





scientific data



OPEN

DATA DESCRIPTOR

A database of freshwater macroinvertebrate occurrence records across Cuba

Yusdiel Torres-Cambas^{1,2}✉, Yoandri S. Megna³, Juan Carlos Salazar-Salina³ , Yander L. Diez^{4,5} , Alejandro Catalá³, Adrian D. Trapero-Quintana⁶, Boris Schröder⁶  & Sami Domisch¹ 

In light of the ongoing freshwater biodiversity crisis, detailed knowledge regarding the spatial distribution of freshwater species is urgently required, especially in biodiversity hotspots. Here we present a database of georeferenced occurrence records of four freshwater invertebrate taxa groups across Cuba, namely flatworms (Platyhelminthes: Tricladida), insects (Ephemeroptera, Odonata, Hemiptera, Trichoptera, Coleoptera, Diptera), crabs and shrimps (Crustacea: Decapoda), and mollusks (Mollusca). We collated the geographic occurrence information from scientific literature, unpublished field records, museum collections and online databases. The database, comprising 6292 records of 457 species at 1075 unique localities, is organized in 32 fields that contain the information about the taxonomic classification of each recorded species, the sex and life stage of collected individuals; the geographic coordinates, location, author and date of the record and a reference to the original data source. This database provides an important basis towards an improved understanding of the spatial distribution of freshwater biodiversity in Cuba.

Background & Summary

The ongoing biodiversity crisis represents a major challenge that requires urgent countermeasures, where freely and universally accessible key scientific information is needed to address and ultimately reverse the loss of biodiversity worldwide. The spatial distribution of biodiversity can be considered such crucial information. In this regard, the lack of knowledge regarding species distributions, the so-called “Wallacean shortfall”, can be considered a bottleneck in the efforts to protect biodiversity¹. To fill this gap, many international initiatives are promoting to digitize, geo-reference and share species distribution information that has been amassed in the past through open-access online repositories. For example, the Global Biodiversity Information Facility (GBIF), the world’s largest open-access biodiversity database, stores more than 2 billion occurrence records (August 2022, <https://www.gbif.org>). Despite this impressive number of records, several extremely diverse taxa groups such as freshwater invertebrates are underrepresented and there is a lack of biodiversity data from freshwater ecosystems in tropical regions^{2,3}.

Freshwater macroinvertebrates consist of a heterogeneous group of organisms that includes, for example, several orders of insects, crustaceans, mollusks, annelids or flatworms. Both in terms of species richness or functional diversity and biomass, freshwater macroinvertebrates are dominant in the communities they are part of. They have in common that at least one stage of their life cycle is associated with aquatic ecosystems. This dependency on freshwater makes them particularly vulnerable to alterations in their habitats, and consequently freshwater macroinvertebrates are considered sentinels regarding the integrity of aquatic ecosystems.

To address the shortcoming regarding the availability of freshwater species occurrences, we present a database⁴ of freshwater macroinvertebrates for Cuba, which represents a biodiversity hotspot that has however not

¹Leibniz Institute of Freshwater Ecology and Inland Fisheries, Department of Community and Ecosystem Ecology, Müggelseedamm 310, D-12489, Berlin, Germany. ²Institute of Geocology, Technische Universität Braunschweig, Braunschweig, D-38106, Germany. ³Departamento de Biología y Geografía, Facultad de Ciencias Naturales y Exactas, Universidad de Oriente, Santiago de Cuba, 90500, Cuba. ⁴Museum of Nature Hamburg - Zoology, Leibniz Institute for the Analysis of Biodiversity Change (LIB), Martin-Luther-King-Platz 3, D-20146, Hamburg, Germany. ⁵Centre for Environmental Sciences, Research Group Zoology: Biodiversity and Toxicology, Universitaire Campus Gebouw D, B-3590, Diepenbeek, Belgium. ⁶Departamento de Biología Animal y Humana, Facultad de Biología de la Universidad de la Habana, La Habana, 10400, Cuba. ✉e-mail: yusdiel.torres-cambas@igb-berlin.de

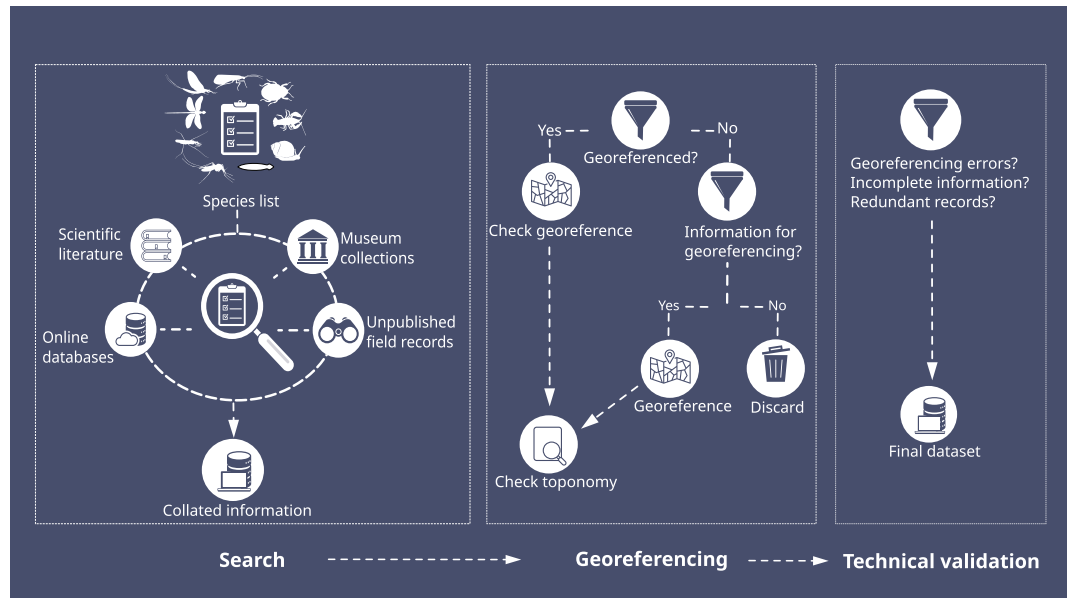


Fig. 1 Workflow followed to compile a dataset of freshwater macroinvertebrates (i.e. Platyhelminthes, Ephemeroptera, Odonata, Hemiptera, Trichoptera, Coleoptera, Decapoda, Mollusca) from Cuba. Occurrence records were obtained from diverse sources, georeferenced and checked for possible errors and omissions.

seen much attention to date. Our new database⁴ consists of georeferenced occurrence records of freshwater flatworms (Platyhelminthes: Tricladida), insects (Ephemeroptera, Odonata, Hemiptera, Trichoptera, Coleoptera, Diptera), crabs and shrimps (Crustacea: Decapoda), and mollusks (Mollusca) from Cuba. So far, information regarding the occurrence of freshwater species in Cuba was difficult to access by the international scientific community, since the data is scattered, or it lacks spatial information (i.e., coordinates) which is however crucial for data reusability. The Fig. 1 represents an schematic overview of the workflow followed to compile our database.

In our database, we have now included the following data sources for which the coordinates for occurrence localities were not published to date. This includes important museum collections which host specimens of Ephemeroptera⁵, Odonata⁶, Hemiptera⁷, Trichoptera⁸ and Coleoptera (Hydrophiloidea, Histeroidea)⁹, but which have not yet been digitized and georeferenced (e.g. *Museo de Historia Natural Charles Ramsden de la Torre, Universidad de Oriente; Museo de Historia Natural Felipe Poey, Universidad de La Habana*), or if digitized, the information is not hosted in online and openly accessible repositories (e.g. *Instituto de Ecología y Sistemática, La Habana*). Likewise, the database for diving beetles (Dytiscidae)¹⁰ with spatial explicit occurrence information, based on a revision of the literature and collections as well as field samplings conducted between 2000 and 2014, has so far not been publicly accessible. For the remaining taxa groups, maps with all known occurrence points (Ephemeroptera³) for each species, list of localities (Hemiptera⁷) or more general revisions of the distribution in Cuba has been published (Odonata^{11,12}, Trichoptera⁸) which we collate under one comprehensive database.

The main contribution of the present database⁴ is that for the first time, occurrence records of major groups of freshwater macroinvertebrates from Cuba, with curated spatial information are compiled under the same standards (Darwin Core standards^{13,14}) and made openly available according to the FAIR principles¹⁵. With a geographic focus on the Cuban archipelago, we expect that the database is of interest to a broader scientific community focused on the distribution and conservation of freshwater biodiversity and macroinvertebrates in particular.

Methods

We created an updated species list⁴ (file “species_list.csv” in the database) of 590 species of tricladid flatworms (Tricladida, phylum Platyhelminthes), the insect orders Ephemeroptera (mayflies), Odonata (dragonflies and damselflies), Trichoptera (caddiflies) and freshwater Hemiptera (water bugs), Coleoptera, Decapoda (crabs and shrimps, subphylum Crustacea) and Mollusca after a revision of the literature (Table 1). We considered only binomial scientific names (genus + species), avoiding the use of subgenus and subspecies categories. This list⁴ (file “species_list.csv” in the database) was our taxonomic reference to search information from four different sources:

- Scientific literature. We conducted a search with Google Scholar using a combination of the keyword “Cuba” with “freshwater”, “Ephemeroptera”, “Odonata” or “Trichoptera” and “Cuba + freshwater” with “Platyhelminthes”, “Coleoptera”, “Heteroptera”, “Decapoda”, “Diptera”, “Mollusca”. Additionally, we consulted a collection of theses between 1980 and 2020 at Universidad de Oriente, Santiago de Cuba. We found occurrence records in 102 journal articles, 11 books and book chapters, 17 thesis dissertations and 3 scientific reports that were published between 1888 and 2022.
- Museum collections, which included the *Museo de Historia Natural Charles Ramsden de la Torre (Universidad de Oriente, Cuba)*, the *Museo de Historia Natural Felipe Poey (Universidad de La Habana, Cuba)*, and the

Taxon	Richness	Reference
Coleoptera: Dryopidae, Dytiscidae, Elmidae, Gyrinidae, Haliplidae, Hydraenidae, Hydrochidae, Hydrophilidae, Lutrochidae, Noteridae, Scirtidae	128	9,10,25–30,47,52–55
Crustacea: Decapoda: Atyidae, Cambaridae, Grapsidae, Palaemonidae, Pseudothelphusidae	37	56,57
Diptera: Blephariceridae, Chironomidae, Culicidae, Simuliidae	95	58,59
Ephemeroptera	34	5
Hemiptera: Heteroptera: Dipsocoromorpha, Leptopodomorpha, Gerromorpha and Nepomorpha	73	7
Mollusca: Basommatophora, Caenogastropoda, Cycloneritomorpha, Littorinomorpha, Neritimorpha, Sorbeoconcha, Unionida, Venerida	44	60,61
Odonata	88	12
Trichoptera	89	8
Platyhelminthes: Tricladida	2	62,63

Table 1. Known species richness of the taxonomic groups included in the dataset of freshwater macroinvertebrates from Cuba and literature consulted to create the species list.

entomological collection of the *Instituto de Ecología y Sistemática (La Habana, Cuba)*. We identified additional 33 institutions that host freshwater invertebrate specimens after the revision of the scientific literature or online databases (Table 2).

- Online databases. We searched in the Global Biodiversity Information Facility portal (GBIF, <https://www.gbif.org>, access date: 25-02-2022) and iNaturalist (<https://www.inaturalist.org/>, access date: 25-02-2022) using the R-package *rgbif*¹⁶ and *rinat*¹⁷ respectively. From iNaturalist, we included only “Research Grade” records in our database, after crosschecking the identification based on the photos included in this portal web. The data obtained from GBIF includes records from 26 different datasets¹⁸.
- Unpublished field records (n = 260), collected by the authors between 2001 and 2022.

Sixteen percent (n = 1016) of records had been georeferenced by the source of information. For these we cross-checked the georeference using the standard web-client of the GEOLocate software (<https://www.geo-locate.org/web/WebGeoref.aspx>) and corrected if required (Table 3). We assigned coordinates to the records that lacked them only when the locality, or information such as municipality name, name of the water body, mountain range, city or town were available. In contrast, we discarded records that had only “Cuba” or a Cuban province name as spatial information. Overall, coordinate uncertainty in the database ranges from 30 m to 90892 m (median = 363 m). We added the coordinate uncertainty as an additional attribute to each record (field “coordinateUncertaintyInMeters” in the database).

To find missing coordinates of the sampling sites, we conducted the following procedure. First, we searched for a record’s locality and coordinates with GEOLocate. Second, and if a locality was not found, we searched for the given locality in a Cuban 1:50 000 map (Instituto de Geodesia y Cartografía, Cuba). We assigned coordinate uncertainty in meters to all localities using GEOLocate (Table 3). We checked locality and water body names and corrected when needed, based on a gazetteer¹⁹ and cartographic sheets (Cuban 1:50 000 map) (Table 3). We checked the municipality and province names and updated these according to the last political-administrative organization of Cuba in 2011²⁰ (<http://www.onei.gob.cu/sites/default/files/dpa.pdf>).

Data Records

The database⁴ consists of 5 files: “species_list.csv” with the species list used as taxonomic reference, “occurrence_records.csv”, with occurrence records; “fields_database.docx”, that provides a complete list of the fields present in “occurrence_records.csv” and two remaining files with the scripting procedures used to query the GBIF and iNaturalist databases and for technical validation (“scr_down_o_gbif_inat.R”, “scr_tech_val.R”). The database is deposited at Figshare and can be downloaded from <https://doi.org/10.6084/m9.figshare.21155419>. In addition, the occurrence records can be queried through an interactive web application created by us with the programming language R at https://y-torres-cambas.shinyapps.io/fw_cu_dash/.

The occurrence records⁴ (“occurrence_records.csv”) are organized in 32 fields that contains information about the taxonomic classification of each recorded species, the sex and life stage of individuals, the geographic coordinates and the location (e.g. locality, water body name), the author and date of the record and the reference (i.e. URL, bibliographic reference or “newly published in this work”) to the original data source. Twenty-nine of these fields are named according to the Darwin Core standards, which is an internationally accepted vocabulary for disseminating biodiversity information^{13,14}. Two of the remaining fields are based on GBIF (i.e., gbifID) and the Freshwater Core Template at the Freshwater Biodiversity Data Portal (i.e., waterBodyType, <https://data.freshwaterbiodiversity.eu/>). The last field consists of a sub-catchment identification number for each occurrence, extracted from the most recent high-resolution global hydrographic dataset Hydrography90m²¹. Coordinates are in decimal degrees of the World Geodetic System 1984 (WGS84).

Our database⁴ (“occurrence_records.csv”) comprises a total of 6292 records of 457 species, at 1075 unique localities from 269 drainage basins of the Hydrography90m²¹ and 1080 sub-catchments (Fig. 2). Most of the records correspond to citations (i.e., a reference or citation in scholarly publications^{13,14}) or preserved specimens (i.e., records from museum collections^{13,14}) and, to a lesser extent, to human observations (referring to the Darwin Core standard, i.e. “an output of a human observation process” or “a record of an occurrence without

Institution	Country
American Museum of Natural History, New York	USA
Colección del Departamento de Zoología, Centro Oriental de Ecosistemas y Biodiversidad, Santiago de Cuba	Cuba
Colorado Entomological Museum	USA
Departamento de Zoología, Universidad de Murcia	Spain
Florida Museum of Natural History (UF)	USA
Florida State Collection of Arthropods	USA
Forschungsinstitut und Natur-Museum Senckenberg (SMF)	Germany
Illinois Natural History Survey	USA
Instituto de Biología de la Universidad Nacional Autónoma de México	Mexico
Instituto de Ecología y Sistemática, La Habana	Cuba
Instituto Nacional de Pesquisas da Amazônia, Manaus	Brazil
Museo de Historia Natural Charles Ramsden de la Torre, Universidad de Oriente, Santiago de Cuba	Cuba
Museo de Historia Natural de Holguín	Cuba
Museo de Historia Natural Felipe Poey, Universidad de La Habana	Cuba
Museo de Zoología, Escuela de Biología, Universidad de Costa Rica, San José (UCR-MZ)	Costa Rica
Museo di Storia Naturale dell'Università di Firenze	Italy
Museum für Naturkunde, Berlin (ZMB)	Germany
Museum of Comparative Zoology, Harvard University	USA
National Museum of Natural History, Smithsonian Institution, Washington, D.C.	USA
National Museum, Prague	Czech R.
Natural History Museum of Rotterdam (NMR)	Netherlands
Naturgucker	Germany
Naturhistorisches Museum Wien	Austria
New Mexico State University (NMSU)	USA
North Carolina Museum of Natural Sciences (NCSM)	USA
Personal collection of Nico Nieser	Netherlands
Speleological Institute of Bucharest	Romania
The Field Museum of Natural History, Chicago	USA
Université Laval	Canada
University of Kansas	USA
University of Michigan Museum of Zoology (UMMZ)	USA
University of Minnesota Insect Collection (UMSP)	USA
Zoological Institute at Leningrad, Russia	Russia
Zoological Muesum La Specola, Florence	Italy
Zoological Museum of the University of Amsterdam	Netherlands
Zoologische Staatssammlung München	Germany

Table 2. List of institutions that served as a source of occurrence records of freshwater macroinvertebrates from Cuba.

Source	Coordinates	Locality
Scientific literature	3185 (92%)	14 (0.4%)
Unpublished records	260 (100.0%)	0 (0%)
Museum collections	1545 (90%)	0 (0%)
Online databases	317 (38%)	181 (22%)
Total	5307	195

Table 3. Occurrence records that underwent additions or updates in their coordinates or locality. The records are organized according to their source of origin. Percent values are relative to the total records in each source class.

physical evidence^{13,14} (Fig. 4). The 23% ($n = 1434$) and 21% ($n = 1324$) of records have information about the sex and life stage (larvae, nymph, exuviae, juvenile, adult, ovigerous female), respectively.

The records were collected between 1888 and 2022, with a highest frequency during the decades of 1960s, 1970s, 1990s, 2000s and 2010s (Fig. 3). Between 1960s–1970s, institutionalization of science in Cuba and an extensive collaboration with European scientists lead to the exploration of Cuban mountain ranges, the description of new species and data collection on species distribution, particularly in the case of freshwater invertebrates. For example, the “Cuban–Romanian Biospeleological Expeditions”, conducted during the

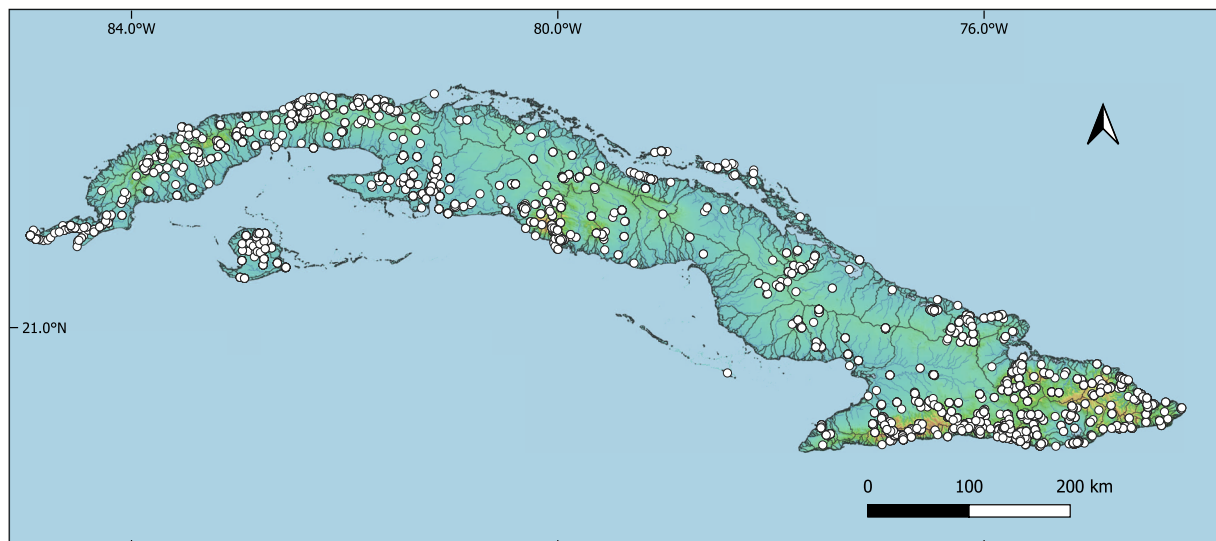


Fig. 2 Geographic distribution of freshwater macroinvertebrates (Platyhelminthes: Tricladida, Ephemeroptera, Odonata, Hemiptera, Trichoptera, Coleoptera, Diptera, Crustacea: Decapoda, Mollusca) occurrence records across Cuba. Catchments are demarcated with black lines.

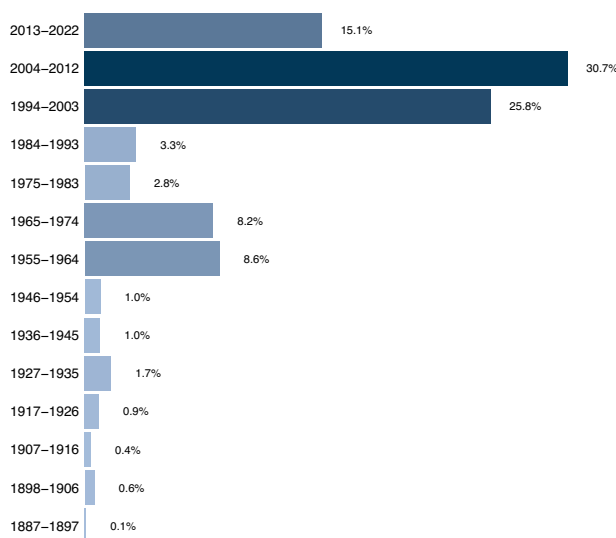


Fig. 3 Temporal distribution of occurrence records in a dataset of freshwater macroinvertebrates from Cuba.

1960s, contributed strongly to the knowledge of Trichoptera²² and freshwater Decapoda²³ and Coleoptera²⁴. From 1990 to 2021 a series of studies conducted at the Department of Biology, Universidad de Oriente, Cuba or by researchers associated to this institution, on the fields of taxonomy and distribution^{5,7,12,25–34}, diversity of freshwater macroinvertebrate assemblages^{35–42} and bio-monitoring⁴³ in rivers of central and eastern Cuba, contributed significantly to the number of records during this period.

The 77% ($n = 457$) of the freshwater macroinvertebrate species listed in our species list reference⁴ (“species_list.csv”) have at least one geographic record. All taxa except Diptera have a high taxonomic representativeness, with 77% to 100% of species with at least one record (Table 4). The number of records per species ranges from 1 (for 87 species), to 189 in the Band-winged Dragonlet *Erythrodiplax umbrata* (Odonata: Libellulidae) (Fig. 5). The median number of records in each taxon is between 1 in Diptera to 23 in Odonata (Table 4). The 35% ($n = 159$) of the species in the file “occurrence_records.csv”⁴ have more than 10 records, however this percent value varies across taxa (Fig. 5).

The order Odonata (dragonflies and damselflies) comprises most entries in the database ($n = 2675$, 43%, Table 4, “occurrence_records.csv”⁴). The taxonomic bias towards this order of insect can be attributed to the high detectability of Odonata in the field, and the ease to achieve a taxonomic identification of a given species. Odonata are therefore the freshwater macroinvertebrates that can be most easily detected and identified both in the laboratory and in the field in Cuba, which facilitates the recording of occurrences, including those by

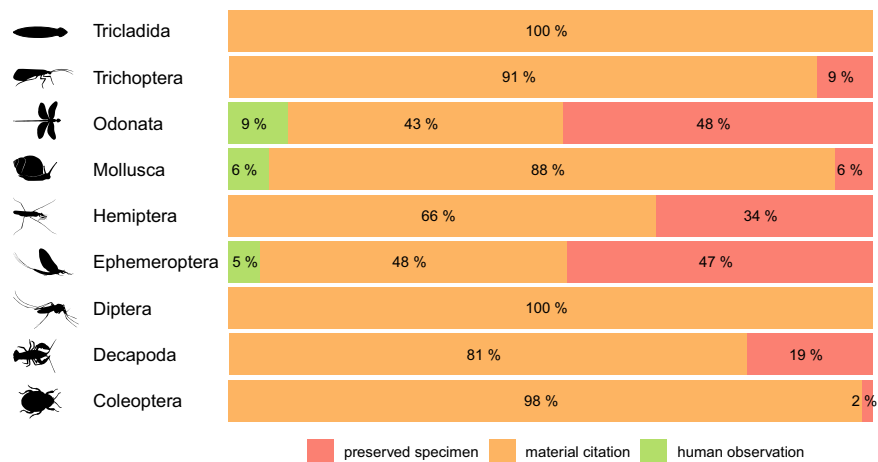


Fig. 4 Proportion of records in different taxonomic groups according to the source of origin in a dataset of freshwater species from Cuba. The classification adopted (i.e. preserved specimen, material citation, human observation) is based on Darwin Core standards^{13,14}.

Taxon	N. occurrences	P. occurrences	Sp. one record	T. representativeness	Median	Med. abs. dev.
Coleoptera	1011	16%	113	88%	5	4
Decapoda	236	4%	34	92%	4	3
Diptera	90	1%	26	27%	1	0
Ephemeroptera	827	13%	29	85%	19	25
Hemiptera	465	7%	66	90%	3	3
Mollusca	481	8%	34	77%	10	11
Odonata	2675	43%	78	89%	23	20
Trichoptera	501	8%	75	84%	3	3
Tricladida	6	0.1%	2	100%	3	3

Table 4. Descriptive statistics of the occurrence records in a dataset of freshwater macroinvertebrates from Cuba. N. occurrences: number of occurrences. P. occurrences: percent of occurrences in Taxon relative to the total occurrence number in the dataset. Sp. one record: number of species with at least one occurrence record in the dataset. T. representativeness: taxonomic representativeness, percent of species with records relative to the known species richness in each taxon. Median: median occurrence records by taxon. Med. abs. dev.: median absolute deviation.

non-expert taxonomists and citizen scientists. For example, Odonata is the taxon with the highest proportion of records by human observation (Fig. 4), and 97% of records downloaded by us from the citizen-scientist-based online database iNaturalist correspond to Odonata (“occurrence_records.csv”⁴).

Contrary to the Odonata and the remaining taxa groups (except for Tricladida), the identification of freshwater dipterans generally requires microscope preparations and taxonomic expertise, which increases the difficulties regarding the identification to the species level. In the literature that holds information about the distribution of freshwater Diptera in Cuba, Diptera specimens, with a few exceptions^{44,45}, are generally identified to genus or family^{35–39} and were therefore not included in our database⁴.

Technical Validation

The occurrence records⁴ were screened for georeferencing errors (e.g. points in the ocean), incomplete information and redundant information with functions of the R-packages *dplyr* and *sf* (file “scr_tech_val.R”⁴). Each record must have a date, the taxonomic (species, genus, family, class, phylum) and geographic (locality, latitude, longitude, municipality and province) information and a reference to the source. A valid date must include at least the year. If two or more records of a same species had the same coordinates and date, we considered these as duplicates and only kept unique records.

Usage Notes

The database⁴ has a high potential in spatial freshwater biodiversity analyses, biogeography and area-based conservation planning. For instance, the species geographic records can be used to model their range-wide distributions using species distribution models (SDM)⁴⁶ with applications in the conservation of freshwater biodiversity in Cuba, for example:

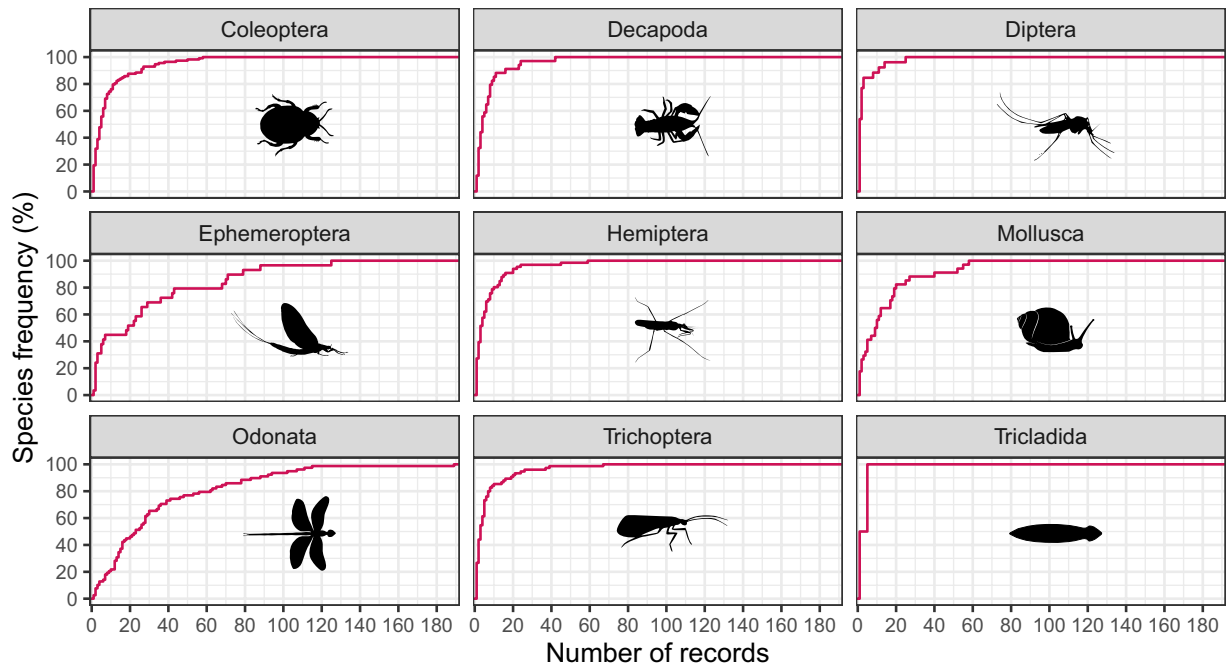


Fig. 5 Cumulative frequency of species across the number of records per species in a dataset of freshwater macroinvertebrates from Cuba.

- The identification of gaps in the National System of Protected Areas (e.g.⁴⁷).
- The assessment of the conservation status of Cuban endemic species (e.g.³³).
- The development of proactive strategies for the protection of freshwater species under future climate change (e.g.⁴⁸).
- The improvement of the National System of Protected Areas to meet the spatial connectivity requirements of freshwater ecosystems (e.g.⁴⁹).

We highlight that the database is welcoming future contributions. Among the taxa analyzed here, Mollusca is the taxa group for which the distribution data is not yet openly available despite that the data has been collated, digitized, georeferenced and stored systematically, such as by the Laboratory of Malacology at the Tropical Medicine Institute of Cuba who maintains a relational database with this information from 1980s to the present^{50,51}.

Code availability

The R codes used to query the GBIF and iNaturalist databases (“scr_down_o_gbif_inat.R”) and for technical validation (“scr_tech_val.R”) are available at <https://doi.org/10.6084/m9.figshare.21155419>.

Received: 9 November 2022; Accepted: 20 March 2023;

Published online: 27 March 2023

References

1. Lomolino, M. V. Conservation biogeography. In Lomolino, M. V. & Heaney, L. R. (eds.) *Frontiers Of Biogeography: New Directions In The Geography Of Nature*, 293–96 (Sinauer Associates, Inc., Sunderland, Massachusetts, 2004).
2. Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. & Legendre, F. Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports* 7, 9132, <https://doi.org/10.1038/s41598-017-09084-6> (2017).
3. Hughes, A. C. *et al.* Sampling biases shape our view of the natural world. *Ecography* 44, 1259–1269, <https://doi.org/10.1111/ecog.05926> (2021).
4. Torres-Cambas, Y. *et al.* Database of freshwater macroinvertebrate occurrences across Cuba. *figshare* <https://doi.org/10.6084/m9.figshare.21155419> (2022).
5. Naranjo-López, C., Peters, J. G. & López, P. Ephemeroptera (Insecta) in Cuba. *Insecta Mundi* 0736, 1–52 (2019).
6. Reyes, E. & Alvarez, A. Odonatos (Insecta) depositados en el Instituto de Ecología y Sistemática, Ciudad de La Habana. *Cocuyo* 11, 11–15 (2001).
7. Riviaux, S. M., Moreira, F. F. F. & Naranjo-López, C. Checklist, distribution, and habitat of the semiaquatic and aquatic bugs from Cuba (Hemiptera: Heteroptera: Dipsocoromorpha, Leptopodomorpha, Gerromorpha and Nepomorpha). *Zootaxa* 23, 1–23 (2010).
8. Naranjo-López, C. & González-Lazo, D. Situación actual del estudio del orden Trichoptera en Cuba. *Boletín de la Sociedad Entomológica Aragonesa* 36, 147–152 (2005).
9. Fernández, I., Reyes, E., Álvarez, D., Deler, A. & Megna, Y. Colección entomológica “Juan C. Gundlach”: Superfamilias Hydrophiloidea e Histeroidea (Coleoptera: Staphyliniformia). *Poeyana* 498, 7–10 (2010).
10. Megna, Y. & Sánchez-Fernández, D. Catálogo faunístico de la familia Dytiscidae (Coleoptera: Adephaga) en Cuba. *Boletín de la Sociedad Entomológica Aragonesa* 59, 159–180 (2016).

11. Trapero-Quintana, A. D. & Naranjo-López, C. Revision of the order Odonata in Cuba. *Bulletin of American Odonatology* **7**, 23–40 (2003).
12. Trapero-Quintana, A., Torres-Cambas, Y. & Martínez, A. Las libélulas de Cuba: una actualización taxonómica 10 años después. *Revista Cubana de Ciencias Biológicas* **10**, 1–8 (2018).
13. Wieczorek, J., Bloom, D., Guralnick, R., Blum, S. & Dö Ring, M. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE* **7**, 29715, <https://doi.org/10.1371/journal.pone.0029715> (2012).
14. Darwin Core Maintenance Group. Darwin Core quick reference guide. Biodiversity Information Standards (TDWG). <https://dwc.tdwg.org/terms/> (2021).
15. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* **3**, 1–9, <https://doi.org/10.1038/sdata.2016.18> (2016).
16. Chamberlain, S. Package ‘rgbif’: Interface to the Global Biodiversity Information Facility API. R package version 3.7.2. <https://cran.r-project.org/package=rgbif> (2022).
17. Barve, V. & Hart, E. rinat: Access ‘iNaturalist’ Data Through APIs. R package version 0.1.9. <https://docs.ropensci.org/rinat/> (2022).
18. Derived dataset GBIF.org. Filtered export of GBIF occurrence data <https://doi.org/10.15468/dd.kdr9xp> (21 September 2022).
19. Comisión Nacional de Nombre Geográficos. *Diccionario Geográfico De Cuba* (Oficina Nacional de Hidrografía y Geodesia, La Habana, 2000).
20. Méndez Delgado, C. E. & Lloret Feijoo, M. D. C. La división político administrativa en Cuba: antecedentes y actualidad. *Territorios y Regionalismos* **5**, 1–11 (2021).
21. Amatulli, G. *et al.* Hydrography90m: A new high-resolution global hydrographic dataset. *Earth System Science Data Discussions* **2022**, 1–43, <https://doi.org/10.5194/essd-2022-9> (2022).
22. Botosaneanu, L. & Sykora, J. L. Sur quelques Trichoptères (Insecta: Trichoptera) de Cuba. In Orghidan T., Núñez, A., Botosaneanu, L., Decu, V., Negrea, St., Viña, N. (ed.) *Résultats Des Expéditions Biospéologiques Cubano-Roumaines Á Cuba*, 379–407 (Editura Academiei Republicii Socialiste Romania, Bucharest, 1973).
23. Holthuis, L. B. On some freshwater and terrestrial Crustacea Decapoda from Cuba. In Orghidan T. *et al.* (ed.) *Résultats Des Expéditions Biospéologiques Cubano-Roumaines Á Cuba*, 271–275 (Editura Academiei Republic Socialiste Romania, Bucarest, 1977).
24. Spangler, P. J. Aquatic Coleoptera, collected by the Biospeliological expedition to Cuba by the Academy of Science of Cuba and Rumania (Gyrinidae: Dytiscidae: Hydrophilidae: Elmthinidae: Psephenidae). In Orghidan T. *et al.* (ed.) *Résultats Des Expéditions Biospéologiques Cubano-Roumaines Á Cuba I*, 353–358 (Editura Academiei Republicii Socialiste Romania, Bucharest, 1973).
25. Deler-Hernández, A. & Delgado, J. A. The Hydraenidae of Cuba (Insecta: Coleoptera) I: Contribution to the fauna of eastern Cuba. *Zootaxa* **3478**, 213–238, <https://doi.org/10.11646/zootaxa.3478.1.23> (2012).
26. Deler-Hernández, A., Fikáček, M. & Cala-riquelme, F. A review of the genus *Berosus* Leach of Cuba. *Zookeys* **273**, 73–106, <https://doi.org/10.3897/zookeys.273.4591> (2013).
27. Deler-Hernández, A. & Delgado, J. A. The Hydraenidae of Cuba (Insecta: Coleoptera) II: Morphology of preimaginal stages of six species and notes on their biology. *Zootaxa* **4238**, 451–498, <https://doi.org/10.11646/zootaxa.4238.4.1> (2017).
28. Megna, Y. S. & Deler-Hernández, A. Composición taxonómica, distribución y biología de la familia Noteridae (Coleoptera: Adepaga) en Cuba. *Revista de la Sociedad Entomológica Argentina* **65**, 69–80 (2006).
29. Megna, Y. S. & Epler, J. H. A review of *Copelatus* from Cuba, with the description of two new species (Coleoptera: Dytiscidae: Copelatinae). *Acta Entomologica Musei Nationalis Pragae* **52**, 1–28 (2012).
30. Megna, Y. S. *La familia Dytiscidae (Coleoptera: Adepaga) en Cuba: taxonomía, distribución y conservación*. Phd thesis, Universidad de Oriente (2014).
31. Riviaux, S. M., López, C. N. & Moreira, F. F. F. New distributional records of Gerromorpha Popov, 1971 (Insecta: Hemiptera: Heteroptera) from Cuba. *Check List* **8**, 307–309, <https://doi.org/10.15560/8.2.307> (2012).
32. Torres-Cambas, Y. *et al.* An update on the distribution of threatened odonate species from the Greater Antilles. *International Journal of Odonatology* **18**, 89–104, <https://doi.org/10.1080/13887890.2014.928241> (2015).
33. Torres-Cambas, Y., Cabana-Otero, M., Lorenzo-Carballa, M. O. & Cordero-Rivera, A. Conservation status and protection of three Antillean endemic damselflies. *Journal of Insect Conservation* **20**, 277–284, <https://doi.org/10.1007/S10841-016-9862-1> (2016).
34. Trapero-Quintana, A., Torres-Cambas, Y., Bello, O. C. & Naranjo-López, C. Annotated list of the Odonates in the Alexander Humboldt National Park, Guantánamo Province, Cuba. *Argia* **16**, 16–19 (2004).
35. Deler, A., Megna, Y. S., González, D. D. & Carcacs, C. N. Insectos acuáticos y áreas prioritarias para la conservación en la cuenca alta del río Cauto (Santiago de Cuba, Cuba). *Boletín de la Sociedad Entomológica Aragonesa* **40**, 451–461 (2007).
36. González-Lazo, D., Trapero-Quintana, A. & Naranjo-López, C. Insectos acuáticos del Parque Nacional Alejandro de Humboldt, Cuba. *Boletín de la Sociedad Entomológica Aragonesa* **36**, 257–261 (2005).
37. López, P., González, D. D., Trapero-Quintana, A. D. & Pérez, J. Insectos acuáticos del Parque Nacional “La Bayamesa”, Cuba. *Boletín de la Sociedad Entomológica Aragonesa* **35**, 255–231 (2004).
38. López, P., Naranjo-López, C., González-Lazo, D. & Naranjo-López, C. Lista de insectos acuáticos de la Reserva Ecológica Alturas de Banao, Sancti Spiritus, Cuba (Insecta). *Boletín de la Sociedad Entomológica Aragonesa* **38**, 201–204 (2006).
39. Muñoz, S., Naranjo-López, C., González-Lazo, D. & Imbert, T. Insectos acuáticos del Área Protegida Resolladero del río Cuzco (Provincia Guantánamo, Cuba). *Boletín de la Sociedad Entomológica Aragonesa* **44**, 497–503 (2009).
40. Trapero-Quintana, A. & Reyes-Tur, B. Patrones de emergencia de odonata en las Cuabas, Santiago de Cuba. *Boletín de la Sociedad Entomológica Aragonesa* **46**, 273–280 (2010).
41. Trapero-Quintana, A., Reyes-Tur, B. & Mateu, J. Distancia sobre el agua durante la emergencia en larvas de Odonata para tres cuerpos dulceacuícolas de Cuba Oriental. *Dugesiana* **17**, 103–111 (2010).
42. Trapero-Quintana, A. & Reyes-Lafuente, Y. Diversidad y emergencia de odonatos en los Cantiles, Provincia Granma, Cuba. *Boletín de la Sociedad Entomológica Aragonesa* **53**, 217–222 (2013).
43. Muñoz-Riveaux, S. *et al.* Evaluación de la calidad del agua utilizando los macroinvertebrados bentónicos como bioindicadores. *Revista Chapingo Serie Ciencias Forestales y del Ambiente* **9**, 147–153 (2003).
44. Izquierdo, K. *et al.* Diversidad de Chironomidae (Insecta: Diptera) en lagunas del humedal Guanahacabibes, Cuba. *ECOVIDA* **6**, 73–84 (2016).
45. Martínez, B. T., Quintana, A. T., Cambas, Y. T. & Roque, F. O. Effects of wood density on aquatic insect communities in a Cuban montane rainforest stream. *Neotropical Entomology* **48**, 527–537, <https://doi.org/10.1007/s13744-019-00670-z> (2019).
46. Domisch, S., Jähnig, S. C., Simaika, J. P., Kuemmerlen, M. & Stoll, S. Application of species distribution models in stream ecosystems: the challenges of spatial and temporal scale, environmental predictors and species occurrence data. *Fundamental and Applied Limnology* **186**, 45–61, <https://doi.org/10.1127/fal/2015/0627> (2015).
47. Megna, Y. S., Pallarés, S. & Sánchez-Fernández, D. Conservation of aquatic insects in Neotropical regions: A gap analysis using potential distributions of diving beetles in Cuba. *Aquatic Conservation: Marine and Freshwater Ecosystems* **31**, 2714–2725, <https://doi.org/10.1002/aqc.3664> (2021).
48. Bush, A. *et al.* Freshwater conservation planning under climate change: demonstrating proactive approaches for Australian Odonata. *Journal of Applied Ecology* **51**, 1273–1281, <https://doi.org/10.1111/1365-2664.12295> (2014).
49. Hermoso, V., Kennard, M. J. & Linke, S. Integrating multidirectional connectivity requirements in systematic conservation planning for freshwater systems. *Diversity and Distributions* **18**, 448–458, <https://doi.org/10.1111/j.1472-4642.2011.00879.x> (2012).

50. Piedra, R. M., de Puga, G. P., Cong, M. Y., Rodríguez, J. C. D. & Ferrer, J. R. A computer program for the geographic distribution of fluvial mollusks of medical interest in Cuba. *Revista Cubana de Medicina Tropical* **46**, 175–182 (1994).
51. Perera, A. A. V. & Valderrama, S. P. Endemic freshwater molluscs of Cuba and their conservation status. *Tropical Conservation Science* **3**, 190–199, <https://doi.org/10.1177/194008291000300206> (2010).
52. Megna, Y. S., Hendrich, L. & Balke, M. A new species of *Halipilus* Latreille, 1802 from Cuba (Coleoptera: Haliplidae). *Aquatic Insects* **39**, 1–9, <https://doi.org/10.1080/01650424.2018.1430834> (2018).
53. Shepard, W. D. & Megna, Y. S. Los byrrhoideos acuáticos (Coleoptera: Byrrhoidea) de Cuba. *Cocuyo* **16**, 32–35 (2006).
54. Peck, S. B. *A Checklist Of The Beetles Of Cuba With Data On Distributions And Bionomics (Insects: Coleoptera). Arthropods Of Florida And Neighboring Land*. Vol. **18** (Florida Department of Agriculture and Consumer Services, Gainesville, 2005).
55. Gustafson, G. & Miller, K. The New World whirligig beetles of the genus *Dineutus* Macleay, 1825 (Coleoptera, Gyrrinidae, Gyrrininae, Dineutini). *ZooKeys* **476**, 1–135, <https://doi.org/10.3897/zookeys.476.8630> (2015).
56. Juarrero, A. An updated checklist of the inland water's shrimps of Cuba. *Cocuyo* **9**, 10–11 (1999).
57. Ng, P. K. L., Guinot, D. & Davie, P. J. F. *Systema Brachyurorum: Part 1. An Annotated checklist of extant brachyuran crabs of the world. Raffles Bulletin of Zoology* **17**, 1–286 (2008).
58. Bello, O. C., Spies, M. & Téllez, B. Estado del conocimiento de la familia Chironomidae (Insecta: Diptera) en Cuba. *Dugesiana* **20**, 233–242 (2013).
59. González Broche, R. & Rodríguez, J. Lista actualizada de los mosquitos de Cuba (Diptera: Culicidae). *Cocuyo* **6**, 17–18 (1997).
60. Hernández Quita, M. *et al.* Moluscos terrestres y dulceacuicolas. In Mancina, C. A. & Cruz, D. D. (eds.) *Diversidad Biológica De Cuba: Métodos De Inventario, Monitoreo Y Colecciones Biológicas*, 168–195 (Editorial AMA, La Habana, 2017).
61. MolluscaBase. VLIZ. <https://doi.org/10.14284/448> (2022).
62. Sluys, R. Synopsis of the freshwater triclads of the Caribbean (Platyhelminthes, Tricladida, Paulodica). *Studies on the Natural History of the Caribbean Region* **71**, 1–23 (1992).
63. Benítez-Álvarez, L., Sluys, R., Leal-Zanchet, A., Leria, L. & Riutort, M. First molecular phylogeny of the freshwater planarian genus *Girardia* (Platyhelminthes: Tricladida) unveils hidden taxonomic diversity and initiates resolution of its historical biogeography. *Zoological Journal of the Linnean Society* **197**, 1–9, <https://doi.org/10.1093/zoolinnean/zlac065> (2022).

Acknowledgements

YT-C and YLD were funded by a Georg Foster Postdoctoral Fellowship of the Alexander von Humboldt Foundation, respectively (Ref 3.2 - CUB - 1212347 - GF-P; Ref 3.2 - CUB - 1226121 - GF-P). YSM was supported by the Alexander von Humboldt Foundation through a Digital Cooperation Fellowship (Ref. 3.4 - CUB/1161268). SD was funded by the Leibniz Competition (J45/2018). The authors acknowledge funding by the German Federal Ministry of Education and Research (BMBF grant agreement number no. 033W034A). Domenico Capolongo (Roccarainola, Naples, Italy) provided valuable literature on *Epilobocera* from Cuba.

Author contributions

Y.T.-C. conceived the data collection and data validation; Y.T.-C., Y.S.M., J.C.S.S., Y.D. and A.C. collated the information. All authors analysed the results. Y.T.-C. wrote the first draft of the manuscript, and all authors contributed to the writing of the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.T.-C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023