EphemBrazil: a curated online database and dashboard to explore the distribution of mayflies (Insecta: Ephemeroptera) from Brazil

Frederico Falcão Salles[1]* (iD), Otávio Luiz Fernandes[2], Rafael Boldrini[3],
Lucas Lima[4] (iD), Myllena Lima[5] (iD), Luiza Hoehne[6],
Emerson Medeiros Del Ponte[7]

[1]Universidade Federal de Viçosa, Museu de Entomologia, Departamento de Entomologia, Viçosa, MG, Brasil.
[2]Universidade Federal de Viçosa, Programa de Pós-graduação em Entomologia, Viçosa, MG, Brasil.
[3]Universidade Federal de Roraima, Centro de Estudos da Biodiversidade, Laboratório de Entomologia, Boa Vista, RR, Brasil.
[4]Universidade Estadual do Piauí, Laboratório de Zoologia, Campo Maior, PI, Brasil.
[5]Universidade Federal do Pará, Instituto de Ciências Biológicas, Laboratório de Ecologia e Conservação, Belém, PA, Brasil.
[6]Universidade Federal do Rio de Janeiro, Instituto de Biologia, Laboratório de Entomologia, Rio de Janeiro, RJ, Brasil.
[7]Universidade Federal de Viçosa, Departamento de Fitopatologia, Viçosa, MG, Brasil.

ARTICLE INFO

ABSTRACT

Despite the growing number of studies on mayflies (Ephemeroptera), especially in terms of description of new species and distributional records, the knowledge generated remains scattered across specialized literature. This dispersion poses a challenge for non-specialists, such as ecologists, hindering their ability to promote studies and conservation actions. In order to fill this gap, we are launching EphemBrazil: an online database focusing on the distribution of mayflies in Brazil. The main objectives of this platform are to compile a dataset of distributional information for the valid species of Ephemeroptera in Brazil, provide open access to this dataset through an interactive and easily updated website, and offer dynamic distributional maps for species, genera, and families of Brazilian mayflies. EphemBrazil aims to facilitate the retrieval and access of distributional and basic taxonomic information for the order.

Introduction

The conservation of biodiversity faces several obstacles, one of which is the lack of adequate knowledge on the description and cataloging of organisms. However, another significant challenge lies in the scarcity of information regarding their distribution (Cayuela et al., 2012). In the case of mayflies, insects belonging to the order Ephemeroptera, addressing this initial deficiency, known as the Linnean shortfall (Brown and Lomolino, 1998; Hortal et al., 2015; Faria et al., 2021), has garnered considerable attention over the past two decades in Brazil.

Since the publication of the first list of Brazilian mayflies (Salles et al., 2004), there has been an exponential increase in the number of papers focusing on the description of new taxa (Domínguez and Santos, 2014). Consequently, the number of mayfly species reported in Brazil has grown from 166 to the current count of 430 (Salles and Boldrini, 2023). However, knowledge about the distribution of mayflies in Brazil has received significantly less attention and faces two major problems: a lack of knowledge itself, referred to as the Wallacean shortfall (Brown and Lomolino, 1998; Hortal et al., 2015; Faria et al., 2021), and a lack of a platform for the integration and access to the existing information.

Despite the growing number of studies on Brazilian mayflies, the knowledge generated remains scattered across specialized literature. This dispersion poses a challenge for non-specialists, such as ecologists, hindering their ability to promote studies and conservation actions. Without sufficient information on species' habitats, it becomes impossible to identify species at risk or concentrate efforts on their preservation (Cardoso et al., 2011).

Mitigating the Wallacean shortfall relies on collecting specimens, conducting species-level identification, and publishing these records in

*Corresponding author.
E-mail: ffsalles@gmail.com (F.F. Salles).*

scientific literature or museum database systems. Similarly, addressing the lack of integration and access necessitates compiling accurate distributional data, curating such data, and making it available on platforms that can be frequently updated and corrected. While there are excellent initiatives for the South American fauna and specifically for Brazil, such as the data availability by Molineri et al. (2019, 2020), the Ephemeroptera da América do Sul website (Salles et al., 2023) presenting distributions of species by country, the Taxonomic Catalog of the Fauna of Brazil (Salles and Boldrini, 2023) allowing verification of species distribution by state, and SALVE (ICMBio, 2023), capable of generating maps for each species occurring in Brazil (at least until the last assessment of the risk of extinction of the Brazilian fauna), none of them precisely fit the aforementioned description. For example, the original source of the data provided by Molineri et al. (2019, 2020) is not traceable; the distribution presented on the websites Ephemeroptera da América do Sul and Taxonomic Catalog of the Fauna of Brazil does not include detailed information; and, in the case of SALVE, it's not possible to get a complete overview of the Brazilian fauna.

To bridge this gap, we are launching EphemBrazil: an online database focusing on the distribution of mayflies (Insecta: Ephemeroptera) in Brazil. The main objectives of this platform are to compile a dataset of distributional information for the valid species of Ephemeroptera in Brazil, provide open access to this dataset through an interactive and easily updated website, and offer dynamic distributional maps for species, genera, and families of Brazilian mayflies. EphemBrazil aims to facilitate the retrieval and access of distributional and basic taxonomic information for the order.

Materials and methods

Taxonomic data

To create an updated list of mayfly species from Brazil, we gathered current species lists from various sources, including the Taxonomic Catalog of the Brazilian Fauna (CTFB) (http://fauna.jbrj.gov.br/fauna/listaBrasil/ConsultaPublicaUC/ConsultaPublicaUC.do), the List of South American Ephemeroptera Species (Domínguez et al., 2018; http://ibn-conicet.gob.ar/), and Ephemeroptera of South America (Salles et al., 2023; http://ephemeroptera.com.br/). Utilizing this data, we compiled a preliminary list of species recorded in Brazil. Finally, we created a list of valid species by reviewing all species descriptions or records from Brazil up until May 2023. In order to maintain the accuracy of the data, the website will be updated at least at the beginning of each semester.

Distributional data

Initially, we accessed a digital database called Biodiversity Conservation Status Assessment System (SALVE), developed by ICMBio (The Chico Mendes Institute for Biodiversity Conservation), to integrate all distributional information on Brazilian mayflies. The team from the Museum of Entomology at the Federal University of Viçosa, Brazil, coordinated by F. F. Salles, in cooperation with ICMBio, conducted this compilation in order to assess the risk of extinction of the Brazilian fauna of Ephemeroptera in 2018. To do that, we conducted a bibliographic search of Ephemeroptera publications from 1843 (the year of the first publication mentioning a mayfly species from Brazil) to 2018 (for assessing the risk of extinction), updated herein to May 2023. Most of these publications were accessed from the following websites: Ephemeroptera Galactica (http://www.ephemeroptera-galactica.com/), Ephemeroptera of the World (http://insecta.bio.spbu.ru/z/Eph-spp/index.htm), Google Scholar (https://scholar.google.com.br/), and Zootaxa

(www.zootaxa.com.br). The coordinates were accessed directly from the articles or from the database presented by Molineri et al. (2019, 2020). In cases where papers lacked coordinates, we used the description of collection sites to find the most accurate coordinates on Google Maps (https://www.google.com.br/maps).

For each species, we tabulated standardized information, including the year and author of the description and the occurrence coordinates. Using SALVE, we generated a summary table containing all geographic coordinates for each species, including country, state, and municipality associated with each coordinate. The system also automatically searched for occurrences in protected areas. Furthermore, we utilized QGIS 3.10 software to overlay all coordinates with a shape file of Brazilian biomes, determining the biome associated with each coordinate.

The data was compiled and stored using a Google sheet™ to facilitate collaborative work. The spreadsheet includes distributional data, such as taxonomic information of valid species (order, family, genus, scientific name, author, and year), occurrence coordinates (latitude, longitude, datum), and information derived from each coordinate (federal and non-federal protected areas, private protected areas, country, state, municipality, title of scientific paper containing the coordinate, author, and year of record). This database will be regularly updated as new information becomes available.

Interactive dashboard

The dashboard was developed as a Shiny application (Chang et al., 2021) built using R (R Core Team, 2021) that can be viewed on a web browser in both desktops and smartphones. It features a structured user interface with a dashboard layout, including a sidebar menu and multiple tabs. The app follows a client-server architecture, where the user interface (UI) defines the layout and components, while the server logic handles data manipulation and interactivity. Several packages are utilized in the app, including shiny (Chang et al., 2021) and shinydashboard (Chang and Borges Ribeiro, 2021) for the user interface design, leaflet (Cheng et al., 2023) for mapping functionality, DT (Xie et al., 2021) for displaying tabular data, RColorBrewer (Neuwirth, 2022) for color palette definition, dplyr (Wickham et al., 2023a, 2023b) for data manipulation, plotly (Sievert et al., 2021) for graphical output, and readxl (Wickham and Bryan, 2019) for importing data from an Excel file generated from a Google Sheet™.

Results

Our updated list, as of May 2023, comprises 10 families, 81 genera, and 428 species. Our dataset includes a total of 3,360 occurrence records with georeferences for most of the species in Brazil. All of this data is available in the 'EphemBrazil' database (https://edelponte.shinyapps.io/ephembrazil/) (Fig. 1). The species *Caenis uruzu* Molineri & Malzacher, *Asthenopodes picteti* Hubbard, *Leptohyphes mollipes* Needham & Murphy, and *Tricorythodes cristatus* Allen do not have occurrence coordinates in our records. The articles describing these species exhibit inconsistencies in their distribution, lack of detailed information, or possible typing error that does not allow a location to be safely determined.

Data from EphemBrazil can be accessed in two forms. Firstly, raw data is presented in grids divided into two tabs: Distributional data and Bibliography (Fig. 2). Both tabs allow for the application of filters to search for specific information such as species, genus, family, biome, and/or by restricting the year of the record. There is also a search box available to filter data using specific words within each grid, and the grids can be downloaded either after filtering or as a whole.
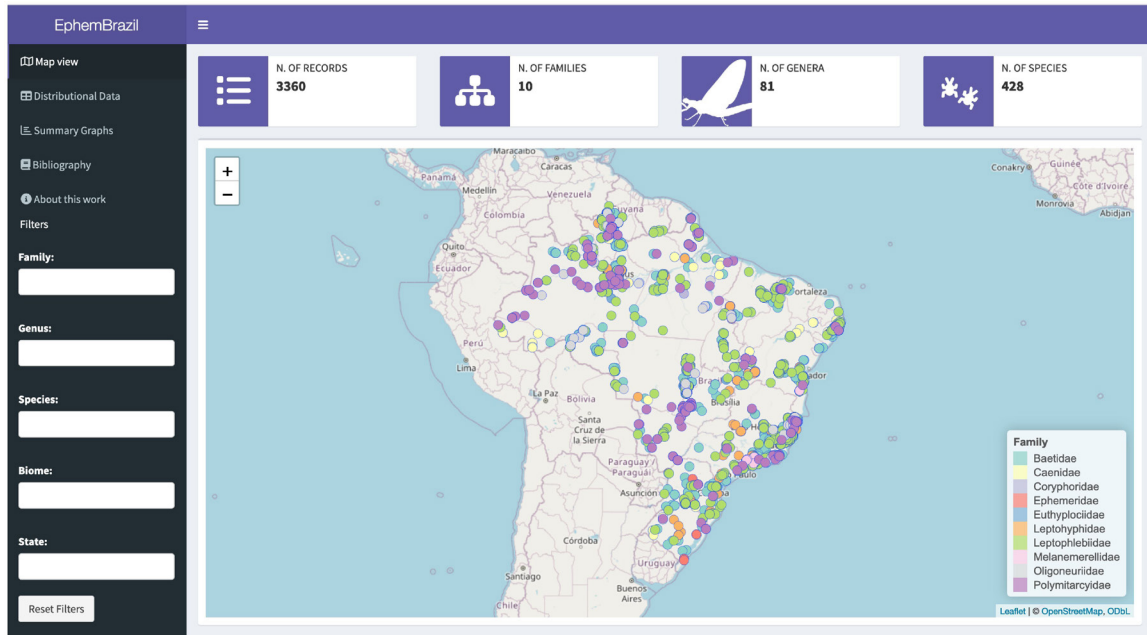
Figure 1 General view of the website and interactive map view tab showing filters on the top and family subtitles in the right corner. Note that no filter is applied and all records are shown.
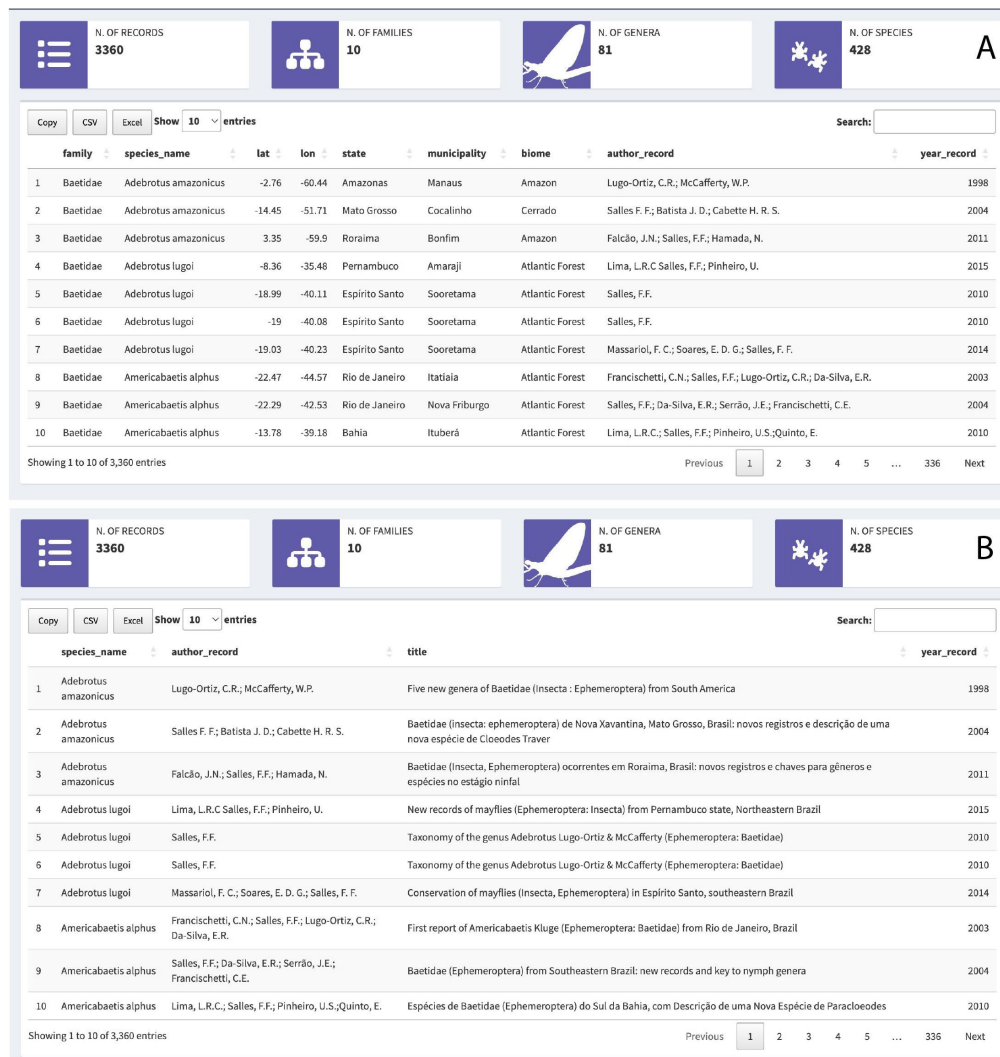

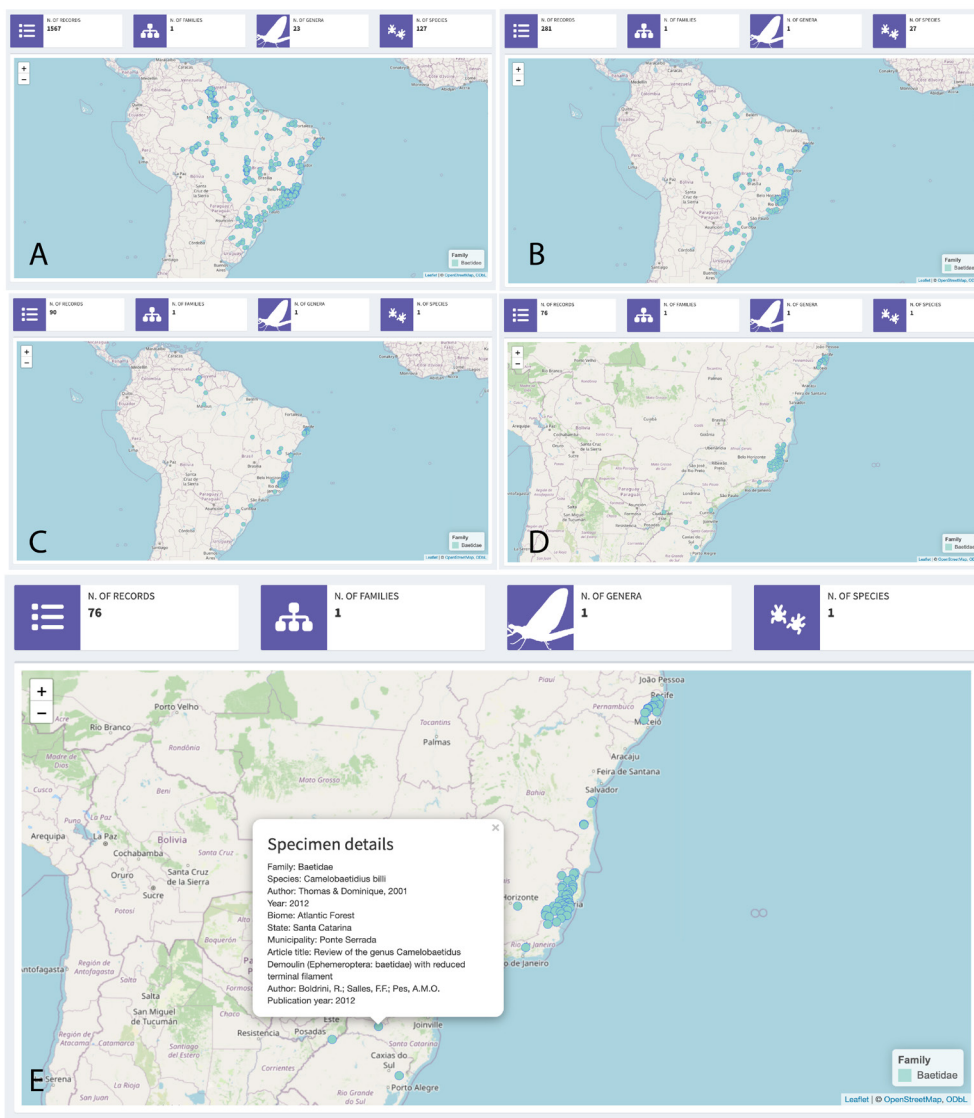**Figure 2** General view of (A) Distributional data grid; and (B) Bibliographic data grid.

Figure 3 Interactive map view of distributional records in Brazil with different filters applied: (A) family Baetidae; (B) genus *Camelobaetidius* Demoulin; (C) species *Camelobaetidius billi* Thomas & Dominique; (D) *Camelobaetidius billi* Thomas & Dominique records in Atlantic Forest biome; (E) Detailed view of a single record of *Camelobaetidius billi* Thomas & Dominique.

Secondly, the data can be accessed through the interactive map view (Fig. 1), where all distributional records are represented by colored dots corresponding to the families. The interactive map view can also be filtered, as explained above for the grids (Fig. 3). Clicking on each bullet point opens a balloon with details of the species record (Fig. 3E). Furthermore, we provide a tab with summary graphs of all records in the database, categorized by family, biome, Brazilian state, and year of record (Fig. 4).

Discussion

The 'EphemBrazil' database represents a unique and pioneering effort to compile and provide access to georeferenced and taxonomic information for the order Ephemeroptera in Brazil. Species-specific distribution data are crucial for predicting geographical ranges and understanding species responses to global change, which are essential aspects of conservation planning (Serra-Diaz and Franklin, 2019). We strongly believe that an open-access, readily available, and comprehensive database can facilitate proactive conservation measures

and enhance scientific endeavors in areas with limited knowledge. As the database continues to grow, it will help identify knowledge gaps for specific taxa, enabling focused resource allocation to maximize research value and address biodiversity shortfalls.

Like other open-access biodiversity databases worldwide (Cayuela et al., 2012; Lopez-Gonzalez et al., 2012; Twardochleb et al., 2021), 'EphemBrazil' can be utilized to investigate ecological communities, analyze species turnover along environmental gradients, and study population dynamics. Specifically for Ephemeroptera, complete datasets of species distribution have been applied in the study of vicariance events (Molineri et al., 2019) and the delimitation of hypothetical biogeographic zones (Molineri et al., 2020). Furthermore, our database can be used to map the taxonomic diversity of mayflies, and when coupled with environmental data, it becomes a valuable resource for quantifying how environmental variables shape diversity patterns.

Moreover, the integration and accessibility of georeferenced species information provided by the website simplify the assessment of extinction risk. The IUCN criteria for Ephemeroptera, for instance, include the Extent of Occurrence (EOO), defined as "the area contained
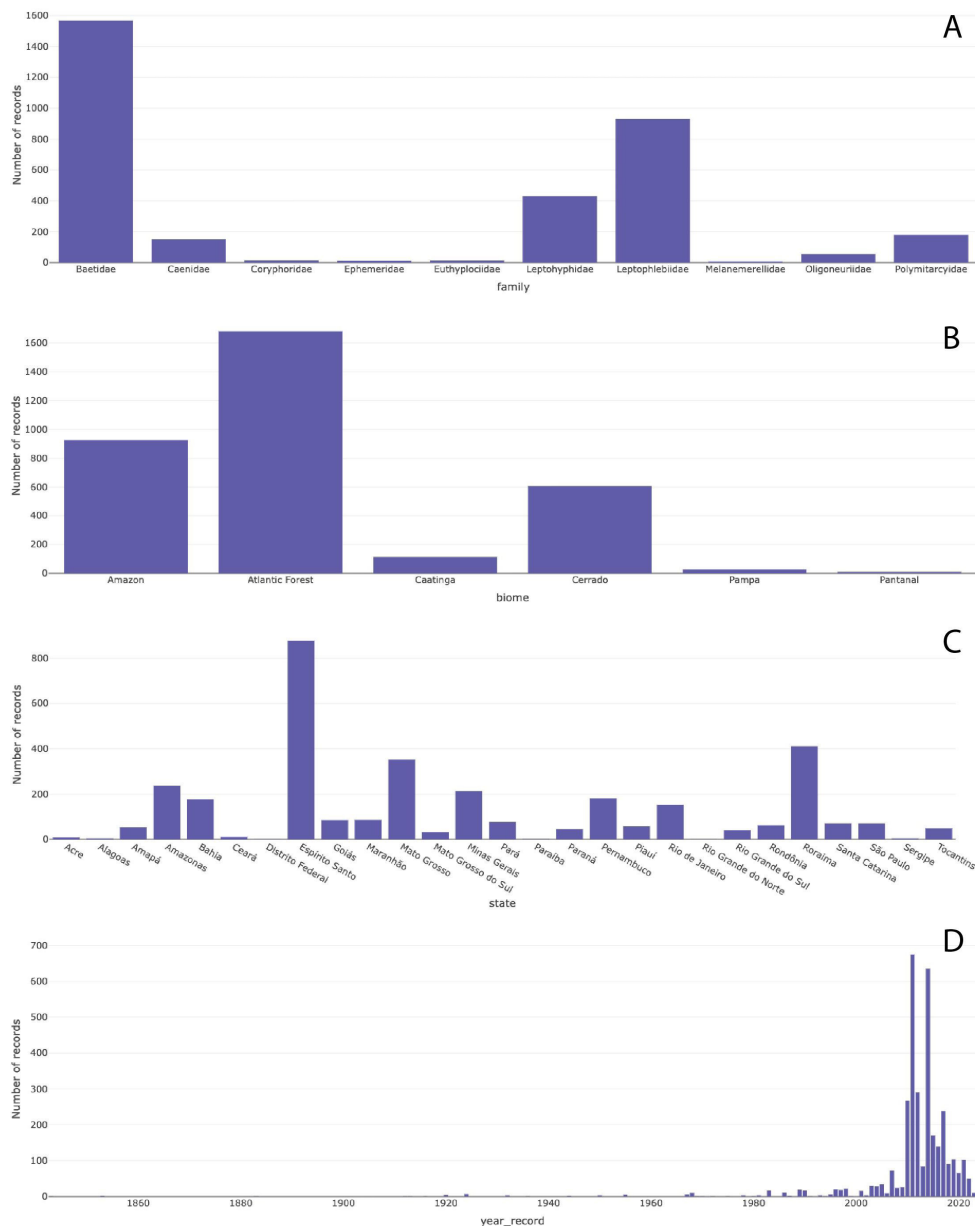
Figure 4 Summary chart of records of Ephemeroptera from Brazil by: (A) family; (B) biome; (C) state; (D) year of record.

within the shortest continuous imaginary boundary that can be drawn to encompass all sites of present occurrence of a taxon" (IUCN, 2019). The compilation of all occurrence records for each species is necessary for determining its EOO. As the database is continually updated with more occurrence records, the extinction risk status of a species may change as its extent of occurrence expands.

Acknowledgments

Conflicts of interest

The authors declare no conflicts of interest.

Author contribution statement

FFS, OLF and EMP conceived the study. All authors contributed with the data. OLF and EMP analyzed data. FFS, OLF and EMP wrote the paper. FFS secured funding. All authors read and approved the manuscript.

References

Brown, J. H., Lomolino, M. V., 1998. Biogeography, 2nd ed. Sinauer Press, Sunderland.

Cardoso, P., Erwin, T. L., Borges, P. A. V., New, T. R., 2011. The seven impediments in invertebrate conservation and how to overcome them. Biol. Conserv. 144 (11), 2647-2655. http://dx.doi.org/10.1016/j.biocon.2011.07.024.

Cayuela, L., Gálvez-Bravo, L., Pérez-Pérez, R., de Albuquerque, F., Golicher, D., Zahawi, R., Ramírez-Marcial, N., Garibaldi, C., Field, R., Rey Benayas, J., González-Espinosa, M., Balvanera, P., Ángel Castillo, M., Figueroa-Rangel, B., Griffith, D., Islebe, G., Kelly, D., Olvera-Vargas, M., Schnitzer, S., Velázquez, E., Williams-Linera, G., Brewer, S., Camacho-Cruz, A., Coronado, I., de Jong, B., del Castillo, R., Granzow-de la Cerda, Í., Fernández, J., Fonseca, W., Galindo-Jaimes, L., Gillespie, T., González-Rivas, B., Gordon, J., Hurtado, J., Linares, J., Letcher, S., Mangan, S., Meave, J., Méndez, E., Meza, V., Ochoa-Gaona, S., Peterson, C., Ruiz-Gutierrez, V., Snarr, K., Tun Dzul, F., Valdez-Hernández, M., Viergever, K., White, D., Williams, J., Bonet, F., Zamora, R., 2012. The Tree Biodiversity Network (BIOTREE-NET): prospects for biodiversity research and conservation in the Neotropics. Biodivers. Ecol. 4, 211-224. http://dx.doi.org/10.7809/b-e.00078.

Chang, W., Borges Ribeiro, B., 2021. Shinydashboard: Create Dashboards with 'Shiny'. R Package Version 0.7.2. R Foundation for Statistical Computing, Vienna. Available in: https://CRAN.R-project.org/package=shinydashboard (accessed 1 May 2023).

Chang, W., Cheng, J., Allaire, J., Xie, Y., McPherson, J., 2021. Shiny: Web Application Framework for R: R Package Version 1.6.0. R Foundation for Statistical Computing, Vienna. Available in: https://CRAN.R-project.org/package=shiny (accessed 1 May 2023).

Cheng, J., Karambelkar, B., Xie, Y., 2023. Leaflet: Create Interactive Web Maps with the Javascript 'Leaflet' Library. R Package Version 2.1.2. R Foundation for Statistical Computing, Vienna. Available in: https://CRAN.R-project.org/package=leaflet (accessed 1 May 2023).

Domínguez, E., Molineri, C., Nieto, C., Zúñiga, M., 2018. Lista de especies de Ephemeroptera Sudamericanos. IBN, (CONICET-UNT), Tucumán, Argentina. Available in: http://ibn.conicet.gov.ar/ (accessed 1 May 2023).

Domínguez, E., Santos, D. A., 2014. Co-authorship networks (and other contextual factors) behind the growth of taxonomy of South American Ephemeroptera: a scientometric approach. Zootaxa 3754 (1), 59-85. http://dx.doi.org/10.11646/zootaxa.3754.1.3.

Faria, L. R. R., Pie, M. R., Salles, F. F., Soares, E. D. G., 2021. The Haecklian shortfall or the tale of missing semaphoronts. Zool. Syst. Evol. Res. 59 (2), 359-369. http://dx.doi.org/10.1111/jzs.12435.

Hortal, J., De Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., Ladle, R. J., 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. Annu. Rev. Ecol. Evol. Syst. 46 (1), 523-549. http://dx.doi.org/10.1146/annurev-ecolsys-112414-054400.

Instituto Chico Mendes de Conservação da Biodiversidade – ICMBio, 2023. SALVE: risco de extinção da fauna brasileira. Available in: https://salve.icmbio.gov.br/#/ (accessed 1 May 2023).

International Union for Conservation of Nature and Natural Resources – IUCN, 2019. Guidelines for Using the IUCN Red List Categories and Criteria. Available in: http://www.iucnredlist.org/documents/RedListGuidelines.pdf (accessed 1 May 2023).

Lopez-Gonzalez, G., Burkitt, M., Lewis, S. L., Phillips, O. L., 2012. Forest-plots.net: managing permanent plot information across the tropics. In: Dengler, J., Oldeland, J., Jansen, F., Chytrý, M., Ewald, J., Finckh, M., Glöck-ler, F., Lopez-Gonzalez, G., Peet, R.K., Schaminée, J.H.J. (Eds.), Vegetation Databases for the 21st Century. Biocentre Klein Flottbek and Botanical Garden, Hamburg, pp. 95-103. (Biodiversity & Ecology, 4).

Molineri, C., Nieto, C., Domínguez, E., 2019. Direct analysis of vicariance in Neotropical mayflies (Ephemeroptera). An. Acad. Bras. Cienc. 91 (4), e20181130. http://dx.doi.org/10.1590/0001-3765201920181130.

Molineri, C., Nieto, C., Santos, D. A., Emmerich, D., Zúñiga, M. C., Fierro, P., Pessacq, P., Gomez, D., Márquez, J. A., Príncipe, R. E., 2020. Do mayflies (Ephemeroptera) support a biogeographic transition zone in South America? J. Biogeogr. 47 (9), 1980-1993. http://dx.doi.org/10.1111/jbi.13868.

Neuwirth, E., 2022. RColorBrewer: ColorBrewer Palettes. R Package Version 1.1-3. R Foundation for Statistical Computing, Vienna. Available in: https://cran.r-project.org/web/packages/RColorBrewer/index.html (accessed 10 August 2023).

R Core Team, 2021. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. Available in: https://www.R- project.org/ (accessed 10 August 2023).

Salles, F. F., Boldrini, R., 2023. Catálogo taxonômica da fauna do Brasil (Ephemeroptera). JBRJ, Rio de Janeiro. Available in: http://fauna.jbrj.gov.br/ (accessed 1 May 2023).

Salles, F. F., Da-Silva, E. R., Hubbard, M. D., Serrão, J. E., 2004. As espécies de Ephemeroptera (Insecta) registradas para o Brasil. Biota Neotrop. 4 (2), 1-34. http://dx.doi.org/10.1590/S1676-06032004000200011.

Salles, F. F., Molineri, C., Nieto, C., Lima, L. R. C., Dias, L. C., Boldrini, R., Mariano, R., Domínguez, E., 2023. Ephemeroptera da América do Sul. Available in: http://ephemeroptera.com.br/ (accessed 1 May 2023).

Serra-Diaz, J. M., Franklin, J., 2019. What's hot in conservation biogeography in a changing climate? Going beyond species range dynamics. Divers. Distrib. 25 (4), 492-498. http://dx.doi.org/10.1111/ddi.12917.

Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., Despouy, P., Sievert, M. C., 2021. Package 'Plotly'. R Foundation for Statistical Computing, Vienna. Available in: https://cran.r-project.org/package=plotly (accessed 10 August 2023).

Twardochleb, L., Hiltner, E., Pyne, M., Zarnetske, P., 2021. Freshwater insects CONUS: a database of freshwater insect occurrences and traits for the contiguous United States. Glob. Ecol. Biogeogr. 30 (4), 826-841. http://dx.doi.org/10.1111/geb.13257.

Wickham, H., Bryan, J., 2019. readxl: Read Excel Files. Available in: https://CRAN.R-project.org/package=readxl (accessed 9 August 2023).

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., 2023a. dplyr: a grammar of data manipulation. Available in: https://dplyr.tidyverse.org (accessed 9 August 2023).

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., 2023b. tidyverse/dplyr. Available in: https://github.com/tidyverse/dplyr (accessed 9 August 2023)

Xie, Y., Cheng, J., Tan, X., 2021. DT: a Wrapper of the JavaScript Library 'DataTables'. R Package Version 0.17. R Foundation for Statistical Computing, Vienna. Available in: https://CRAN.R-project.org/package=DT (accessed 10 August 2023).