occ\_checkName\_wcvp" "summary"

head(res\_batch\_checkName\_wcvp$summary)

## wcvp\_plant\_name\_id wcvp\_taxon\_rank wcvp\_taxon\_status wcvp\_family  
## 4 500156 Species Accepted Achatocarpaceae  
## 22 500161 Form Accepted Achatocarpaceae  
## 15 500146 Species Accepted Achatocarpaceae  
## 9 500163 Species Accepted Achatocarpaceae  
## 2 500150 Species Accepted Achatocarpaceae  
## 16 500149 Species Accepted Achatocarpaceae  
## wcvp\_taxon\_name wcvp\_taxon\_authors  
## 4 Achatocarpus nigricans Triana  
## 22 Achatocarpus praecox f. obovatus (Schinz & Autran) Hauman  
## 15 Achatocarpus balansae Schinz & Autran  
## 9 Achatocarpus pubescens C.H.Wright  
## 2 Achatocarpus gracilis H.Walter  
## 16 Achatocarpus brevipedicellatus H.Walter  
## wcvp\_accepted\_plant\_name\_id wcvp\_reviewed  
## 4 500156 Y  
## 22 500161 Y  
## 15 500146 Y  
## 9 500163 Y  
## 2 500150 Y  
## 16 500149 Y  
## wcvp\_searchedName wcvp\_taxon\_status\_of\_searchedName  
## 4 Achatocarpus nigricans Triana <NA>  
## 22 Achatocarpus obovatus Schinz & Autran Synonym  
## 15 Achatocarpus balansae Schinz & Autran <NA>  
## 9 Achatocarpus pubescens C.H.Wright <NA>  
## 2 Achatocarpus gracilis H.Walter <NA>  
## 16 Achatocarpus brevipedicellatus H.Walter <NA>  
## wcvp\_plant\_name\_id\_of\_searchedName wcvp\_taxon\_authors\_of\_searchedName  
## 4 NA <NA>  
## 22 500158 Schinz & Autran  
## 15 NA <NA>  
## 9 NA <NA>  
## 2 NA <NA>  
## 16 NA <NA>  
## wcvp\_verified\_author wcvp\_verified\_speciesName wcvp\_searchNotes  
## 4 100 100 Accepted  
## 22 100 100 Updated  
## 15 100 100 Accepted  
## 9 100 100 Accepted  
## 2 100 100 Accepted  
## 16 100 100 Accepted

head(res\_batch\_checkName\_wcvp$occ\_checkName\_wcvp)

## wcvp\_plant\_name\_id wcvp\_taxon\_rank wcvp\_taxon\_status wcvp\_family  
## 1 500156 Species Accepted Achatocarpaceae  
## 2 500156 Species Accepted Achatocarpaceae  
## 3 500156 Species Accepted Achatocarpaceae  
## 4 500156 Species Accepted Achatocarpaceae  
## 5 500156 Species Accepted Achatocarpaceae  
## 6 500156 Species Accepted Achatocarpaceae  
## wcvp\_taxon\_name wcvp\_taxon\_authors wcvp\_accepted\_plant\_name\_id  
## 1 Achatocarpus nigricans Triana 500156  
## 2 Achatocarpus nigricans Triana 500156  
## 3 Achatocarpus nigricans Triana 500156  
## 4 Achatocarpus nigricans Triana 500156  
## 5 Achatocarpus nigricans Triana 500156  
## 6 Achatocarpus nigricans Triana 500156  
## wcvp\_reviewed wcvp\_searchedName wcvp\_taxon\_status\_of\_searchedName  
## 1 Y Achatocarpus nigricans Triana <NA>  
## 2 Y Achatocarpus nigricans Triana <NA>  
## 3 Y Achatocarpus nigricans Triana <NA>  
## 4 Y Achatocarpus nigricans Triana <NA>  
## 5 Y Achatocarpus nigricans Triana <NA>  
## 6 Y Achatocarpus nigricans Triana <NA>  
## wcvp\_plant\_name\_id\_of\_searchedName wcvp\_taxon\_authors\_of\_searchedName  
## 1 NA <NA>  
## 2 NA <NA>  
## 3 NA <NA>  
## 4 NA <NA>  
## 5 NA <NA>  
## 6 NA <NA>  
## wcvp\_verified\_author wcvp\_verified\_speciesName wcvp\_searchNotes  
## 1 100 100 Accepted  
## 2 100 100 Accepted  
## 3 100 100 Accepted  
## 4 100 100 Accepted  
## 5 100 100 Accepted  
## 6 100 100 Accepted

There are two ways to load WCVP name database:

* ParsGBIF::get\_wcvp() function
* rWCVPdata package

#### 3.1. Getting WCVP database from ParsGBIF::get\_wcvp() function

There is a potion to save the local copy, indicated to optimize future loads. Or always read the latest database version from which always requires a download.

help(get\_wcvp)  
  
path\_data <- tempdir() # you can change this folder  
  
wcvp <- get\_wcvp(url\_source = 'http://sftp.kew.org/pub/data-repositories/WCVP/',  
 read\_only\_to\_memory = FALSE,  
 path\_results = path\_data,  
 update = FALSE,  
 load\_distribution = TRUE)  
  
names(wcvp)  
  
head(wcvp$wcvp\_names)  
  
colnames(wcvp$wcvp\_names)  
  
head(wcvp$wcvp\_distribution)  
  
colnames(wcvp$wcvp\_distribution)

#### 3.2. Getting WCVP database from rWCVPdata package

Use the database version available in the package. After installing the package, loading the database is very fast. Small adjustments are needed for the data to be used by the ParsGBIF functions.

# devtools::install\_github(“matildabrown/rWCVPdata”) library(rWCVPdata)  
  
library(rWCVPdata)  
  
wcvp\_names <- rWCVPdata::wcvp\_names %\>%  
  
# Small adjustments  
dplyr::mutate(TAXON\_NAME\_U = toupper(taxon\_name),   
 TAXON\_AUTHORS\_U = toupper(taxon\_authors))

1. Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais, [pablopains@yahoo.com.br](mailto:pablopains@yahoo.com.br) [↑](#footnote-ref-20)
2. Natural History Museum, London, [n\_bystriakova@yahoo.com](mailto:n_bystriakova@yahoo.com) [↑](#footnote-ref-22)
3. Royal Botanic Gardens, Kew, [a.monro@kew.org](mailto:a.monro@kew.org) [↑](#footnote-ref-24)