Received: 9 May 2022 | Revised: 18 January 2023 | Accepted: 24 January 2023

DOI: 10.1111/geb.13648

DATA ARTICLE

Global Ecology and Biogeography A Journal of Macroecology WILEY

The global EPTO database: Worldwide occurrences of aquatic insects

Afroditi Grigoropoulou[1,2] | Suhaila Ab Hamid[3] | Raúl Acosta[4] | Emmanuel Olusegun Akindele[5] | Salman A. Al-Shami[6] | Florian Altermatt[7,8] | Giuseppe Amatulli[1,9] | David G. Angeler[10] | Francis O. Arimoro[11] | Jukka Aroviita[12] | Anna Astorga-Roine[13] | Rafael Costa Bastos[14,15] | Núria Bonada[4,16] | Nikos Boukas[17] | Cecilia Brand[18,19] | Vanessa Bremerich[1] | Alex Bush[20] | Qinghua Cai[21,22] | Marcos Callisto[23] | Kai Chen[24,25] | Paulo Vilela Cruz[26] | Olivier Dangles[27] | Russell Death[28] | Xiling Deng[29] | Eduardo Domínguez[30] | David Dudgeon[31] | Tor Erik Eriksen[32] | Ana Paula J. Faria[15] | Maria João Feio[33] | Camino Fernández-Aláez[34] | Mathieu Floury[1,35] | Francisco García-Criado[34] | Jorge García-Girón[34,36] | Wolfram Graf[37] | Mira Grönroos[38] | Peter Haase[29,39] | Neusa Hamada[40] | Fengzhi He[1] | Jani Heino[36] | Ralph Holzenthal[41] | Kaisa-Leena Huttunen[42] | Dean Jacobsen[43] | Sonja C. Jähnig[1,44] | Walter Jetz[9] | Richard K. Johnson[10] | Leandro Juen[15] | Vincent Kalkman[45] | Vassiliki Kati[17] | Unique N. Keke[11] | Ricardo Koroiva[46,47] | Mathias Kuemmerlen[48] | Simone Daniela Langhans[49] | Raphael Ligeiro[15] | Kris Van Looy[50] | Alain Maasri[1,51] | Richard Marchant[52] | Jaime Ricardo Garcia Marquez[1] | Renato T. Martins[40] | Adriano S. Melo[53] | Leon Metzeling[54] | Maria Laura Miserendino[18,19] | S. Jannicke Moe[32] | Carlos Molineri[30] | Timo Muotka[42] | Kaisa-Riikka Mustonen[42] | Heikki Mykrä[12] | Jeane Marcelle Cavalcante do Nascimento[40,55] | Francisco Valente-Neto[56] | Peter J. Neu[54] | Carolina Nieto[30] | Steffen U. Pauls[29] | Dennis R. Paulson[57] | Blanca Rios-Touma[58] | Marciel Elio Rodrigues[59] | Fabio de Oliveira Roque[60] | Juan Carlos Salazar Salina[61] | Dénes Schmera[62] | Astrid Schmidt-Kloiber[37] | Deep Narayan Shah[63] | John P. Simaika[64] | Tadeu Siqueira[65,66] | Ram Devi Tachamo-Shah[67] | Günther Theischinger[68] | Ross Thompson[69] | Jonathan D. Tonkin[66,70,71] | Yusdiel Torres-Cambas[1] | Colin Townsend[72] | Eren Turak[73] | Laura Twardochleb[74] | Beixin Wang[24] | Liubov Yanygina[75] | Carmen Zamora-Muñoz[76,†] | Sami Domisch[1]

†Deceased.

Dennis R. Paulson is now retired.

[1]Department of Community and Ecosystem Ecology, Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany

[2]Department of Biology, Chemistry, Pharmacy, Institute of Biology, Freie Universität Berlin, Berlin, Germany

[3]School of Biological Sciences, Universiti Sains Malaysia, Penang, Malaysia

[4]FEHM-Lab (Freshwater Ecology, Hydrology and Management), Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Facultat de Biologia, Universitat de Barcelona (UB), Barcelona, Spain

[5]Department of Zoology, Obafemi Awolowo University, Ile-Ife, Nigeria

[6]Indian River Research and Education Center, IFAS, University of Florida, Fort Pierce, Florida, USA

[7]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

[8]Department of Aquatic Ecology, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

[9]Center for Biodiversity and Global Change, EEB Department, Yale University, New Haven, Connecticut, USA

[10]Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Sweden

[11]Department of Animal Biology, Federal University of Technology, Minna, Nigeria

[12]Finnish Environment Institute, Freshwater Centre, Oulu, Finland

[13]Centro de Investigacion en Ecosistemas de la Patagonia, CIEP, Coyhaique, Chile

[14]Universidade Federal do Maranhão, Codó, Brazil

[15]Laboratório de Ecologia e Conservação, Universidade Federal do Pará, Belém, Brazil

[16]Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona (UB), Barcelona, Spain

[17]Department of Biological Applications and Technologies, University of Ioannina, Ioannina, Greece

[18]CIEMEP (CONICET-UNPSJB), Esquel, Argentina

[19]Facultad de Ciencias Naturales y Ciencias de la Salud, Universidad Nacional de la Patagonia San Juan Bosco, Esquel, Argentina

[20]Lancaster Environment Centre, Lancaster University, Lancaster, UK

[21]State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China

[22]University of Chinese Academy of Sciences, Beijing, China

[23]Departamento de Genética, Ecologia e Evolução, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

[24]Department of Entomology, Nanjing Agricultural University, Nanjing, China

[25]State Key Laboratory of Marine Resource Utilization in South China Sea, Hainan University, Haikou, China

[26]Laboratório de Biodiversidade e Conservação, Universidade Federal de Rondônia – UNIR, Rolim de Moura, Brazil

[27]Centre d'Ecologie Fonctionnelle et Evolutive, Université de Montpellier, UMR 5175, CNRS, Université Paul Valéry Montpellier, EPHE, IRD, Montpellier, France

[28]Institute of Natural Resources – Ecology, Massey University, Palmerston North, New Zealand

[29]Senckenberg Research Institute and Natural History Museum, Frankfurt, Germany

[30]Instituto de Biodiversidad Neotropical- CONICET, Facultad de Ciencias Naturales, Universidad Nacional de Tucuman, Yerba Buena, Argentina

[31]Division of Ecology & Biodiversity, School of Biological Sciences, The University of Hong Kong, Hong Kong, China

[32]Norwegian Institute for Water Research, Oslo, Norway

[33]Department Life Sciences, FCTUC, Marine and Environmental Sciences Centre, Associate Laboratory ARNET, University of Coimbra, Coimbra, Portugal

[34]Department of Biodiversity and Environmental Management, University of León, León, Spain

[35]Univ Lyon, Université Claude Bernard Lyon 1, CNRS, ENTPE, UMR 5023 LEHNA, Villeurbanne, France

[36]Geography Research Unit, University of Oulu, Oulu, Finland

[37]University of Natural Resources and Life Sciences, Vienna, Austria

[38]Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

[39]Faculty of Biology, University of Duisburg-Essen, Essen, Germany

[40]Coordenação de Biodiversidade, Instituto Nacional de Pesquisas da Amazônia, Manaus, Brazil

[41]Department of Entomology, University of Minnesota, St Paul, Minnesota, USA

[42]Ecology and Genetics Research Unit, University of Oulu, Oulu, Finland

[43]Freshwater Biological Section, Department of Biology, University of Copenhagen, Copenhagen, Denmark

[44]Geography Department, Humboldt-Universität zu Berlin, Berlin, Germany

[45]Naturalis Biodiversity Center, Leiden, The Netherlands

[46]Universidade Federal da Paraíba – UFPB, João Pessoa, Brazil

[47]Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, Brazil

[48][NA], Bonn, Germany

[49]Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark

[50]OVAM, Mechelen, Belgium

[51]The Academy of Natural Sciences of Drexel University, Philadelphia, Pennsylvania, USA

[52]Museum of Victoria, Melbourne, Victoria, Australia

[53]Departamento de Ecologia – IB, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

[54]Independent Scholar

[55]Programa de Pós Graduação em Zoologia, Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, Brazil

[56]Departamento de Biologia Animal, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Campinas, São Paulo, Brazil

[57]Slater Museum of Natural History, University of Puget Sound, Tacoma, Washington State, USA

[58]Facultad de Ingenierías y Ciencias Aplicadas, Grupo de Investigación en Biodiversidad, Medio Ambiente y Salud (BIOMAS), Universidad de Las Américas-Ecuador, Quito, Ecuador

[59]Departamento de Ciências Exatas e Tecnológicas, Universidade Estadual do Sudoeste da Bahia, Vitória da Conquista, Brazil

[60]Institute of BioScience, Universidade Federal de Mato Grosso do Sul, Mato Grosso do Sul, Brazil

[61]Departamento de Biología y Geografía, Facultad de Ciencias Naturales, Universidad de Oriente, Santiago de Cuba, Cuba

[62]Balaton Limnological Research Institute, Tihany, Hungary

[63]Central Department of Environmental Science, Tribhuvan University, Kirtipur, Nepal

[64]Department of Water Resources and Ecosystems, IHE Delft Institute for Water Education, Delft, The Netherlands

[65]Institute of Biosciences, São Paulo State University (UNESP), Rio Claro, Brazil

[66]School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

[67]Department of Life Sciences and Aquatic Ecology Centre, Kathmandu University, Dhulikhel, Nepal

[68]Australian Museum, Darlinghurst, New South Wales, Australia

[69]Centre for Applied Water Science, University of Canberra, Canberra, Australian Capital Territory, Australia

[70]Te Pūnaha Matatini Centre of Research Excellence, University of Canterbury, Christchurch, New Zealand

[71]Bioprotection Aotearoa Centre of Research Excellence, University of Canterbury, Christchurch, New Zealand

[72]Department of Zoology, University of Otago, Dunedin, New Zealand

[73]Department of Planning and Environment, NSW Government, Parramatta, New South Wales, Australia

[74]California Department of Water Resources, West Sacramento, California, USA

[75]Altai State University, Barnaul, Russia

[76]Departamento de Biología Animal, Universidad de Granada, Granada, Spain

Correspondence

Afroditi Grigoropoulou and Sami Domisch, Department of Community and Ecosystem Ecology, Leibniz Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, Berlin 12587, Germany.

Email: afroditi.grigoropoulou@igb-berlin.de and sami.domisch@igb-berlin.de

Funding information

INCT ADAPTA II funded by CNPq, Grant/Award Number: 465540/2014-7; Amazonas State Research Foundation - FAPEAM, Grant/Award Number: 0621187/2017; ANID Programa Regional, Grant/Award Number: R20F0002; BIODIVERSA/FAPEAM; Universal/CNPq, Grant/Award Number: 403758/2021-1; Bundesministerium für Bildung und Forschung, Grant/Award Number: 033W034A; CAPES-Coordination for the Improvement of Higher Education Personnel; FAPEAM-Program POSGRAD; Fundação de Amparo à Pesquisa do Estado de Minas Gerais, Grant/Award Number: PPM 00104-18; INPA/ MCTI; Leibniz Competition, Grant/Award Number: J45/2018; Leibniz-Gemeinschaft; National Council of Development for Scientific and Technological Development - CNPq, Grant/Award Number: 304060/2020-8 and 308970/2019-5; Instituto Nacional de Pesquisas da Amazônia, unidade de

Abstract

Motivation: Aquatic insects comprise 64% of freshwater animal diversity and are widely used as bioindicators to assess water quality impairment and freshwater ecosystem health, as well as to test ecological hypotheses. Despite their importance, a comprehensive, global database of aquatic insect occurrences for mapping freshwater biodiversity in macroecological studies and applied freshwater research is missing. We aim to fill this gap and present the *Global EPTO Database*, which includes worldwide geo-referenced aquatic insect occurrence records for four major taxa groups: Ephemeroptera, Plecoptera, Trichoptera and Odonata (EPTO).

Main type of variables contained: A total of 8,368,467 occurrence records globally, of which 8,319,689 (99%) are publicly available. The records are attributed to the corresponding drainage basin and sub-catchment based on the Hydrography90m dataset and are accompanied by the elevation value, the freshwater ecoregion and the protection status of their location.

Spatial location and grain: The database covers the global extent, with 86% of the observation records having coordinates with at least four decimal digits (11.1 m precision at the equator) in the World Geodetic System 1984 (WGS84) coordinate reference system.

Time period and grain: Sampling years span from 1951 to 2021. Ninety-nine percent of the records have information on the year of the observation, 95% on the year and month, while 94% have a complete date. In the case of seven sub-datasets, exact dates can be retrieved upon communication with the data contributors.

pesquisa Ministério da Ciência, Tecnologia e Inovações (INPA/MCTI); Programa Peixe Vivo of the Companhia Energética de Minas Gerais, Grant/Award Number: P&D Aneel-Cemig GT 599, and GT 611; Rutherford Discovery Fellowship administered by the Royal Society Te Apārangi, Grant/Award Number: RDF-18-UOC-007; Tertiary Education Commission; São Paulo Research Foundation (FAPESP), Grant/Award Number: 2021/13299-0; Foundation for Science and Technology, Grant/Award Number: UIDB/04292/2020 and UIDP/04292/2020; Associate Laboratory ARNET, Grant/Award Number: LA/P/0069/2020; Principal Investigator CEEC

Handling Editor: Ines Martins

Major taxa and level of measurement: Ephemeroptera, Plecoptera, Trichoptera and Odonata, standardized at the genus taxonomic level. We provide species names for 7,727,980 (93%) records without further taxonomic verification.
Software format: The entire tab-separated value (.csv) database can be downloaded and visualized at https://glowabio.org/project/epto_database/. Fifty individual datasets are also available at https://fred.igb-berlin.de, while six datasets have restricted access. For the latter, we share metadata and the contact details of the authors.

KEYWORDS
aquatic insects, biodiversity, Ephemeroptera, freshwater ecosystems, global dataset, observation records, Odonata, Plecoptera, species distributions, Trichoptera

1 | INTRODUCTION

Freshwater ecosystems account for less than 0.01% of the Earth's surface, yet they play a vital role in the survival of people and wildlife (Gleick, 1998). Besides supporting the various human needs for freshwater use, freshwater bodies consist of numerous habitats naturally exchanging organisms, matter and resources, while hosting approximately 10% of the world's biodiversity. However, persistent and ever-increasing anthropogenic disturbance has placed them among the most endangered habitats in an ongoing global freshwater biodiversity crisis (IPBES, 2019; Living Planet Report, 2022; Tickner et al., 2020). The need for mitigation and conservation efforts focusing on inland waters and their biodiversity is thus becoming increasingly urgent (Maasri et al., 2022; Reid et al., 2019; Tickner et al., 2020; van Rees et al., 2021).

Knowledge of the spatio-temporal distributions of aquatic insects is vital for assessing water quality and overall ecosystem health (Bonada et al., 2006; Eriksen et al., 2021). In particular, given that the aquatic life stages of mayflies (Ephemeroptera), stoneflies (Plecoptera), caddisflies (Trichoptera), as well as dragon- and damselflies (Odonata), hereafter referred to as EPTO, have different sensitivities to water pollution and habitat degradation (Schmidt-Kloiber & Hering, 2015), the occurrence of these orders in an ecosystem is commonly used as an indicator of overall water quality and, therefore, can also be considered relevant for conservation planning (Bonada et al., 2006; Eriksen et al., 2021). In addition, these groups are useful surrogates for whole macroinvertebrate assemblages in terms of species richness estimates (Brito et al., 2018; Martins et al., 2022). Many EPTO feed on algae and/or decaying organic matter, while some species are predators. Moreover, they are prey to other organisms, such as fish and birds. They occupy multiple positions in food webs, and through leaf litter decomposition, contribute to nutrient transport between habitats (Macadam & Stockan, 2015; Schmidt-Kloiber & Hering, 2015; Wallace & Webster, 1996). Consequently, changes in the diversity and abundances of aquatic insects affect both terrestrial and other aquatic organisms (Sullivan & Manning, 2019). Moreover, these insect orders highly contribute to freshwater taxonomic, phylogenetic and functional biodiversity as they are among the most diverse in terms of number of species and functional trait variation (Dijkstra et al., 2014). However, 33% of known EPTO species are characterized as threatened according to International Union for Conservation of Nature (IUCN) criteria due to global change, while the extinction rate (i.e., percent of species not observed in over 50 years) is estimated to be around 9% (see table 1 in Sánchez-Bayo & Wyckhuys, 2019). The conflicting evidence on the spatio-temporal patterns of their populations at the global scale (Jähnig et al., 2021; van Klink et al., 2020; van Klink et al., 2021) calls for additional efforts in assessing their distributions worldwide that could effectively guide the implementation of conservation strategies (Maasri et al., 2022).

An inventory of EPTO observation records at the global scale is of high importance, as it would enable further analyses on biodiversity patterns from both theoretical and applied perspectives (Collen et al., 2014). In previous years, considerable efforts have been made towards this aim with, for example, the Distribution Atlas of European Trichoptera (DAET; Neu et al., 2018; Schmidt-Kloiber et al., 2017), the CONUS database for the contiguous United States (Twardochleb et al., 2021) and Macro-MED for macroinvertebrates of mediterranean countries (Blanco-Garrido et al., 2013), as well as online repositories such as Odonata Central (https://www.odonatacentral.org/; Abbott, 2018) the Atlas of Living Australia (ALA; http://www.ala.org.au) and the Global Biodiversity Information Facility (GBIF.org, 2020). However, these are either focused on one taxonomic order (e.g., DAET for caddisflies, Odonata Central for dragon- and damselflies) or cover subcontinental and continental scales (e.g., DAET in Europe, CONUS in the USA, ALA in Australia).

The majority of studies on the macroecology of freshwater insects has focused on regional or continental scales (Altermatt et al., 2013; Boyero et al., 2014; Grigoropoulou et al., 2022; Heino, 2011; Kaelin & Altermatt, 2016; Simaika et al., 2013), while worldwide macroecological assessments and meta-analyses remain scarce (but see Heino et al., 2015). Moreover, global analyses have been mainly based on compiled data drawn from individual publications, constantly highlighting the need for more detailed information on invertebrate species distributions (Collen et al., 2014; Vinson & Hawkins, 2003). Species distribution models are now increasingly used to predict aquatic macroinvertebrate distributions at regional scales (Bush et al., 2014; Dias-Silva et al., 2021; Domisch et al., 2013; Kaelin & Altermatt, 2016; Nieto et al., 2017). These could greatly benefit from more occurrence records and larger-scale data, as manifested in Sánchez-Fernández et al. (2011) over a decade ago, not least to decrease geographic and environmental sampling biases. Overall, the *Global EPTO Database* aims to fill these gaps, assisting future studies to disentangle broad-scale spatial patterns and determinants of insect distributions and detect biodiversity hot- and coldspots across the world.

2 | METHODS

2.1 | Data acquisition

The *Global EPTO Database* is a comprehensive compilation of heterogeneous data sources. As well as the 93 co-authors who contributed with occurrence records via personal communication up to the end of 2021, our database has also been supported by many researchers and taxonomists who contributed to parts of databases, such as the DAET (Neu et al., 2018; Schmidt-Kloiber et al., 2017), the CONUS database covering the contiguous USA (Twardochleb et al., 2021), Mayflies of South America (Dos Santos et al., 2018; Molineri et al., 2020), and Macro-MED (Blanco-Garrido et al., 2013), as well as the public databases GBIF (2020) and ALA (http://www.ala.org.au). Overall, we collated an initial number of 10,069,801 occurrence records, comprising any georeferenced data on EPTO records that had been made available until December 2021.

2.2 | Quality control

2.2.1 | Taxonomy

We standardized the database at the genus level. We validated and complemented the taxonomic classification of the occurrence records using the package "taxize" v.0.9.99 in R (R Core Team, 2020), which interacts with the Integrated Taxonomic Information System (ITIS) and the National Center for Biotechnology Information (NCBI) databases to verify genus names and correct spelling, as well as to obtain taxonomic hierarchies (Chamberlain & Szöcs, 2013). For cases in which names were not found in these databases because of misspellings or database deficiencies, we manually verified them through online sources, such as the Catalogue of Life (Bánki et al., 2022), Zootaxa (http://www.biotaxa.org), the Encyclopedia of Life (https://eol.org/), and primary publications. We also provide the binomial nomenclature regarding species names, available for 7,727,980 (93%) records, without further taxonomic verification because of heterogenous country-level checklists. Overall, we discarded 837,603 records of non-EPTO orders and records lacking genus names.

2.2.2 | Dates

We included records taken between 1951 and 2021 in the database and harmonized the sampling dates to the yyyymmdd format in R, split into separate columns of year, month and day. We removed records with a date format that impeded the differentiation between month and day (e.g., 03082000) or with a year later than 2021. Additionally, we removed records lacking year information, with the exception of the dataset "Mayflies of South America" (Nieto et al., 2022), for which we kept the year range 1970–2019. In total, we discarded 331,585 records with inconsistent date information.

2.2.3 | Coordinate treatment

We transformed all coordinates to the World Geodetic System 1984 (WGS84) coordinate reference system using the package "sp" in R (Bivand et al., 2013). We approximated the spatial precision of records using the number of decimal digits of the coordinates. We counted the decimal digits using the function "format.info" in R, taking 10 digits into account. Subsequently, we evaluated the validity of the records' coordinates through the package "CoordinateCleaner" by using geographic gazetteers to identify possibly erroneous sets of coordinates or imprecise records, which are common in biological collections (Zizka et al., 2019). We discarded records with zero or equal latitude and longitude coordinate attributes and coordinates assigned to the location of biodiversity institutions or universities, instead of exact sampling locations. We fixed swapped coordinates manually. In contrast, we flagged coordinates that were assigned to country centroids, capitals as well as cases of coordinate–country mismatches with the value "1" in the corresponding columns cen_fl (centroid), cap_fl (capital), coun_fl (country), else we assigned the value "0" in these columns. Finally, we treated coordinates located in the sea following the procedure described in Section 2.3. Overall, we discarded 504,607 records with invalid coordinates.

2.3 | Integration with the Hydrography90m dataset

We attributed the EPTO occurrence records to their corresponding hydrographic units at two spatial scales, namely the drainage basin and sub-catchment, as delineated in the Hydrography90m dataset (https://hydrography.org/hydrography90m/hydrography90m_layers/; Amatulli et al., 2022). In Hydrography90m, a drainage basin is defined as any area of land where precipitation is collected and drained into a common outlet, either into the sea or into an inland depression. Further, every basin is divided into smaller sub-catchments, each of them representing the land area between two stream segment nodes that contributes to the local flow accumulation of a given stream segment (Amatulli et al., 2022). Every drainage basin and sub-catchment worldwide holds a separate, unique ID. By combining the *Global EPTO Database* with the

Hydrography90m dataset, we aimed to yield global statistics regarding the number of observation records and unique genera per drainage basin.

Moreover, we acquired the elevation value of every record's location based on the Multi-Error-Removed Improved-Terrain DEM (MERIT) Hydro digital elevation model (DEM), which is the most accurate, globally seamless, 3 arc-second resolution (~90 m at the equator) DEM to date (Yamazaki et al., 2019).

For cases in which sea coordinates were located within a 10-km radius off the coast or of an island, we attributed/snapped them to the closest sub-catchment using the function “r.stream.snap” in Geographic Resources Analysis Support System (GRASS GIS) version 7.8 (GRASS Development Team, 2008), or else we discarded the record. We snapped 62,498 points in total.

2.4 | Integration with the Freshwater Ecoregions of the World and the World Protected Areas datasets

We assigned the ID of the corresponding freshwater ecoregion to each record using the Freshwater Ecoregions of the World dataset (FEOW, Abell et al., 2008), after rasterizing the layer to a 90-m resolution using the function “gdal_rasterize” (GDAL/OGR Contributors, 2021). This allows quick filtering of the records based on the biogeographic realm. We included the IDs of all the freshwater ecoregions present in each dataset in its metadata file.

Moreover, we included information regarding the overlap of the point record and protected areas using the latest layer (November 2022) of the Protected Planet WDPA WD-OECM dataset [Protected Planet: World Database on Protected Areas (WDPA) and World Database on Other Effective Area-based Conservation Measures (WD-OECM), UNEP-WCMC and IUCN, 2022]. We downloaded the layer using the “wdpa_fetch” function of the “wdpar” package in R (Hanson, 2022) and filtered the global file based on the protection status, using the function “ogr2ogr” (GDAL/OGR contributors, 2021). We retained “Designated”, “Inscribed” and “Established” protected areas to ensure that those that are not currently implemented are excluded (e.g., proposed protected areas). We rasterized the layer to a 90-m resolution and extracted the information of the “Status Year” (STATUS_YR) field, which identifies the year in which a site's current status was implemented. We assigned each record a value of “0” if it was not located in a protected area, “1” if it was in a protected site on the date of the observation, or “2”, in the case that the location was protected after the observation event (Table 1, “WDPA_protection_status” column). Information regarding the past protection status of an area was not available.

We used the function “gdallocationinfo” in the Geospatial Data Abstraction Library (GDAL), version number 3.1.0 (GDAL/OGR contributors, 2021) for all spatial extractions at point level at a global extent.

3 | RESULTS AND DISCUSSION

3.1 | Database formatting

The database comprises the entire table including 8,319,689 publicly available WGS84 geo-referenced records of EPTO genera globally, from 50 individual datasets, for download at https://doi.org/10.18728/igb-fred-811.0 and https://glowabio.org/project/epto_database/. In addition, we provide the metadata and authors' contact details for six datasets that have restricted access, comprising 48,728 records, which can be found in the metadata file (Supporting Information Table S1).

All data are organized in tables (i.e., tab separated .csv format; see Supporting Information Table S1 for the URLs to the individual datasets). Within the dataset tables, each row corresponds to one observation record defined as the presence of an individual of a given genus observed at a particular location on a given date, while the 28 columns include a variety of taxonomic, geographic and additional site-specific information (Table 1). Duplicate observation records, possibly derived from overlapping data sources, were collapsed to one. The metadata of each dataset include information on the availability of abundances derived from the sampling itself and site-specific information regarding physicochemical, hydrological or land cover characteristics of the sampling location.

3.2 | Database completeness

3.2.1 | Spatial coverage and precision

The *Global EPTO Database* includes 8,368,467 occurrence records, covering 1,505,302 unique locations worldwide and 39,584 out of 1.6 million drainage basins of the Hydrography90m dataset (Table 2; Amatulli et al., 2022). The records are spread across 417 out of 426 Freshwater Ecoregions of the World (FEOW, Abell et al., 2008), while 20% of them occur in protected areas. We mapped the number of observations of the whole database on a 10-km regular grid and per drainage basin (Figure 1a,c, respectively) and the numbers of observation records of each order individually per drainage basin (Figure 1e). The spatial patterns arising from the maps of Figure 1a and the respective strong peaks in the marginal plots reveal a major sampling bias towards North America, central Europe, and Oceania.

To further explore the spatial distribution of the records, we plotted the number of observation records within elevational bands of 500 m (Figure 2a). Records cover an elevation range of more than 4,000 m, although the majority is concentrated in lower altitudes of up to 500 m. The distribution of the data across the elevational gradient follows the global pattern of declining species richness as altitude increases (Gaston, 2000). However, there is contrasting evidence suggesting that taxon richness might well increase along an elevation gradient (Chiu et al., 2020). Therefore, increasing sampling effort at higher altitudes is crucial, especially because headwater

TABLE 1 Column names of the individual dataset tables available in the *Global EPTO Database* and their explanations.

Column name	Column content
order_ID	One of the E, P, T, O
dataset_ID	Unique ID of the dataset of a given data source, e.g., 003
occurrence_ID	Unique ID of the individual occurrence record, e.g., 000423769
species	Species name (not taxonomically verified)
genus	Genus name
family	Family name
order	Order name
longitude	Longitude in decimal degrees, in WGS84
latitude	Latitude in decimal degrees, in WGS84
no_decimals_published	Number of decimal points of the coordinates that are published
no_decimals_available	Number of decimal points of the coordinates that are available upon communication with the authors (could include more detailed coordinates)
location_river	Locality or river of the observation record
country	Country of the observation record
sampling_day	Day of the observation
month	Month of the observation
year	Year of the observation
source	Data source (e.g., name or DOI of the original database)
FRED_link-DOI	Dataset link to FRED, where all metadata are available, including dataset authors' names
notes	Dataset authors' notes
cen_fl	Centroid flag, Boolean. If "1", then the coordinates represent the centroid of a country instead of the exact location of the observation
cap_fl	Capital flag, Boolean. If "1", then the coordinates belong to the capital of a country instead of the exact location of the observation
coun_fl	Country flag, Boolean. If "1", then the coordinates of the record do not match with the assigned country
snapped	Boolean. If "1", then the original coordinates of the record, that fell into the sea, have been snapped to the closest sub-catchment
subcatchment_Idy9	Unique ID of the sub-catchment where the record is located, based on the Hydrography90m dataset (Amatulli et al., 2022)
basin_ID	Unique ID of the drainage basin where the record is located, based on the Hydrography90m dataset (Amatulli et al., 2022)
elevation	Elevation of the point of the record, in metres, based on the MERIT Hydro DEM (Yamazaki et al., 2019)
WDPA_protection_status	Status of the record's location based on the Protected Planet WDPA WD-OECM dataset, updated in November 2022 (UNEP-WCMC and IUCN, 2022). If "0", the record was not located in a protected area. If "1", the record was in a protected site. If "2", the record's location was protected on a date following the observation
FEOW_ID	Freshwater Ecoregions of the World ID (Abell et al., 2008)

Note: DEM = digital elevation model; EPTO = Ephemeroptera, Plecoptera, Trichoptera and Odonata; FRED = IGB Freshwater Research and Environmental Database; MERIT = Multi-Error-Removed Improved-Terrain DEM; WD-OECM = World Database on Other Effective Area-based Conservation Measures; WDPA = World Database on Protected Areas; WGS84 = World Geodetic System 1984. The FRED link of each specific dataset is only included in the comprehensive table of all orders.

streams and springs harbour high levels of biodiversity (Clarke et al., 2006; Heino et al., 2013).

3.2.2 | Taxonomic representation

The observation records include 1,473 genera and 142 families (Table 2). Sixty-two percent of all occurrence records correspond to observations of Odonata, which is probably related to their being considered as charismatic, not least because their specimens are easily observable due to their size, colour patterns, and flying style (Vorster et al., 2020). The database includes more than 15 records for most genera of each order for which we have occurrence data (216 genera of Ephemeroptera out of 310 have at least 15 records, Plecoptera: 138/199, Trichoptera: 281/391, Odonata: 422/573 genera, Figure 2b). Although the number of presence observations

TABLE 2 Summary statistics of the *Global EPTO Database* for the number of records, drainage basins and sub-catchments, the geographic location and precision as well as the temporal precision of the samplings.

	Ephemeroptera	Plecoptera	Trichoptera	Odonata	EPTO
Observation records	933,070	538,236	1,640,723	5,256,438	8,368,467
Genera	310	199	391	573	1,473
Described genera	442	313	632	694	2,081
Families	41	17	46	38	142
Described families	42	17	63	45	167
Occurrence locations	114,711	58,939	164,985	1,328,255	1,505,302
Sub-catchments covered	89,521	42,182	129,658	593,135	702,409
Drainage basins covered	5,220	3,318	8,343	36,623	39,584
% of records in protected areas	12	10	16	22	20
% of coordinates with ≥ 4 decimal digits	95	95	91	82	86
% of records including sampling year	99	99	99	100	99
% of records including sampling month	93	95	88	97	95
% of records including sampling day	92	94	86	96	94

Note: Information on the numbers of the described extant genera and families of Ephemeroptera, Plecoptera, Trichoptera, and Odonata (EPTO) was retrieved from Barber-James et al. (2013), DeWalt et al. (2021), Morse (2022) and Paulson et al. (2022), respectively.

required for building robust predictive models, such as species distribution models (SDMs), may vary depending on the prevalence and spatial representativeness of each taxon (van Proosdij et al., 2016), it has been proposed that 15 presence observations may be enough to capture the general response of species to important environmental gradients (Mateo et al., 2010; Støa et al., 2019). Following this guideline, the *Global EPTO Database* provides enough occurrences for modelling spatial patterns and their underlying mechanisms of most EPTO genera.

We mapped the number of EPTO genera for the whole database and for each order individually in a similar way as for the number of occurrence records (Figure 1b,f). The apparent lack of congruence between the number of observations and the number of genera per basin (Figure 1a,c,e and b,d,f, respectively) highlights the fact that some of the richest areas in terms of taxonomic diversity are being overlooked regarding sampling and/or data digitalization/mobilization (e.g., the Amazon basin, or basins in Africa and south-western Asia). Likewise, due to the spatial sampling bias, genera from under-sampled areas may be missing (Figure 1, grey areas in the maps).

3.2.3 | Time frame

All records are accompanied by information on the year of observation, with 95 and 94% also having information on the month and day of the observation, respectively (Table 2). The dated occurrences can serve in defining EPTO distributions at a given time using species distribution modelling and therefore serve in conservation planning. The increase in record numbers after the year 2000 follows the development of GIS technology and data digitization (Figure 2c). The steeper increase in the number of records for Odonata is due to a strong increase in the number of observations submitted by citizen scientists, especially in North America and western Europe. Given the lack of a consistent sampling effort through time, the database is generally unsuitable for investigating temporal trends at the global scale. However, we encourage users to explore the data according to their needs, as this trend might change at local scales in the best sampled areas.

4 | CONCLUSIONS

By integrating heterogeneous data sources of EPTO occurrence records, we attempted to respond to the priority need for a comprehensive compilation of biological data, as recently suggested by Maasri et al. (2022). As a result, the *Global EPTO Database* is easily accessible and developed according to the FAIR principles regarding Findability, Accessibility, Interoperability and Reusability of data (Wilkinson, 2016). At the same time, it highlights the spatial patterns of sampling efforts and/or data digitization and mobilization worldwide, which are mostly concentrated in the Northern Hemisphere, in accordance with the findings of Meyer et al. (2015) for vertebrate species and Murphy et al. (2019) for aquatic plants. It has been shown that gaps in inventories exist in most countries of the Global South, while species data are mainly limited by distance to researchers, available research funding for a given region and participation in data-sharing networks (Meyer et al., 2015). Moreover, the notably higher number of Odonata observations, in contrast to EPT orders, leads to a taxonomic bias in available inventories. These geographic or taxonomic gaps need to be tackled by guiding future sampling campaigns towards under-represented areas, opening up previously inaccessible repositories and enhancing taxonomic research, especially on EPT (Balian et al., 2008).

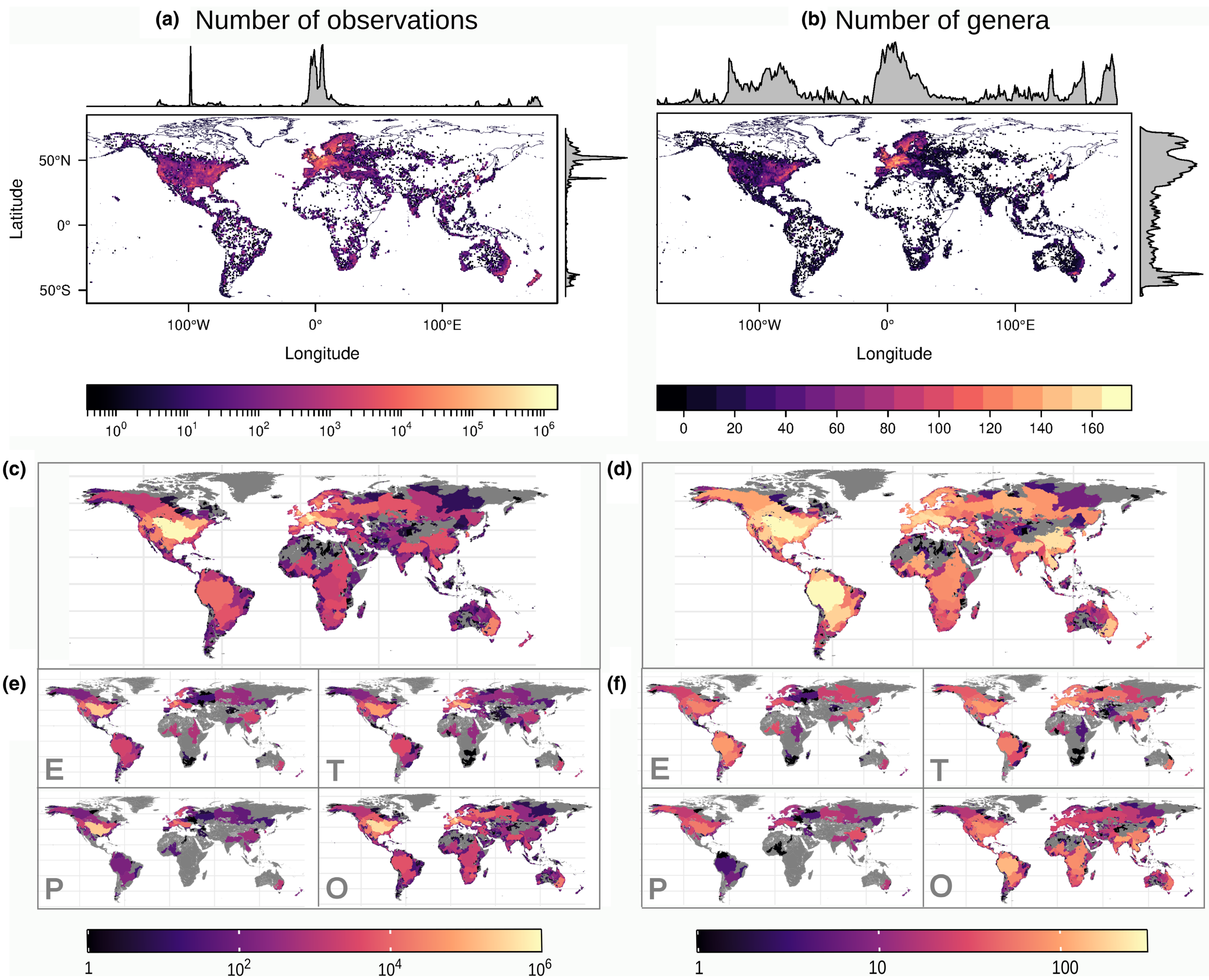

FIGURE 1 Number of observation records (a) and genera (b) for all orders, aggregated in a 10-km regular grid. Number of observation records (c) and genera (d) for all orders per drainage basin. Number of observation records (e) and genera (f) per drainage basin for each order individually. E = Ephemeroptera; P = Plecoptera; T = Trichoptera; O = Odonata. White and grey areas lack information. A more detailed interactive view of the maps (e) and (f) is available at https://glowabio.org/project/epto_database/.

In the light of the ongoing global freshwater biodiversity decline (Reid et al., 2019), mobilizing and openly sharing existing data will be crucial for advancing biodiversity and conservation research. More detailed species level and trait information (Sarremejane et al., 2020; Schmidt-Kloiber & Hering, 2015) is required for more in-depth analyses and informed decision-making (Maasri et al., 2022). The *Global EPTO Database* in combination with the Hydrography90m dataset aim to provide a solid, open-source information base for freshwater macroecological analyses and species distribution modelling on a high spatial resolution.

ACKNOWLEDGMENTS

The *Global EPTO Database* is part of the Global Freshwater Biodiversity, Biogeography and Conservation project (https://glowabio.org), and AG, JRGM and SD acknowledge funding by the Leibniz Competition to SD (J45/2018). We acknowledge funding by the German Federal Ministry of Education and Research (BMBF grant agreement number 033W034A). We thank the Yale Center for Research Computing for the use of research computing infrastructure. We also sincerely thank Daniel Langenhaun and Thomas Tomiczek for their help in uploading the files to https://fred.igb-berlin.de and https://geo.igb-berlin.de. Open Access funding enabled and organized by Projekt DEAL.

NH received research grants from National Council of Development for Scientific and Technological Development – CNPq (308970/2019-5) and Instituto Nacional de Pesquisas da Amazônia, unidade de pesquisa Ministério da Ciência, Tecnologia e Inovações (INPA/MCTI). Collecting campaign of Aquatic Insects – LACIA/INPA dataset was supported in part by INCT ADAPTA II funded by CNPq (465540/2014-7), Amazonas State Research Foundation – FAPEAM (0621187/2017), CAPES – Coordination for the Improvement of Higher Education Personnel, FAPEAM-Program POSGRAD, BIODIVERSA/FAPEAM, and Universal/CNPq (403758/2021-1). RTM received a fellowship from BIODIVERSA/FAPEAM.

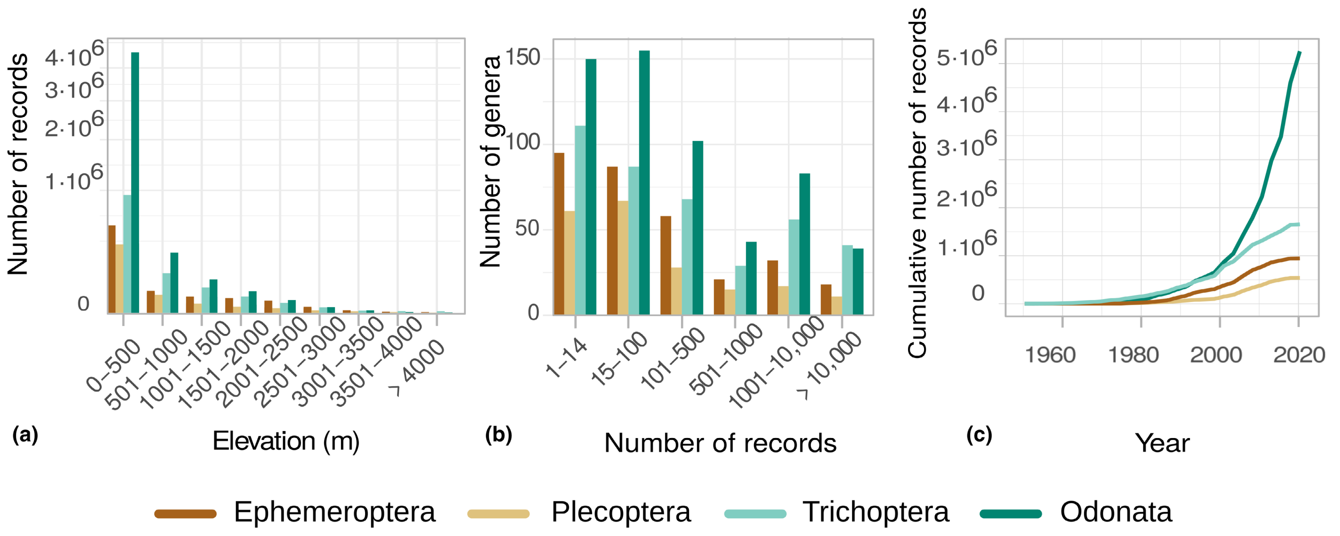

FIGURE 2 (a) Number of observation records per 500-m elevational band. (b) Number of genera and the number of their observation records in the database. (c) Cumulative number of observation records of each order over the last 70 years (only data with exact sampling years included).

FOR received research grants from National Council of Development for Scientific and Technological Development – CNPq, Brazil.

FVN was supported by grant #2021/13299-0, São Paulo Research Foundation (FAPESP).

SCJ acknowledges funding through the Leibniz Association for the project Freshwater Megafauna Futures.

JDT is supported by a Rutherford Discovery Fellowship administered by the Royal Society Te Apārangi (RDF-18-UOC-007), and Bioprotection Aotearoa and Te Pūnaha Matatini, both Centres of Research Excellence funded by the Tertiary Education Commission, New Zealand.

AA-R acknowledges funding from the programme ANID Programa Regional R20F0002.

MC is grateful for continued funding by Programa Peixe Vivo of the Companhia Energética de Minas Gerais (P&D Aneel-Cemig GT 599, and GT 611), Coordination for the Improvement of Higher Education Personnel (CAPES), National Council for Scientific & Technological Development (CNPq) research productivity grant 304060/2020-8, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (grant PPM 00104-18).

MJF acknowledges the financial support of the Foundation for Science and Technology through the projects UIDB/04292/2020, UIDP/04292/2020 granted to MARE, and LA/P/0069/2020, granted to the Associate Laboratory ARNET and the Principal Investigator CEEC contract granted to MJF.

We thank the PIs Jon Gelhaus, Barbara Hayford, Riley Nielson and John Morse, and Boloroo Erdenee, Suvdtsetseg Chuluunbat and Oyunchuluun Yadamsuren from Mongolia, who were instrumental in getting the Mongolian Aquatic Insect Survey data sampled and identified.

AS-K, PJN and WG sincerely thank all the caddisfly data providers who contributed to the Distribution Atlas of European Trichoptera and are listed here: https://doi.org/10.15504/fmj.2015.9.

We dedicate this work to the memory of Carmen Zamora-Muñoz and thank her for her support and enthusiasm for this project.

CONFLICT OF INTEREST STATEMENT

There are no conflicts of interest.

DATA AVAILABILITY STATEMENT

The entire tab-separated value (.csv) database can be downloaded and visualized at https://glowabio.org/project/epto_database/ and https://doi.org/10.18728/igb-fred-811.0. The URLs to the individual dataset entries can be found in the column "FRED_link-DOI" of Supporting Information Table S1. For dataset visualization, a link to https://geo.igb-berlin.de/ can be found in the field "GeoNode references" in the FRED entry of each dataset. Fifty of the datasets are openly available for download while for six datasets we share the metadata and contact details of their authors (see Supporting Information Table S1). The R code used for data cleaning is available at https://gitlab.com/afrogri37/global_aquatic_insect_data. We welcome further data contributions for future versions of the database.

We recommend using the function "fread" of the "data.table" package (Dowle & Srinivasan, 2022) to load the table in R (e.g., data <- fread("Global_EPTO_Database.csv")).

ORCID

Afroditi Grigoropoulou https://orcid.org/0000-0002-7884-097X
Suhaila Ab Hamid https://orcid.org/0000-0003-0932-3677
Raúl Acosta https://orcid.org/0000-0003-0923-8237
Emmanuel Olusegun Akindele https://orcid.org/0000-0002-4911-8857
Salman A. Al-Shami https://orcid.org/0000-0001-9134-4172
Florian Altermatt https://orcid.org/0000-0002-4831-6958
Giuseppe Amatulli https://orcid.org/0000-0002-8341-2830
David G. Angeler https://orcid.org/0000-0003-2197-7470

Francis O. Arimoro https://orcid.org/0000-0001-6100-4011
Jukka Aroviita https://orcid.org/0000-0003-3330-0731
Anna Astorga-Roine https://orcid.org/0000-0003-4600-4384
Rafael Costa Bastos https://orcid.org/0000-0001-7581-7099
Núria Bonada https://orcid.org/0000-0002-2983-3335
Cecilia Brand https://orcid.org/0000-0002-6611-5033
Vanessa Bremerich https://orcid.org/0000-0002-7657-1534
Alex Bush https://orcid.org/0000-0002-0679-6666
Qinghua Cai https://orcid.org/0000-0002-0224-7256
Marcos Callisto https://orcid.org/0000-0003-2341-4700
Kai Chen https://orcid.org/0000-0002-5332-5920
Paulo Vilela Cruz https://orcid.org/0000-0003-2426-4628
Xiling Deng https://orcid.org/0000-0002-0806-1428
Eduardo Domínguez https://orcid.org/0000-0002-4201-7869
David Dudgeon https://orcid.org/0000-0003-4632-3473
Tor Erik Eriksen https://orcid.org/0000-0003-2033-6399
Ana Paula J. Faria https://orcid.org/0000-0003-2729-5358
Maria João Feio https://orcid.org/0000-0003-0362-6802
Camino Fernández-Aláez https://orcid.org/0000-0001-9385-1354
Mathieu Floury https://orcid.org/0000-0002-4952-5807
Francisco García-Criado https://orcid.org/0000-0003-3419-7086
Jorge García-Girón https://orcid.org/0000-0003-0512-3088
Wolfram Graf https://orcid.org/0000-0001-6559-0644
Mira Grönroos https://orcid.org/0000-0002-8210-8837
Peter Haase https://orcid.org/0000-0002-9340-0438
Neusa Hamada https://orcid.org/0000-0002-3526-5426
Fengzhi He https://orcid.org/0000-0002-7594-8205
Jani Heino https://orcid.org/0000-0003-1235-6613
Ralph Holzenthal https://orcid.org/0000-0003-1853-6340
Kaisa-Leena Huttunen https://orcid.org/0000-0003-0488-1274
Dean Jacobsen https://orcid.org/0000-0001-5137-297X
Sonja C. Jähnig https://orcid.org/0000-0002-6349-9561
Walter Jetz https://orcid.org/0000-0002-1971-7277
Richard K.Johnson https://orcid.org/0000-0001-7979-6563
Leandro Juen https://orcid.org/0000-0002-6188-4386
Vincent Kalkman https://orcid.org/0000-0002-1484-7865
Vassiliki Kati https://orcid.org/0000-0003-3357-4556
Unique N. Keke https://orcid.org/0000-0002-9528-0599
Ricardo Koroiva https://orcid.org/0000-0002-6658-0824
Mathias Kuemmerlen https://orcid.org/0000-0003-1362-3701
Simone Daniela Langhans https://orcid.org/0000-0001-9581-3183
Raphael Ligeiro https://orcid.org/0000-0001-9717-5461
Kris Van Looy https://orcid.org/0000-0001-5436-2320
Alain Maasri https://orcid.org/0000-0003-1236-8374
Richard Marchant https://orcid.org/0000-0001-7387-2609
Renato T. Martins https://orcid.org/0000-0003-3464-7905
Adriano S. Melo https://orcid.org/0000-0002-4695-2854
Maria Laura Miserendino https://orcid.org/0000-0002-8414-0734
S. Jannicke Moe https://orcid.org/0000-0002-3681-3551
Carlos Molineri https://orcid.org/0000-0003-2662-624X
Timo Muotka https://orcid.org/0000-0002-2268-5683
Kaisa-Riikka Mustonen https://orcid.org/0000-0003-3717-0911
Heikki Mykrä https://orcid.org/0000-0001-6576-5563
Jeane Marcelle Cavalcante do Nascimento https://orcid.org/0000-0002-5428-7495
Francisco Valente-Neto https://orcid.org/0000-0002-5298-3753
Carolina Nieto https://orcid.org/0000-0003-1909-1750
Steffen U. Pauls https://orcid.org/0000-0002-6451-3425
Blanca Rios-Touma https://orcid.org/0000-0002-3921-0908
Marciel Elio Rodrigues https://orcid.org/0000-0001-8161-6234
Fabio de Oliveira Roque https://orcid.org/0000-0001-5635-0622
Juan Carlos Salazar Salina https://orcid.org/0000-0003-3796-8691
Dénes Schmera https://orcid.org/0000-0003-1248-8413
Astrid Schmidt-Kloiber https://orcid.org/0000-0001-8839-5913
Deep Narayan Shah https://orcid.org/0000-0001-8436-7560
John P. Simaika https://orcid.org/0000-0002-8073-2804
Tadeu Siqueira https://orcid.org/0000-0001-5069-2904
Günther Theischinger https://orcid.org/0000-0002-5207-2626
Ross Thompson https://orcid.org/0000-0002-5287-2455
Jonathan D. Tonkin https://orcid.org/0000-0002-6053-291X
Yusdiel Torres-Cambas https://orcid.org/0000-0003-2312-2329
Eren Turak https://orcid.org/0000-0001-7383-9112
Laura Twardochleb https://orcid.org/0000-0002-8804-9399
Beixin Wang https://orcid.org/0000-0002-5253-8799
Liubov Yanygina https://orcid.org/0000-0001-6738-2769
Sami Domisch https://orcid.org/0000-0002-8127-9335

REFERENCES

Abbott, J. (2018). *An online resource for the distribution and identification of Odonata*. Odonata Central. https://www.odonatacentral.org/

Abell, R., Thieme, M. L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., Coad, B., Mandrak, N., Balderas, S. C., Bussing, W., Stiassny, M. L., Skelton, P., Allen, G. R., Unmack, P., Naseka, A., Ng, R., Sindorf, N., Robertson, J., Armijo, E., … Petry, P. (2008). Freshwater ecoregions of the world: A new map of biogeographic units for freshwater biodiversity conservation. *Bioscience*, *58*(5), 403–414.

Altermatt, F., Seymour, M., & Martinez, N. (2013). River network properties shape α-diversity and community similarity patterns of aquatic insect communities across major drainage basins. *Journal of Biogeography*, *40*(12), 2249–2260. https://doi.org/10.1111/jbi.12178

Amatulli, G., Garcia Marquez, J., Sethi, T., Kiesel, J., Grigoropoulou, A., Üblacker, M. M., Shen, L. Q., & Domisch, S. (2022). Hydrography90m: A new high-resolution global hydrographic dataset. *Earth System Science Data*, *14*(10), 4525–4550.

Atlas of Living Australia. Retrieved April 28, 2022. http://www.ala.org.au

Balian, E. V., Segers, H., Lévèque, C., & Martens, K. (2008). The freshwater animal diversity assessment: An overview of the results. *Hydrobiologia*, *595*(1), 627–637. https://doi.org/10.1007/s10750-007-9246-3

Bánki, O., Roskov, Y., Döring, M., Ower, G., Vandepitte, L., Hobern, D., et al. (2022). *Catalogue of Life Checklist (Version 2022-04-26)*. Catalogue of Life. https://doi.org/10.48580/dfpk

Barber-James, H., Sartori, M., Gattolliat, J.-L., & Webb, J. (2013, February). World checklist of freshwater Ephemeroptera species. In O. Bánki, Y. Roskov, M. Döring, G. Ower, L. Vandepitte, D. Hobern, et al. (Eds.), *Catalogue of life checklist*. https://doi.org/10.48580/dfp3-3cn

Bivand, R. S., Pebesma, E., & Gómez-Rubio, V. (2013). *Applied spatial data analysis with R* (2nd ed.). Springer. https://doi.org/10.1007/978-1-4614-7618-4

Blanco-Garrido, F., Zamora-Muñoz, C., & Bonada, N. (2013). MacroMED: Macroinvertebrates in Mediterranean Climate Watercourses.

Bonada, N., Prat, N., Resh, V. H., & Statzner, B. (2006). Developments in aquatic insect biomonitoring: A comparative analysis of recent approaches. *Annual Review of Entomology*, *51*, 495–523. https://doi.org/10.1146/annurev.ento.51.110104.151124

Boyero, L., Ramirez, A., & Pearson, R. G. (2014). Gradients in regional diversity of freshwater taxa gradients in regional diversity of freshwater taxa. *Journal of the North American Benthological Society*, *28*(2), 504–514. https://doi.org/10.1899/08-118.1

Brito, J. G., Martins, R. T., Oliveira, V. C., Hamada, N., Nessimian, J. L., Hughes, R. M., Ferraz, S. F., & de Paula, F. R. (2018). Biological indicators of diversity in tropical streams: Congruence in the similarity of invertebrate assemblages. *Ecological Indicators*, *85*, 85–92. https://doi.org/10.1016/j.ecolind.2017.09.001

Bush, A. A., Nipperess, D. A., Duursma, D. E., Theischinger, G., Turak, E., & Hughes, L. (2014). Continental-scale assessment of risk to the Australian Odonata from climate change. *PLoS One*, *9*(2), e88958. https://doi.org/10.1371/journal.pone.0088958

Chamberlain, S. A., & Szöcs, E. (2013). Taxize: Taxonomic search and retrieval in R [version 2; peer review: 3 approved]. F1000Research, 2:191. https://doi.org/10.12688/f1000research.2-191.v2

Chiu, M. C., Ao, S., He, F., Resh, V. H., & Cai, Q. (2020). Elevation shapes biodiversity patterns through metacommunity-structuring processes. *Science of the Total Environment*, *743*, 140548. https://doi.org/10.1016/j.scitotenv.2020.140548

Clarke, R. T., Lorenz, A., Sandin, L., Schmidt-Kloiber, A., Strackbein, J., Kneebone, N. T., & Haase, P. (2006). Effects of sampling and subsampling variation using the STAR-AQEM sampling protocol on the precision of macroinvertebrate metrics. *Hydrobiologia*, *566*(1), 441–459. https://doi.org/10.1007/s10750-006-0078-3

Collen, B., Whitton, F., Dyer, E. E., Baillie, J. E. M., Cumberlidge, N., Darwall, W. R. T., Pollock, C., Richman, N. I., Soulsby, A. M., & Böhm, M. (2014). Global patterns of freshwater species diversity, threat and endemism. *Global Ecology and Biogeography*, *23*(1), 40–51. https://doi.org/10.1111/geb.12096

DeWalt, R. E., Hopkins, H., Maehr, M. D., Neu-Becker, U., & Stueber, G. (2021, August). Plecoptera species file. In O. Bánki, Y. Roskov, M. Döring, G. Ower, L. Vandepitte, D. Hobern, et al. (Eds.), *Catalogue of life checklist*. https://doi.org/10.48580/dfq8-39r

Dias-Silva, K., Vieira, T. B., Moreira, F. F. F., Juen, L., & Hamada, N. (2021). Protected areas are not effective for the conservation of freshwater insects in Brazil. *Scientific Reports*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-00700-0

Dijkstra, K. D. B., Monaghan, M. T., & Pauls, S. U. (2014). Freshwater biodiversity and aquatic insect diversification. *Annual Review of Entomology*, *59*(1), 143–163. https://doi.org/10.1146/annurev-ento-011613-161958

Domisch, S., Araújo, M. B., Bonada, N., Pauls, S. U., Jähnig, S. C., & Haase, P. (2013). Modelling distribution in European stream macroinvertebrates under future climates. *Global Change Biology*, *19*(3), 752–762. https://doi.org/10.1111/gcb.12107

Dos Santos, D. A., Molineri, C., Nieto, C., Zuñiga, M. C., Emmerich, D., Fierro, P., Pessacq, P., Rios-Touma, B., Márquez, J., Gomez, D., Salles, F. F., Encalada, A. C., Príncipe, R., Gómez, G. C., & Dominguez, E. (2018). Cold/warm stenothermic freshwater macroinvertebrates along altitudinal and latitudinal gradients in Western South America: A modern approach to an old hypothesis with updated data. *Journal of Biogeography*, *45*(7), 1571–1581.

Dowle, M., & Srinivasan, A. (2022). *data.table: Extension of 'data.frame'*. R package version 1.14.4. https://CRAN.R-project.org/package=data.table

Encyclopedia of Life. Retrieved July 15, 2018, http://eol.org

Eriksen, T. E., Brittain, J. E., Søli, G., Jacobsen, D., Goethals, P., & Friberg, N. (2021). A global perspective on the application of riverine macroinvertebrates as biological indicators in Africa, South-Central America, Mexico and Southern Asia. *Ecological Indicators*, *126*, 107609. https://doi.org/10.1016/j.ecolind.2021.107609

Gaston, K. J. (2000). Global patterns in biodiversity. *Nature*, *405*, 220–227. https://doi.org/10.1038/35012228

GBIF.org. (2020). GBIF home page. https://www.gbif.org

GDAL/OGR Contributors. (2021). *GDAL/OGR geospatial data abstraction software library*. Open Source Geospatial Foundation. https://doi.org/10.5281/zenodo.5884351

Gleick, P. H. (1998). The human right to water. *Water Policy*, *1*(5), 487–503. https://doi.org/10.1016/S1366-7017(99)00008-2

GRASS Development Team. (2008). Geographic Resources Analysis Support System (GRASS GIS) Software. http://grass.osgeo.org

Grigoropoulou, A., Schmidt-Kloiber, A., & Múrria, C. (2022). Incongruent latitudinal patterns of taxonomic, phylogenetic and functional diversity reveal different drivers of caddisfly community assembly across spatial scales. *Global Ecology and Biogeography*, *31*(5), 1006–1020. https://doi.org/10.1111/geb.13479

Hanson, J. O. (2022). *Wdpar: Interface to the world database on protected areas*. R package version 1.3.3. https://CRAN.R-project.org/package=wdpar

Heino, J. (2011). A macroecological perspective of diversity patterns in the freshwater realm. *Freshwater Biology*, *56*(9), 1703–1722. https://doi.org/10.1111/j.1365-2427.2011.02610.x

Heino, J., Grönroos, M., Ilmonen, J., Karhu, T., Niva, M., & Paasivirta, L. (2013). Environmental heterogeneity and β diversity of stream macroinvertebrate communities at intermediate spatial scales. *Freshwater Science*, *32*(1), 142–154. https://doi.org/10.1899/12-083.1

Heino, J., Melo, A. S., Bini, L. M., Altermatt, F., Al-Shami, S. A., Angeler, D. G., Bonada, N., Brand, C., Callisto, M., Cottenie, K., Dangles, O., Dudgeon, D., Encalada, A., Göthe, E., Grönroos, M., Hamada, N., Jacobsen, D., Landeiro, V. L., Ligeiro, R., & Townsend, C. R. (2015). A comparative analysis reveals weak relationships between ecological factors and beta diversity of stream insect metacommunities at two spatial levels. *Ecology and Evolution*, *5*(6), 1235–1248. https://doi.org/10.1002/ece3.1439

IPBES. (2019). Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. S. Díaz, J. Settele, E. S. Brondízio, H. T. Ngo, M. Guèze, J. Agard, A. Arneth, P. Balvanera, K. A. Brauman, S. H. M. Butchart, K. M. A. Chan, L. A. Garibaldi, K. Ichii, J. Liu, S. M. Subramanian, G. F. Midgley, P. Miloslavich, Z. Molnár, D. Obura, A. Pfaff, S. Polasky, A. Purvis, J. Razzaque, B. Reyers, R. Roy Chowdhury, Y. J. Shin, I. J. Visseren-Hamakers, K. J. Willis, & C. N. Zayas (Eds.). IPBES Secretariat, Bonn, Germany. 56 pages.

Jähnig, S. C., Baranov, V., Altermatt, F., Cranston, P., Friedrichs-Manthey, M., Geist, J., He, F., Heino, J., Hering, D., Hölker, F., Jourdan, J., Kalinkat, G., Kiesel, J., Leese, F., Maasri, A., Monaghan, M. T., Schäfer, R. B., Tockner, K., Tonkin, J. D., & Domisch, S. (2021). Revisiting global trends in freshwater insect biodiversity. *Wiley Interdisciplinary Reviews: Water*, *8*(2), 1–5. https://doi.org/10.1002/wat2.1506

Kaelin, K., & Altermatt, F. (2016). Landscape-level predictions of diversity in river networks reveal opposing patterns for different groups of macroinvertebrates. *Aquatic Ecology*, *50*, 283–295.

Maasri, A., Jähnig, S. C., Adamescu, M. C., Adrian, R., Baigun, C., Baird, D. J., Batista-Morales, A., Bonada, N., Brown, L. E., Cai, Q., Campos-Silva, J. V., Clausnitzer, V., Contreras-MacBeath, T., Cooke, S. J., Datry, T., Delacámara, G., De Meester, L., Dijkstra, K.-D. B., Do, V. T., … Worischka, S. (2022). A global agenda for advancing freshwater biodiversity research. *Ecology Letters*, *25*(2), 255–263. https://doi.org/10.1111/ele.13931

Macadam, C. R., & Stockan, J. A. (2015). More than just fish food: Ecosystem services provided by freshwater insects. *Ecological Entomology*, *40*(S1), 113–123. https://doi.org/10.1111/een.12245

Martins, R. T., Brito, J., Dias-Silva, K., Leal, C. G., Leitão, R. P., Oliveira, V. C., Oliveira-Júnior, J. M. B., de Paula, F. R., Roque, F. O., Hamada, N., Juen, L., Nessimian, J. L., Pompeu, P. S., & Hughes, R. M. (2022). Congruence and responsiveness in the taxonomic compositions of Amazonian aquatic macroinvertebrate and fish assemblages. *Hydrobiologia*, *849*, 1–18.

Mateo, R. G., Felicísimo, Á. M., & Muñoz, J. (2010). Effects of the number of presences on reliability and stability of MARS species distribution models: The importance of regional niche variation and ecological heterogeneity. *Journal of Vegetation Science*, *21*(5), 908–922. https://doi.org/10.1111/j.1654-1103.2010.01198.x

Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. *Nature Communications*, *6*, 1–8. https://doi.org/10.1038/ncomms9221

Molineri, C., Nieto, C., Dos Santos, D. A., Emmerich, D., Zúñiga, M., Fierro, P., Pessacq, P., Gomez, D., Márquez, J. A., Príncipe, R. E., Valdovinos Zarges, C., & Domínguez, E. (2020). Do mayflies (Ephemeroptera) support a biogeographic transition zone in South America? *Journal of Biogeography*, *47*(9), 1980–1993. https://doi.org/10.1111/jbi.13868

Morse, J. C. (Ed.). (2022). Trichoptera World Checklist. Retrieved November 2, 2022, from http://entweb.clemson.edu/database/trichopt/index.htm

Murphy, K., Efremov, A., Davidson, T. A., Molina-Navarro, E., Fidanza, K., Crivelari Betiol, T. C., Chambers, P., Grimaldo, J. T., Martins, S. V., Springuel, I., Kennedy, M., Mormul, R. P., Dibble, E., Hofstra, D., Lukács, B. A., Gebler, D., Baastrup-Spohr, L., Urrutia-Estrada, J., & Urrutia-Estrada, J. (2019). World distribution, diversity and endemism of aquatic macrophytes. *Aquatic Botany*, *158*, 103127. https://doi.org/10.1016/j.aquabot.2019.06.006

Neu, P. J., Malicky, H., Graf, W., & Schmidt-Kloiber, A. (2018). *Distribution atlas of European Trichoptera*. ConchBooks.

Nieto, C., Molineri, C., & Dominguez, E. (2022). Mayflies of South America. IGB Leibniz-Institute of Freshwater Ecology and Inland Fisheries. dataset. https://doi.org/10.18728/igb-fred-795.2

Nieto, C., Ovando, X. M. C., Loyola, R., Izquierdo, A., Romero, F., Molineri, C., Rodríguez, J., Rueda Martin, P., Fernández, H., Manzo, V., & Miranda, M. J. (2017). The role of macroinvertebrates for conservation of freshwater systems. *Ecology and Evolution*, *7*(14), 5502–5513.

Paulson, D., Schorr, M., & Deliry, C. (2022). *World Odonata list*. Slater Museum of Natural History. https://www2.pugetsound.edu/academics/academic-resources/slater-museum/biodiversity-resources/dragonflies/world-odonata-list2/

R Core Team. (2020). R: A language and environment for statistical computing. https://www.r-project.org/

Reid, A. J., Carlson, A. K., Creed, I. F., Eliason, E. J., Gell, P. A., Johnson, P. T. J., Kidd, K. A., TJ, M. C., Olden, J. D., Ormerod, S. J., Smol, J. P., Taylor, W. W., Tockner, K., Vermaire, J. C., Dudgeon, D., & Cooke, S. J. (2019). Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews*, *94*(3), 849–873. https://doi.org/10.1111/brv.12480

Sánchez-Bayo, F., & Wyckhuys, K. A. G. (2019). Worldwide decline of the entomofauna: A review of its drivers. *Biological Conservation*, *232*, 8–27. https://doi.org/10.1016/j.biocon.2019.01.020

Sánchez-Fernández, D., Lobo, J. M., & Hernández-Manrique, O. L. (2011). Species distribution models that do not incorporate global data misrepresent potential distributions: A case study using Iberian diving beetles. *Diversity and Distributions*, *17*(1), 163–171. https://doi.org/10.1111/j.1472-4642.2010.00716.x

Sarremejane, R., Cid, N., Stubbington, R., Datry, T., Alp, M., Cañedo-Argüelles, M., Cordero-Rivera, A., Csabai, Z., Gutiérrez-Cánovas, C., Heino, J., Forcellini, M., & Bonada, N. (2020). DISPERSE, a trait database to assess the dispersal potential of European aquatic macroinvertebrates. *Scientific Data*, *7*(1), 1–9.

Schmidt-Kloiber, A., & Hering, D. (2015). Www.freshwaterecology.info—An online tool that unifies, standardises and codifies more than 20,000 European freshwater organisms and their ecological preferences. *Ecological Indicators*, *53*, 271–282. https://doi.org/10.1016/j.ecolind.2015.02.007

Schmidt-Kloiber, A., Neu, P. J., Malicky, M., Pletterbauer, F., Malicky, H., & Graf, W. (2017). Aquatic biodiversity in Europe: A unique dataset on the distribution of Trichoptera species with important implications for conservation. *Hydrobiologia*, *797*(1), 11–27. https://doi.org/10.1007/s10750-017-3116-4

Simaika, J. P., Samways, M. J., Kipping, J., Suhling, F., Dijkstra, K. D. B., Clausnitzer, V., Boudot, J. P., & Domisch, S. (2013). Continental-scale conservation prioritization of African dragonflies. *Biological Conservation*, *157*, 245–254.

Støa, B., Halvorsen, R., Stokland, J. N., & Gusarov, V. I. (2019). How much is enough? Influence of number of presence observations on the performance of species distribution models. *Sommerfeltia*, *39*(1), 1–28. https://doi.org/10.2478/som-2019-0001

Sullivan, S. M. P., & Manning, D. W. P. (2019). Aquatic–terrestrial linkages as complex systems: Insights and advances from network models. *Freshwater Science*, *38*(4), 936–945. https://doi.org/10.1086/706071

Tickner, D., Opperman, J. J., Abell, R., Acreman, M., Arthington, A. H., Bunn, S. E., Cooke, S. J., Dalton, J., Darwall, W., Edwards, G., Harrison, I., Hughes, K., Jones, T., Leclère, D., Lynch, A. J., Leonard, P., McClain, M. E., Muruven, D., Olden, J. D., … Young, L. (2020). Bending the curve of global freshwater biodiversity loss: An emergency recovery plan. *Bioscience*, *70*(4), 330–342. https://doi.org/10.1093/biosci/biaa002

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. *Global Ecology and Biogeography*, *30*(4), 826–841. https://doi.org/10.1111/geb.13257

UNEP-WCMC and IUCN. (2022, November). *Protected planet: The world database on protected areas (WDPA) and world database on other effective area-based conservation measures (WD-OECM)* [Online]. www.protectedplanet.net

van Klink, R., Bowler, D. E., Gongalsky, K. B., & Chase, J. M. (2021). Revisiting global trends in freshwater insect biodiversity: A reply. *Wiley Interdisciplinary Reviews: Water*, *8*(2), e1501.

van Klink, R., Bowler, D. E., Gongalsky, K. B., Swengel, A. B., Gentile, A., & Chase, J. M. (2020). Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science*, *368*(6489), 417–420.

van Proosdij, A. S. J., Sosef, M. S. M., Wieringa, J. J., & Raes, N. (2016). Minimum required number of specimen records to develop accurate species distribution models. *Ecography*, *39*(6), 542–552. https://doi.org/10.1111/ecog.01509

van Rees, C. B., Waylen, K. A., Schmidt-Kloiber, A., Thackeray, S. J., Kalinkat, G., Martens, K., Domisch, S., Lillebø, A. I., Hermoso, V., Grossart, H. P., Schinegger, R., Decleer, K., Adriaens, T., Denys, L., Jarić, I., Janse, J. H., Monaghan, M. T., De Wever, A., Geijzendorffer, I., … Jähnig, S. C. (2021). Safeguarding freshwater life beyond 2020: Recommendations for the new global biodiversity framework from the European experience. *Conservation Letters*, *14*(1), e12771. https://doi.org/10.1111/conl.12771

Vinson, M. R., & Hawkins, C. P. (2003). Broad-scale geographical patterns in local stream insect genera richness. *Ecography*, *26*(6), 751–767. https://doi.org/10.1111/j.0906-7590.2003.03397.x

Vorster, C., Samways, M. J., Simaika, J. P., Kipping, J., Clausnitzer, V., Suhling, F., & Dijkstra, K.-D. (2020). Development of a new continental-scale index for freshwater assessment based on dragonfly assemblages. *Ecological Indicators*, *109*, 105819.

Wallace, J. B., & Webster, J. R. (1996). The role of macroinvertebrates in stream ecosystem function. *Annual Review of Entomology*, *41*(1), 115–139. https://doi.org/10.1146/annurev.en.41.010196.000555

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

WWF. (2022). Living planet report 2022—Building a naturepositive society. In R. E. A. Almond, M. Grooten, D. Juffe Bignoli, & T. Petersen (Eds.). WWF.

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G., & Pavelsky, T. (2019). MERIT hydro: A high-resolution global hydrography map based on latest topography datasets. *Water Resources Research*, *55*, 5053–5073. https://doi.org/10.1029/2019WR024873

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., & Antonelli, A. (2019). CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, *10*(5), 744–751. https://doi.org/10.1111/2041-210X.13152

BIOSKETCH

Afroditi Grigoropoulou is a doctoral student at the Leibniz Institute of Freshwater Ecology and Inland Fisheries in Berlin, working on spatial patterns of freshwater insect biodiversity and developing hydrographic data processing tools. The EPTO database is a result of a global collaboration among many freshwater scientists who share an interest in freshwater ecology and biogeography over multiple years. This effort is part of the Global Freshwater Biodiversity, Biogeography and Conservation (GLOWABIO) project (https://glowabio.org/), aiming towards an improved global spatial representation of freshwater biodiversity.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Grigoropoulou, A., Hamid, S. A., Acosta, R., Akindele, E. O., Al-Shami, S. A., Altermatt, F., Amatulli, G., Angeler, D. G., Arimoro, F. O., Aroviita, J., Astorga-Roine, A., Bastos, R. C., Bonada, N., Boukas, N., Brand, C., Bremerich, V., Bush, A., Cai, Q., Callisto, M. … Domisch, S. (2023). The global EPTO database: Worldwide occurrences of aquatic insects. *Global Ecology and Biogeography*, *00*, 1–14. https://doi.org/10.1111/geb.13648