

SHORT COMMUNICATION

faunabr: An R package to explore taxonomic data and map species distributions using the Catálogo Taxonômico da Fauna do Brasil

Weverton Carlos Ferreira Trindade¹

¹Laboratório de Ecologia de Plantas, Departamento de Botânica, Universidade Federal do Paraná.
81531-980 Curitiba, PR, Brazil. (wevertonf1993@gmail.com)

<https://zoobank.org/0444C8D7-0C83-4BBA-8D9F-2FFDE3DB7D12>

ABSTRACT. The Catálogo Taxonômico da Fauna do Brasil (CTFB) is a comprehensive digital biodiversity database that provides standardized taxonomic and ecological information for animal species in Brazil. Although widely used in biodiversity studies, its accessibility and reproducibility can be enhanced through integration with computational tools like R packages. Here, we introduce *faunabr*, an R package designed to facilitate the retrieval, exploration, and spatialization of taxonomic data from CTFB. I demonstrate its application in key biogeographic and conservation tasks, including taxonomic verification, selection of target species based on traits and distribution, and identification of erroneous occurrence records outside their known ranges. By simplifying these processes, *faunabr* enhances the accessibility and practical use of CTFB data for zoologists, biogeographers, ecologists, taxonomists, and curators of biological collections, ultimately supporting more rigorous and reproducible biodiversity research.

KEYWORDS. Biodiversity database, Brazilian fauna, species information, taxonomic verification.

The Catálogo Taxonômico da Fauna do Brasil (CTFB, <http://fauna.jbrj.gov.br/>) is a comprehensive digital biodiversity database that provides standardized taxonomical, ecological, and geographical distribution data for both native and non-native animal species in Brazil. Publicly launched in 2015, the project is the result of a collaborative effort by over 800 zoologists, each specializing in different taxonomic groups of Brazilian fauna (Boeger et al. 2024). By February 2025, the CTFB had cataloged 137,571 nominal species, of which 127,203 were considered valid.

Digital biodiversity databases, such as the CTFB, play a fundamental role in advancing biodiversity research and conservation efforts (Peterson et al. 2018). CTFB has supported a broad range of studies in taxonomy and ecology (Boeger et al. 2024). For instance, it has provided data that support the discovery of new animal taxa and biodiversity inventories (Victorino et al. 2023) and served as a primary source of biodiversity data for ecological and biogeographic research (Benetti et al. 2024). Although the CTFB has already been widely used in various studies, the accessibility and

reproducibility of biodiversity databases can be enhanced by integrating their data with R packages (Lai et al. 2019).

R is a programming language and software environment specifically designed for statistical computing, data analysis, and graphical representation. Owing to its robust statistical capabilities, advanced data visualization tools, open-source nature, and overall accessibility and reproducibility, R has become the preferred choice for data analysis in biodiversity and conservation research (Lai et al. 2023). One of R's greatest strengths lies in its packages, which extend its functionality by providing datasets and specialized functions to address specific research questions. For instance, *florabr* (Trindade 2024) is an R package designed to integrate data from Flora e Funga do Brasil (BFG et al. 2021, <https://floradobrasil.jbrj.gov.br/consulta/>), a sister project of CTFB, into R. Inspired by this package, I present here *faunabr*, an R package designed to explore taxonomic information and spatialize species distribution using CTFB.

The *faunabr* package provides a suite of functions designed to retrieve, filter, and spatialize data from the



Figure 1. The four main steps of the workflow proposed here for using *faunabr* to assist with common tasks in biogeography and conservation studies. Each of the four steps is represented in brown boxes, while the corresponding functions are illustrated in off-white boxes.

CTFB (Fig. 1). It includes a proposed workflow that can be applied to common tasks in taxonomy, biogeography, and conservation studies. These tasks include checking and resolving taxonomic issues, selecting target species based on their traits and distribution, and identifying erroneous occurrence records that deviate from the species' natural ranges according to the CTFB. The workflow is available on the project's website (<https://wevertonbio.github.io/faunabr/>) in both English and Portuguese.

The workflow is divided into four vignettes. The first demonstrates how to install the package and load data from CTFB. All data in CTFB are stored as a Darwin Core archive via the Integrated Publishing Toolkit (IPT; https://ipt.jbrj.gov.br/jbrj/resource?r=catalogo_taxonomico_da_fauna_do_brasil), which is updated monthly. The *get_faunabr()* function downloads these datasets, optionally translates the information from Portuguese to English, and merges them into a single file. By default, the function downloads the most recent version available. However, users can check which versions are available and download older ones via the

IPT link mentioned above. To use the other functions of the package, the data must be loaded into the R environment with the *load_faunabr()* function.

Selecting a list of target species is a common task in macroecological and conservation studies, and the second vignette demonstrates how to accomplish this using the package. The *select_fauna()* function allows users to combine various filters related to taxonomy (phylum, class, order, family, genus, and taxonomic status), ecology (life form and habitat), and distribution (Brazilian states and countries with confirmed occurrences). For instance, this vignette shows how to retrieve a list of all native arboreal insects with valid names and confirmed occurrences in Rio de Janeiro, which results in 114 species. In some cases, users may already have a list of species and wish to retrieve additional information about them. In such cases, the package can be used to check the spelling, nomenclatural status, and taxonomic status of species names (*check_fauna_names* function); retrieve species information on taxonomy, ecology, and distribution (*subset_fauna*); and retrieve their synonyms (*fauna_synonym*).

Since the distributions of species in CTFB are based on the expertise of taxonomists, they represent valuable information for validating occurrence records retrieved from online databases, such as Global Biodiversity Information Facility (GBIF; <https://www.gbif.org/>) or SpeciesLink (<https://specieslink.net/>). In the third vignette, I demonstrate how to use the `filter_faunabr()` function to flag and/or remove records that fall outside the Brazilian states or countries with confirmed occurrences according to CTFB. In one of the examples in the package, we retrieved 2,793 occurrence records for *Panthera onca* from GBIF and checked if the records fall within the Brazilian states with confirmed occurrences (Fig. 2). To highlight the example, I added four fake records from countries and states without confirmed occurrences. According to CTFB, *Panthera onca* is a terrestrial felid with confirmed occurrences in 22 of the 27 Brazilian states (absent in Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, and Sergipe). Of the 2,797 records tested, 917 fall in Brazil, with two of them flagged because they fall in the states of Alagoas and Pernambuco.

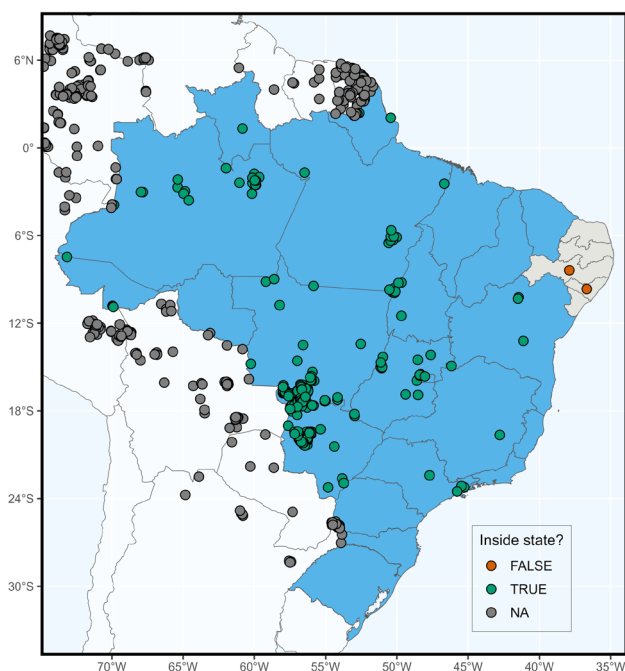


Figure 2. Map of flagged records for *Panthera onca* using the `filter_faunabr()` function. The blue-shaded area represents the states with confirmed occurrences of *Panthera onca* in Brazil. Green dots indicate records consistent with the distribution provided by specialists in CTFB. Orange dots represent records flagged as potentially incorrect due to falling outside the confirmed states. Grey dots represent records that were not tested because they do not fall within Brazil.

The last vignette demonstrates how to convert species distribution information in CTFB to generate a presence-absence matrix (PAM), where the rows represent sites (i) and the columns represent species (j). In a PAM, a value of one indicates the presence of species j in site i, while a value of zero signifies its absence. PAMs enable the estimation of various metrics that summarize biodiversity patterns, including richness, range size, and community composition (Soberon and Cavner 2015, Soberón et al. 2021). The `fauna_pam()` function returns a PAM where the rows represent Brazilian states or countries, and the columns represent the subset of species of interest. In addition to the PAM, the function also provides a summary and a map (SpatVector) containing the number of species in each site. This summary is useful, for example, for describing the biodiversity of different groups in Brazilian states, as done in several studies (Ament and Pereira 2024, Duarte and Lecci 2024, Krolow et al. 2024, Machado and Martins 2023, Santarém and Felipe-Bauer 2024). In the example from the vignette, I generated a PAM and a summary of species richness including all mammal species native to Brazil (Fig. 3).

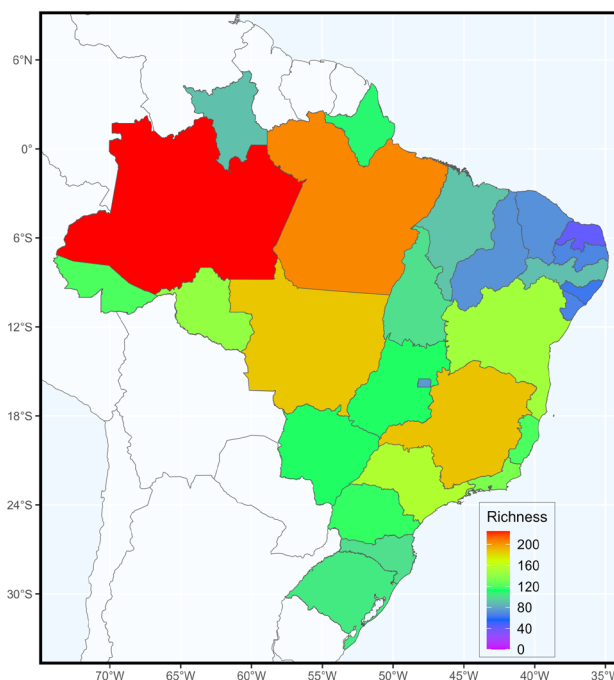


Figure 3. SpatVector returned by the `fauna_pam()` function indicating the number of species in each site. In this context, each site represents a state in Brazil. Warmer colors indicate higher species richness.

To the best of my knowledge, *faunabr* is the only R package that connects to the CTFB, the most comprehensive resource for reliably documenting Brazilian fauna diversity. In addition to interacting with the latest version (or any other version) of the CTFB, the package also facilitates the selection of target species based on multiple criteria, as well as the option to use distribution information to generate a presence-absence matrix (PAM) and flag occurrence records based on expert knowledge. I encourage and welcome user feedback to address issues and suggest new features.

ACKNOWLEDGMENTS

I thank the coordinators and researchers of the CTFB for their continued commitment to maintaining and enhancing this essential database.

LITERATURE CITED

- Ament DC, Pereira TPL (2024) Diversity, distribution and general knowledge of the Brazilian Phoridae (Diptera), the family of flies with most species known in the country. *Zoologia* 41: e24010. <https://doi.org/10.1590/S1984-4689.v41.e24010>
- Benetti CJ, Alencar JBR, Hamada N (2024) Taxonomic Catalog of the Brazilian Fauna: Hydraenidae (Insecta: Coleoptera), diversity and distribution. *Zoologia* 41: e23069. <https://doi.org/10.1590/S1984-4689.v41.e23069>
- BFG, Gomes-da-Silva J, Filardi FLR, Barbosa MRV, Baumgratz JFA, et al. (2021) Brazilian Flora 2020: Leveraging the power of a collaborative scientific network. *TAXON* 71: 178–198. <https://doi.org/10.1002/tax.12640>
- Boeger WA, Valim MP, Zaher H, Rafael JA, Forzza RC, et al. (2024) Catálogo Taxonômico da Fauna do Brasil: Setting the baseline knowledge on the animal diversity in Brazil. *Zoologia* 41: e24005. <https://doi.org/10.1590/S1984-4689.v41.e24005>
- Duarte T, Lecci LS (2024) A Scientometric Approach to the Taxonomy of Brazilian Plecoptera: An Overview of Data. *Revista Brasileira de Entomologia* 67: e20230056. <https://doi.org/10.1590/1806-9665-RBENT-2023-0056>
- Krolow TK, Carmo DDD, Oliveira LP, Henriques AL (2024) The Tabanidae (Diptera) in Brazil: Historical aspects, diversity and distribution. *Zoologia* 41: e23074. <https://doi.org/10.1590/S1984-4689.v41.e23074>
- Lai J, Cui D, Zhu W, Mao L (2023) The Use of R and R Packages in Biodiversity Conservation Research. *Diversity* 15: 1202. <https://doi.org/10.3390/d15121202>
- Lai J, Lortie CJ, Muenchen RA, Yang J, Ma K (2019) Evaluating the popularity of R in ecology. *Ecosphere* 10: e02567. <https://doi.org/10.1002/ecs2.2567>
- Machado RJP, Martins CC (2023) The extant fauna of Neuroptera (Insecta) from Brazil: diversity, distribution and history. *Revista Brasileira de Entomologia* 66: e20220083. <https://doi.org/10.1590/1806-9665-RBENT-2022-0083>
- Peterson AT, Asase A, Canhos DAL, de Souza S, Wiczorek J (2018) Data leakage and loss in biodiversity informatics. *Biodiversity Data Journal* 6: e26826. <https://doi.org/10.3897/BDJ.6.e26826>
- Santarém MCA, Felipe-Bauer ML (2024) Analysis of Brazilian Ceratopogonidae (Diptera: Culicomorpha) species diversity and knowledge assessment. *Zoologia* 41: e23066. <https://doi.org/10.1590/S1984-4689.v41.e23066>
- Soberon J, Cavner J (2015) Indices of Biodiversity Pattern Based on Presence-Absence Matrices: A GIS Implementation. *Biodiversity Informatics* 10: 22–34. <https://doi.org/10.17161/bi.v10i0.4801>
- Soberón J, Cobos ME, Nuñez-Penichet C (2021) Visualizing species richness and site similarity from presence-absence matrices. *Biodiversity Informatics* 16: 20–27. <https://doi.org/10.17161/bi.v16i1.14782>
- Trindade WCF (2024) florabr: An R package to explore and spatialize species distribution using Flora e Funga do Brasil. *Applications in Plant Sciences* 12(6): e11616. <https://doi.org/10.1002/aps3.11616>
- Victorino BI, Penteado-Dias AM, Dias Filho MM (2023) New species and new records of Ichneumonidae (Hymenoptera) in Southeast Brazil. *Brazilian Journal of Biology* 83: e266746. <https://doi.org/10.1590/1519-6984.266746>

Submitted: April 22, 2025

Accepted: August 7, 2025

Editorial responsibility: Walter A.P. Boeger

Author Contributions

The author conducted all aspects of the research, analysis, and writing.

Competing Interests

The author has declared that no competing interests exist.

Data Availability

No datasets were generated or analyzed in this study. Data sharing is not applicable to this article. The R package developed in this study is publicly available on CRAN (<https://cran.r-project.org/package=faunabr>), and the source code is openly accessible

on GitHub at <https://github.com/wevertonbio/faunabr/>

Funding

This study was supported by Fundação Araucária as part of the NAPI Biodiversity: Ecosystem Services Project.

How to cite this article

Trindade WCF (2025) faunabr: An R package to explore taxonomic data and map species distributions using the

Catálogo Taxonômico da Fauna do Brasil. Zoologia 42: e25027. <https://doi.org/10.1590/S1984-4689.v42.e25027>

Published by

Sociedade Brasileira de Zoologia at Scientific Electronic Library Online – <https://www.scielo.br/zool>

Copyright

© 2025 The Author.