Package 'ppendemic'

May 9, 2026

Title A Glimpse at the Diversity of Peru's Endemic Plants

Version 0.2.1

Description Introducing a novel and updated database showcasing Peru's endemic plants. This meticulously compiled and revised botanical collection encompasses a remarkable assemblage of over 7,898 distinct species. The data for this resource was sourced from the work of Govaerts, R., Nic Lughadha, E., Black, N. et al., titled 'The World Checklist of Vascular Plants: A continuously updated resource for exploring global plant diversity', published in Sci Data 8, 215 (2021) <doi:10.1038/s41597-021-00997-6>.

License MIT + file LICENSE

URL https://github.com/PaulESantos/ppendemic/

BugReports https://github.com/PaulESantos/ppendemic/issues/

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

Encoding UTF-8

LazyData true

RoxygenNote 7.3.3

Depends R (>= 4.1.0),

Config/testthat/edition 3

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

Imports assertthat, cli, dplyr, fuzzyjoin, memoise, progress, purrr, readr, stringr, tibble, tidyr

NeedsCompilation no

Author Paul E. Santos Andrade [aut, cre] (ORCID: <https://orcid.org/0000-0002-6635-0375>), Lucely L. Vilca Bustamante [aut] (ORCID: <https://orcid.org/0000-0002-5559-1296>)

Repository CRAN

Date/Publication 2026-02-26 08:40:07 UTC

Contents

`is_ppendemic` *Check if species are endemic in the ppendemic database*

Description

This function checks if a list of species names are endemic in the ppendemic database. The function allows fuzzy matching for species names with a maximum distance threshold to handle potential typos or variations in species names.

Usage

```
is_ppendemic(
  splist,
  max_dist = 2,
  save_ambiguous = FALSE,
  ambiguous_path = "ambiguous_genera.csv"
)
```

Arguments

`splist`	A character vector containing the list of species names to be checked for endemic in the ppendemic database.
`max_dist`	Maximum edit distance used in fuzzy matching steps. Defaults to 2.
`save_ambiguous`	Logical flag. If `TRUE`, ambiguous fuzzy genus matches are exported to disk.
`ambiguous_path`	File path used when `save_ambiguous = TRUE`. Defaults to `"ambiguous_genera.csv"`.

Value

A character vector indicating if each species is endemic or not endemic.

Examples

```
is_ppendemic(c("Aa aurantiaca", "Aa aurantiaaia", "Werneria nubigena"))
```

ppendemic_tab14 *ppendemic_tab14: Endemic Plant Database of Peru*

Description

The ppendemic_tab14 dataset is a tibble (data frame) that provides easy access to a comprehensive database of Peru's endemic plant species. It contains a total of 7,898 records with essential botanical information, including the accepted name, accepted family, genus, species, infraspecific information, taxon authors, primary author, place of publication, volume and page, publication years, and version details.

Usage

```
ppendemic_tab14
```

Format

A tibble (data frame) with 7,898 rows and 18 columns:

taxon_name Character vector. The accepted name of the endemic plant species.

taxon_status Character vector. The taxonomic status of the species (e.g., "Accepted").

family Character vector. The family of the accepted name of the endemic plant species.

genus Character vector. The genus of the endemic plant species.

species Character vector. The specific epithet of the endemic plant species.

infraspecific_rank Character vector. The infraspecific rank (e.g., "subsp.", "var.") when applicable.

infraspecies Character vector. The infraspecific epithet when applicable.

taxon_authors Character vector. The author(s) of the accepted name of the endemic plant species.

primary_author Character vector. The primary author(s) of the publication containing the endemic plant species information.

place_of_publication Character vector. The place of publication of the endemic plant species information.

volume_and_page Character vector. The volume and page number of the publication containing the endemic plant species information.

first_published Character vector. The first published year of the publication containing the endemic plant species information.

year_actual Numeric vector. The actual year of publication extracted from first_published.

year_nominal Numeric vector. The nominal year of publication extracted from first_published.

both_years Character vector. Both actual and nominal years when different, extracted from first_published.

has_different_years Logical vector. Indicates whether the actual and nominal publication years differ (TRUE when both_years contains the pattern "YYYY|YYYY").

version Character vector. The version identifier "V-14" of the ppendemic database.

version_date Character vector. The version date "28-05-2025" indicating when this version was created.

Details

The dataset provides a curated and up-to-date collection of Peru's endemic plant species, gathered from reputable botanical sources and publications. The data for this database was extracted and compiled from the World Checklist of Vascular Plants (WCVP) database, which is a comprehensive and reliable repository of botanical information.

This version (ppendemic_tab14) includes enhanced temporal information with separate numeric fields for actual and nominal publication years. This allows for more precise bibliographic tracking and citation accuracy. The dataset also includes improved infraspecific taxonomy handling with dedicated fields for ranks and epithets.

The year extraction process uses sophisticated pattern matching to distinguish between actual publication years and nominal years, with the has_different_years field automatically flagging records where these differ. This is particularly important for historical botanical publications where publication delays were common.

Source

The dataset has been carefully compiled and updated to offer the latest insights into Peru's endemic plant species. The data is sourced from the World Checklist of Vascular Plants (WCVP) database, an international collaborative programme initiated in 1988 by Rafaël Govaerts that provides high-quality expert-reviewed taxonomic data on all vascular plants.

For detailed methodology, see Govaerts et al. (2021) "The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity" in Nature Scientific Data.

Examples

```
# Load the package
library(ppendemic)

# Access the dataset
data("ppendemic_tab14")

# View the structure of the dataset
str(ppendemic_tab14)

# View first few rows
head(ppendemic_tab14)

# Check for species with different actual and nominal years
different_years <- subset(ppendemic_tab14, has_different_years == TRUE)
nrow(different_years)

# View records with both years information
head(ppendemic_tab14$both_years[ppendemic_tab14$has_different_years])
```

ppendemic_tab15 *ppendemic_tab15: Endemic Plant Database of Peru (based on WCVP v15)*

Description

The ppendemic_tab15 dataset is a tibble (data frame) providing a curated and taxonomically validated list of vascular plant species that occur exclusively in Peru. The dataset is derived from version 15 of the World Checklist of Vascular Plants (WCVP), facilitated by the Royal Botanic Gardens, Kew, and corresponds to the extraction performed on 06 January 2026.

Usage

```
ppendemic_tab15
```

Format

A tibble (data frame) with 7,892 rows and 18 columns:

taxon_name Character vector. The full scientific name of the accepted endemic species (binomial or trinomial), constructed from genus, species and, where applicable, infraspecific epithets, following WCVP standards.

taxon_status Character vector. The taxonomic status of the name according to WCVP (e.g., "Accepted").

family Character vector. The botanical family to which the endemic taxon belongs, following WCVP family circumscription.

Genus Character vector. The genus name of the endemic taxon.

Species Character vector. The specific epithet of the endemic taxon.

infraspecific_rank Character vector. The infraspecific rank (e.g., "subsp.", "var.") when applicable.

infraspecies Character vector. The infraspecific epithet when applicable.

taxon_authors Character vector. The standardized authorship of the taxon name, concatenating parenthetical and primary authors following botanical nomenclature conventions.

primary_author Character vector. The author(s) who validly published the scientific name.

place_of_publication Character vector. The journal, book or other publication in which the taxon name was effectively published.

volume_and_page Character vector. The volume and page reference of the original publication of the taxon name (e.g., "13(2): 99").

first_published Character vector. The year of publication of the name, enclosed in parentheses, as reported by WCVP.

year_actual Numeric vector. The actual year of publication extracted from first_published.

year_nominal Numeric vector. The nominal year of publication extracted from first_published.

both_years Character vector. A concatenation of actual and nominal years when these differ (format "YYYY|YYYY").

has_different_years Logical vector. Indicates whether the actual and nominal publication years differ.

version Character vector. The database version identifier "V-15", corresponding to WCVP version 15.

version_date Character vector. The extraction date of the WCVP source dataset ("06-01-2026").

Details

This database contains only accepted endemic species names, following the taxonomic backbone of WCVP and filtered by geographic distribution records indicating occurrence restricted to Peru.

The dataset was constructed from the World Checklist of Vascular Plants (WCVP) version 15, a continuously updated global taxonomic resource curated by the Royal Botanic Gardens, Kew. Only taxa of rank species and below with accepted status were considered.

Geographic distribution fields from WCVP were used to identify taxa whose native range is restricted exclusively to Peru. Introduced, cultivated, misapplied, synonymic and unplaced names were excluded.

The nomenclature, authorship, family assignment and publication details strictly follow the International Code of Nomenclature for algae, fungi and plants (ICN) and WCVP editorial standards.

Temporal metadata was processed to extract both nominal and actual publication years, allowing detailed historical bibliographic analysis of Peruvian endemic flora.

Source

Data derived from:

Govaerts, R. (ed.). 2026. WCVP: World Checklist of Vascular Plants. Facilitated by the Royal Botanic Gardens, Kew. Version 15. Extracted 06 January 2026. https://doi.org/10.34885/rvc3-4d77

For methodological details:

Govaerts, R., Nic Lughadha, E., et al. (2021). The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. Scientific Data, 8, 215. https://doi.org/10.1038/s41597-021-00997-6

References

Royal Botanic Gardens, Kew (2026). World Checklist of Vascular Plants (WCVP), Version 15. https://wcvp.science.kew.org

Index