Knowl. Manag. Aquat. Ecosyst. 2025, 426, 7
© D. Mameri *et al.*, Published by EDP Sciences 2025
https://doi.org/10.1051/kmae/2025002

www.kmae-journal.org

Knowledge & Management of Aquatic Ecosystems

Journal fully supported by Office français de la biodiversité

DATA PAPER OPEN ACCESS

Fishspotting: freshwater fish species presence in European river basins – RivFISH database

Daniel Mameri[1,*], Gonçalo Duarte[1,*], João Cabo[1], Rui Figueira[2,3], Pedro Segurado[1], José Maria Santos[1], Maria Teresa Ferreira[1] and Paulo Branco[1]

[1] Forest Research Centre, Associate Laboratory TERRA, School of Agriculture, University of Lisbon, Tapada da Ajuda, 1349-017 Lisbon, Portugal
[2] CIBIO/InBio, Centro de Investigação em Biodiversidade e Recursos Genéticos, Laboratório Associado, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisbon, Portugal
[3] CIBIO/InBio, Centro de Investigação em Biodiversidade e Recursos Genéticos, Laboratório Associado, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal

Received January 7, 2025 – Accepted February 17, 2025

Abstract – Biodiversity assessments at the river basin level represent a crucial step in drawing conservation strategies for freshwater biota. As river basins are independent functional units, it is fundamental to ensure that spatial information on species presence is fitted to the basins where they effectively occur. To maximize quality and fitness for data usage, scientific names recorded during fish sampling, which may be outdated (synonyms), need to be associated with currently accepted species names. The RivFISH database shows the presence of all native freshwater-dependent fish species in 1554 European river basins, linking data from both synonyms and accepted scientific species names. Spatial data was compiled from a total of 68 references and confronted with the accepted names lists from FishBase, thus ensuring the correspondence of the recorded species name with its current accepted names and river basin presence. RivFISH is composed of six normalized tables reflecting the association between species taxonomy, presence in each river basin, and IUCN risk category, and one denormalized table including all fields for user convenience. Overall, RivFISH includes a total of 707 scientific species names, 667 of which are currently accepted, and aims to facilitate large-scale studies in Europe and promote the interoperability with conservation-related datasets.

Keywords: Biodiversity assessments / river basins / species presence / synonyms / interoperability

1 Introduction

Rivers are among the most biodiverse ecosystems worldwide, harbouring one-quarter of all known vertebrate species (Dudgeon *et al.*, 2006; Reid *et al.*, 2019). Among vertebrates, fish are the most representative group, with nearly 17,800 species described so far, and a high proportion of endangered species (Collen *et al.*, 2014; Tedesco *et al.*, 2012). Particularly in Europe, over 500 native freshwater-dependent fish species have been reported to occur in rivers (Freyhof and Brooks, 2011), and currently, over a third of these taxa raise some level of conservation concern (IUCN, 2024).

The high degree of imperilment of European fish species is coupled with the current threats that freshwater ecosystems are facing, including the presence of barriers, water management, droughts, and the introduction and spread of alien species (Costa *et al.*, 2021). To develop effective strategies for freshwater ecosystem management and fish conservation, the nested and hierarchical nature of rivers must be taken into consideration. As river basins function independently from each other, acting as "islands" for the biological instream communities they host, these can be regarded as the main units for freshwater conservation and management.

As fish are confined in river basins, it is fundamental to ensure that spatial data on species occurrence is correct. Imprecisions in spatial references may lead to an allocation of a species to a basin where it does not occur. Specifically in the case of species assessed by the IUCN, the distribution maps available for each species are often built based on convex hulls – the smallest convex set joining the occurrence data (Burgman and Fox, 2003). While this method fulfils the purpose of standardizing the resolution of the distribution maps for all species assessed by the IUCN, it may not reflect the real occurrences of freshwater fish.

* Corresponding authors: dmameri@isa.ulisboa.pt; goncalo.f.duarte@edu.ulisboa.pt

Table 1. List of tables (and corresponding description) composing the RivFISH database.

Table	Description
Current_species	List of currently accepted species names, validated using the FishBase information system. The unique code used by FishBase for each of these scientific names ('Syn_code') was also included in the table to facilitate interoperability between the two datasets.
Species_names	Relationship between known recorded species names and corresponding current accepted species names. This correspondence was made by verifying each synonym in FishBase.
IUCN_assessments	Risk assessment for each currently accepted species name evaluated by the International Union for Conservation of Nature (IUCN).
CCM_sea_outlets	River basin network draining to the sea (sea outlets) of the Catchment Characterisation and Modelling (CCM), and the corresponding main basin from the HydroSHEDS layer. The country where the largest area of each river basin is located is also presented (ISO code).
Species_basins	Species presence per river basin, at the resolution of the Catchment Characterisation and Modelling (CCM) and main basin of the HydroSHEDS layer.
References	References used to confirm the presence of each species in each river basin.

Another issue related to spatial data on species occurrence is the taxonomic change that species undergo throughout time. A species that was recorded in a certain river basin may have had its scientific name changed, or it may have split into two species, and the previously recorded names are now misused (synonyms). If the link between the synonym and the current species name is not established, the time and spatial information may be lost. Hence, a correspondence between species synonyms and their current scientific name is essential to maximize data usage and adequately reflect species occurrence. FishBase, an online information system containing data on all described fish species and which feeds the fish data in Catalogue of Life (Bánki *et al.*, 2024), presents a list of synonyms for each accepted scientific name (Froese and Pauly, 2024a). However, an association at the river basin level is required, as the geographical description for each species (*e.g.* "Adriatic basin") does not allow a systematic allocation of each species to river basins.

Hence, the goal of the RivFISH database is to compile presence records, per river basin, of freshwater-dependent fish species, using all the recorded species names up to date, while also linking this information to data with conservation relevance, such as the IUCN Red List Categories. The RivFISH database is structured to account for updates, including the addition of new taxa (both synonyms and new species), but also new records of species occurrences in river basins.

2 Materials and methods

2.1 Data collection

Data was acquired for native freshwater-dependent fish species (resident and obligatory migrants) currently occurring in European river basins. The Ural and Caucasus Mountains were defined as the geographical boundaries for including river basins in Europe, plus Turkey (delimited on the south by the Orontes river basin), to account for diadromous species (that migrate between fresh waters and sea) that also occur in European river basins draining to the Mediterranean Sea, Black Sea and Caspian Sea.

The list of recorded species names associated with spatial data was elaborated by consulting existing data sources and literature, from a total of 68 references, including the IUCN Red List (IUCN, 2024) and the Freshwater Fish Ecology & Distribution (FFED) database (Tedesco *et al.*, 2017). The majority of the presence records were obtained from the FFED database (38.0%), followed by the datasets compiled under the BioFresh project (Schmidt *et al.*, 2019), published through the Global Information Biodiversity Platform (GBIF, 22.0%), scientific articles (13.0%), the IUCN Red List (11.1%), Kottelat and Freyhof's European Handbook of Freshwater Fish Species (9.8%), and other data sources including books, project reports and other datasets (6.1%).

2.2 Data validation and standardization

Following the inventory of fish presence per river basin, each species name was confronted with the accepted and synonym species names lists from FishBase, thus ensuring the correspondence of the recorded species name with spatial data and its current validated names. All species names were verified in October 2024 using the Checklist Bank portal, selecting the FishBase dataset (Froese and Pauly, 2024a). For presence records, data was retrieved from the sources listed in the Table "References" of the RivFISH database (Tab. 1).

The bibliographic reference used to confirm the presence of a species in each river basin was also included in the database, in the table "Species_basins" (Tab. 1). Spatial data for native fish species presence available at both the IUCN Red List and FFED is based on the watershed boundaries represented in HydroBASINS, derived from the HydroSHEDS database, at a 15-second resolution (Lehner & Grill, 2013). While the data retrieved from FFED comes in a tabular format (csv files) with the name of the main river basins in HydroBASINS, species occurrence in the IUCN may be provided through polygon layers based on the convex hull of HydroSHEDS level 8 units, whenever fish distribution cannot be allocated to the individual HydroSHEDS units. This may result in overestimating species presence in river basins if the spatial data is not previously processed. Hence, spatial data retrieved from the IUCN Red List was intersected with a vectorized river basins layer, and

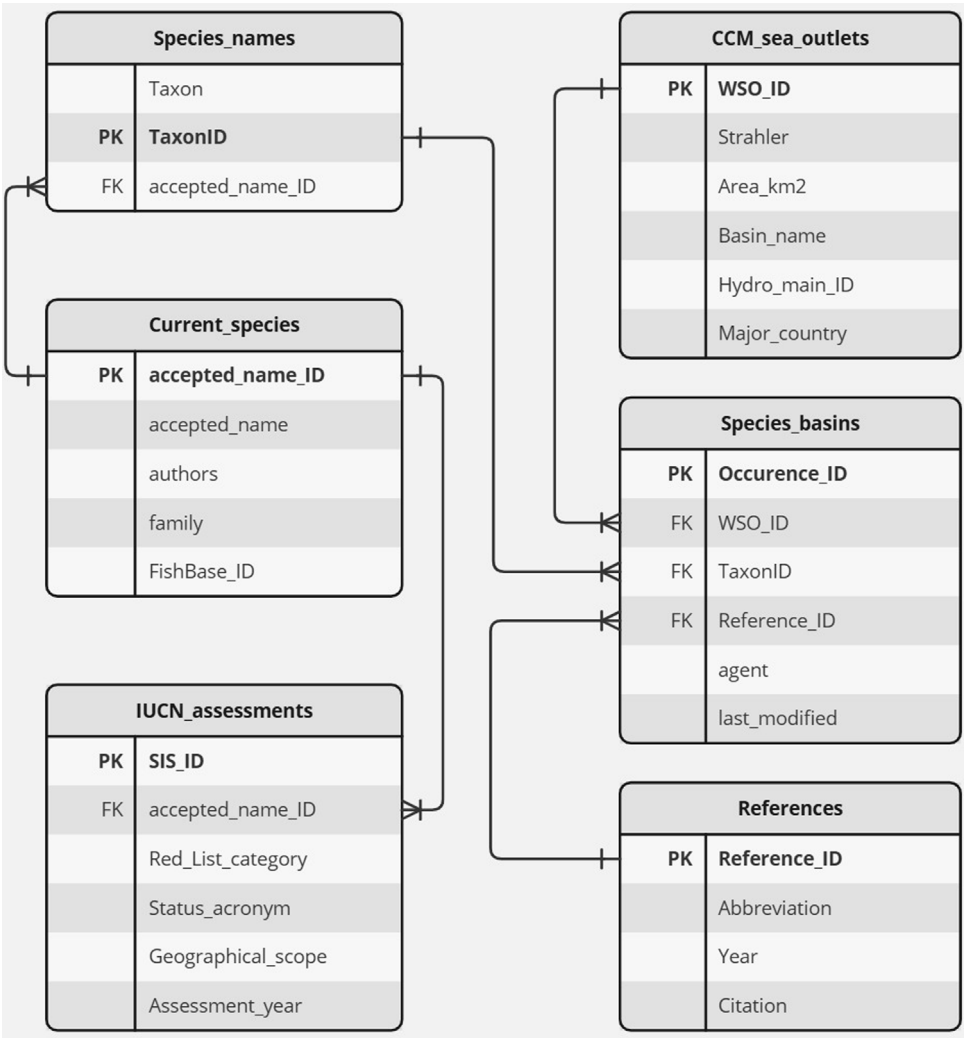

Fig. 1. Entity-relationship diagram for the RivFISH database. Each table contains one primary key (PK), with unique values that may link to other tables through a foreign key (FK). RivFISH includes a total of six tables, covering species' taxonomy ("Current_species", "Species_names"), spatial data ("CCM_sea_outlets" and "Species_basins", with the last associated to the references from the literature) and IUCN information ("IUCN_assessments").

the presence of each species per basin was verified in the literature.

The spatial reference for the sea outlet river basins was taken from the Catchment Characterisation and Modelling (CCM2) – River and Catchment Database v2.1 (de Jager and Vogt, 2007). Overall, the CCM2 layer of river basins (where each basin has a unique ID – "WSO_ID") presents a better adjustment to European basins and enables interoperability with other datasets requiring river representations at finer scales (*e.g.*, river instream barriers). It should be noted that, for the RivFISH database, only CCM2 river basins with maximum Strahler stream order above 2 were considered (representing 95.8% of the study area), as these have been identified as thresholds for freshwater biodiversity (Vorste *et al.*, 2017).

The RivFISH database seeks to increase interoperability with reference datasets, by adopting their persistent identifiers. This is possible for basin data (CCM2 and HydroSHEDS), and for the IUCN status, but not for FishBase, as this database actively decided to not adopt persistent identifiers (Froese and Pauly 2024b). To ensure interoperability with the retrieved data on species presence from the IUCN Red List and FFED, an association between the two catchment datasets was established. Each HydroSHEDS level 8 unit (for the spatial data retrieved in the IUCN Red List) corresponds to a sub-basin linked to a main river basin with a unique identifier, which was used to establish an association with the CCM river basins. Complementarily, the similarity of the main basin name (available in FFED) in both catchment models was used to match the basins. Finally, a spatial overlap between the CCM and HydroSHEDS basins was performed to link the remaining basins and validate species presence. This match allowed correcting species presence, by excluding polygons in vector layers that fell into adjacent river basins due to the changes in spatial resolution, and that are not supported by consulted datasets and literature (table "References", see Tab. 1). All spatial analyses were conducted in ArcGIS Pro (version 3.2.2), using the coordinate reference system ETRS89-extended/ LAEA Europe (EPSG: 3035).

Table 2. Relations between fields in the RivFISH database with the original data sources and other data standards. Fields related to taxonomy are linked with the Darwin Core standard. FishBase's 'SynCode' can be consulted through the package 'rfishbase' (Boettiger et al., 2012) available for R.

Table	Field	Associated source	Term/Field
Current_species	Accepted_name	Darwin Core standard	AcceptedNameUsage
Current_species	Accepted_name_ID	Darwin Core standard	AcceptedNameUsageID
Current_species	FishBase_ID	FishBase	SynCode
Species_names	Taxon	Darwin Core standard	ScientificName
Species_names	TaxonID	Darwin Core standard	TaxonID
IUCN_assessments	SIS_ID	IUCN Species Information Service (SIS)	SIS_ID
CCM_sea_outlets	WSO_ID	CCM2	WSO_ID
CCM_sea_outlets	Hydro_main_ID	HydroSHEDS	Main_bas

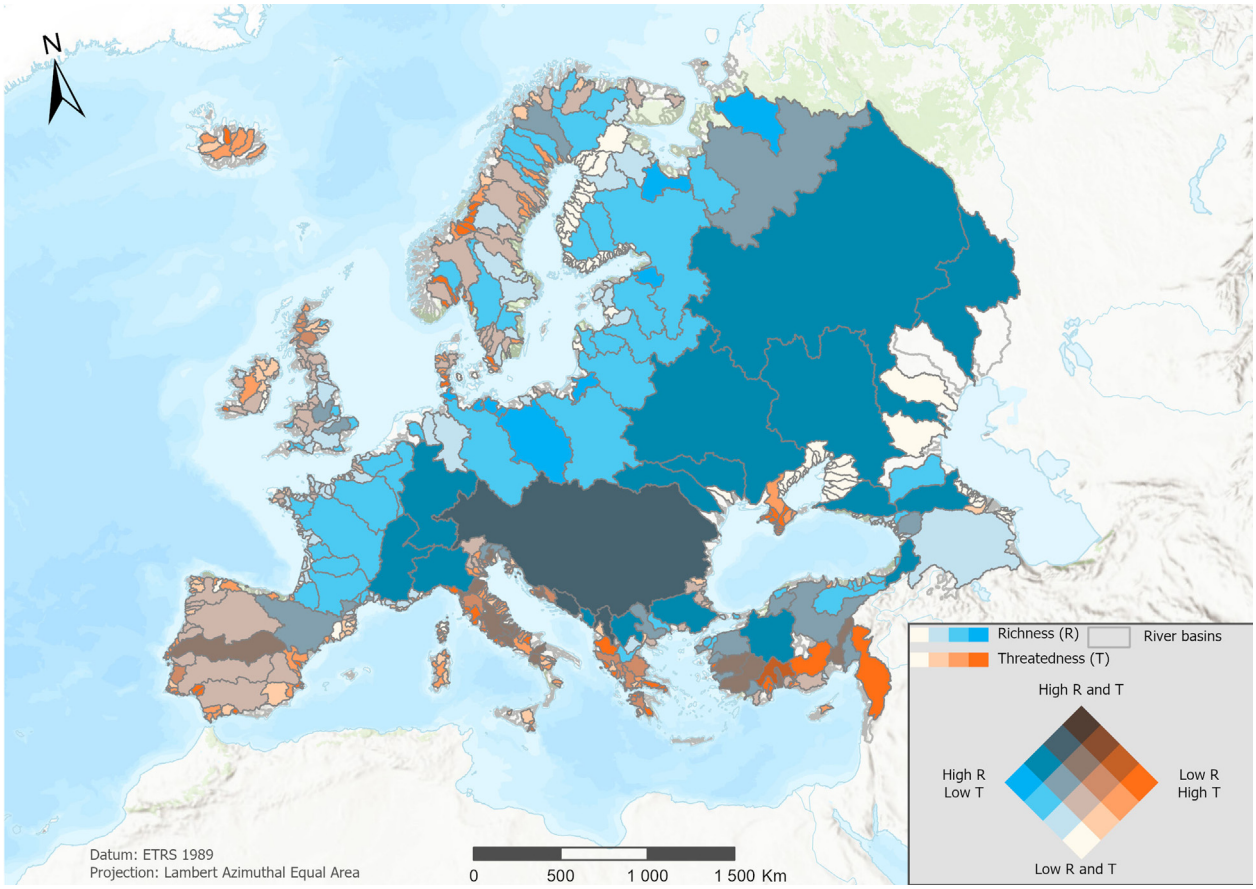

Fig. 2. Map representing species richness versus the ratio of threatened species (listed as "Vulnerable", "Endangered" or "Critically Endangered") per river basin, considering the risk category in the IUCN (International Union for Conservation of Nature) Red List. The map was produced in ArcGIS Pro (version 3.2.2) with the European sea outlet basins layer from the Catchment Characterisation and Modelling 2 (CCM2) dataset (de Jager and Vogt, 2007). All basins with a maximum Strahler above 2 and harbouring more than 2 native species were considered for this analysis, including 756 river basins, representing 90.9% of RivFISH study area.

2.3 Data structure and fields

The RivFISH database is organized into six tables, reflecting the taxonomic, spatial and conservation status data on all native freshwater-dependent fish species occurring in Europe, including Turkey. The database architecture implemented a normalized relational model, implemented in SQL. A short description of each table is presented in Table 1. A complete metadata file, following the DarwinCore biodiversity standard and including each field's description, is also provided in the Supplementary Information. These tables are linked through primary and foreign keys, based on unique identifiers (IDs), including an ID for each recorded species name (*"TaxonID"*), current accepted scientific name (*"accepted_name_ID"*), and CCM basin ID (*"WSO_ID"*), as depicted in Figure 1.

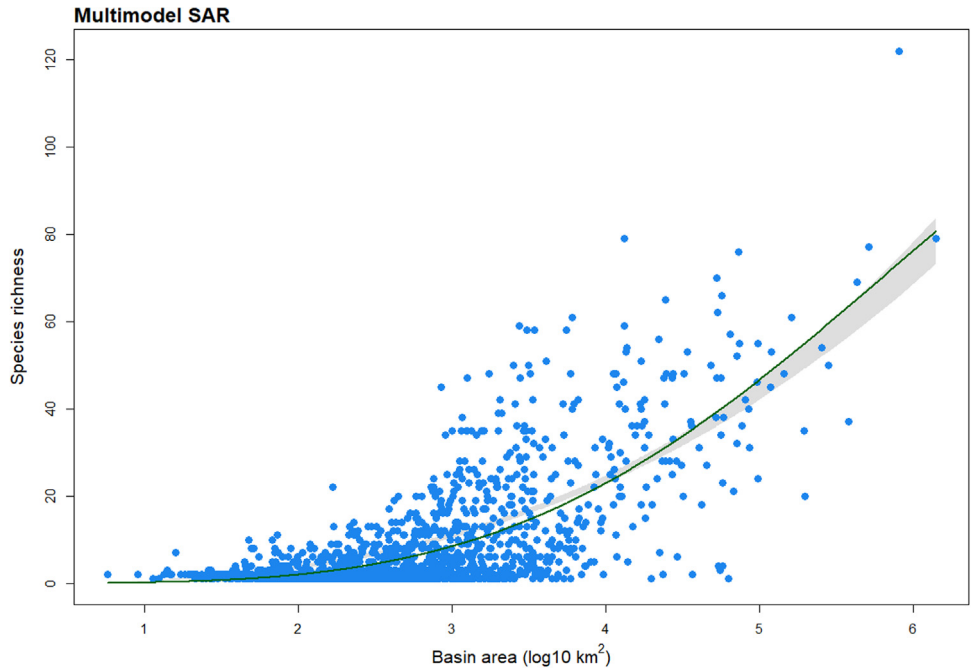

Fig. 3. Species-area relationship (SAR), based on the log-transformed basin area. A multi-model inference approach was followed to determine a multi-model averaged curve with the best fit to the data (green curve, $R^2 = 0.542$). Grey boundaries represent the bootstrap 95% confidence interval of the multi-model averaged curve. The SAR plot was produced in R, version 4.1.0 (R Core Team, 2021), using the "sars" package, version 1.3.6 (Matthews, Triantis, Whittaker, and Guilhaumon, 2019). All basins with a maximum Strahler above 2 were considered for this analysis, including 1554 river basins, representing 95.8% of RivFISH study area.

The interoperability to other databases is ensured by the fields and links listed in Table 2. For user convenience, a denormalised flat table named "RivFISH_denormalised" was also included, containing all fields of the related tables.

3 Using the database

A total of 707 recorded taxa names, including 667 currently accepted scientific names, were included in the RivFISH database for Europe, including Turkey. All 667 species have spatial data associated with the CCM2 river basin resolution (1554 basins with maximum Strahler above 2, covering 95.8% of the total drainage area within the study area covered by RivFISH).

Regarding the conservation status in the IUCN, 633 freshwater-dependent fish species have a risk category assigned (94.9% of the currently accepted species included in RivFISH). Figure 2 provides a usage example that allows visualizing freshwater-dependent fish species richness per basin across Europe and where communities are most threatened, according to the IUCN. Southern Europe, including the Iberian, Adriatic and Balkan peninsulas, as well as southern European Turkey, stand out as the areas where there is a higher proportion of endangered species, given the higher levels of threatedness and lower species richness (Fig. 2).

Figure 3 shows another usage example of the database, displaying a positive association between species richness and basin area ($R^2 = 0.542$). Scale dependence and positive species-area relationship (SAR) is a well-known pattern in ecology (Lomolino, 2001), which also applies to stream fishes (Angermeier and Schlosser, 1989; Oberdorff *et al.*, 2011). It should be noted, however, that the SAR may be diluted due to their environmental condition, such as habitat complexity and environmental variation (Oberdorff *et al.*, 2011).

The Danube River basin (basin area of 800,000 km^2, the uppermost point of Fig. 3), for instance, presents a high heterogeneity across its sub-basins, with a total of 122 freshwater fish-dependent species, including several endemic species, thus standing out from the SAR pattern evidenced in Figure 3. The delimitation of the Danube drainage in the CCM2 layer also incorporates karstic (sinking) freshwater systems, including the Jadova and Lika rivers in the Western Balkans, where endemic species such as *Cobitis jadovensis* occur (Mustafic *et al.*, 2008), contributing to the high biodiversity found in the Danube. Contrastingly, the biggest river basin (1,400,000 km^2) in Europe, the Volga (the rightmost point of Fig. 3), harbours 79 freshwater-dependent fish species.

4 Discussion

The RivFISH database aggregates the available data on freshwater-dependent fish presence in Europe, validated at the river basin level and considering taxonomical synonyms for species names. This allows for a maximization of data usage and robustness. Having a correct species presence per basin enables an accurate starting point for proper management and conservation actions to be planned at a finer scale, namely at the sub-basin and river segment scale. It is, as far as the authors know, the most up-to-date and comprehensive database on the

presence of freshwater-dependent fish species in European river basins, with the inclusion of 1554 river basins and 667 freshwater-dependent fish species. This represents an increase of over 120 species described in Europe since the publication of the Handbook of European Freshwater Fishes, which reported a total of 546 species by 2011 (Freyhof & Brooks, 2011).

RivFISH can be further used to investigate species occurrence, fish biodiversity and community vulnerability at the European extent, by crossing the taxonomic and spatial layers of information with other available datasets, such as climate data and anthropogenic pressures. The structure of the database is also prepared to deal with future alterations in species taxonomy, as well as new records of species occurrence in river basins, including recently described species (2023 and 2024) that are yet to be included and validated in both FishBase and the Catalogue of Life (*e.g.* (Doadrio, Sousa-Santos, Robalo, & Perea, 2024).

Acknowledgments

The authors would like to thank Iro Sipsa, Miguel Silva and Juliana Sêco for their contribution to data collection. This study was funded by the project "Dammed Fish – Impact of structural and functional river network connectivity losses on fish biodiversity – Optimizing management solutions" (PTDC/CTA-AMB/4086/2021, DOI: 10.54499/PTDC/CTA-AMB/4086/2021), from Fundação para a Ciência e Tecnologia, I.P. (FCT). DM and GD have been financed by FCT within the Dammed Fish project. PB was financed by national funds via FCT (LA/P/0092/2020). Forest Research Centre (CEF) is a research unit funded by FCT through project reference UID/00239: Centro de Estudos Florestais, DOI 10.54499/UIDB/00239/2020. The Associate Laboratory TERRA (LA/P/0092/2020, DOI: 10.54499/LA/P/0092/2020) is also funded by FCT.

Data availability statement

The RivFISH database (version 1.2) is published in tabular format (csv files) in the public repository Zenodo (https://doi.org/10.5281/zenodo.13848976), with the corresponding metadata, according to the EML 2.1.1 standard, in xml format. A shapefile containing the data on the presence of fish species included in the database ('rivfish_shapefile') is also made available. All files are compressed in a single zipped folder.

Supplementary material

The Supplementary Material is available at https://www.kmae-journal.org/10.1051/kmae/2025002/olm.

References

Angermeier PL, Schlosser IJ. 1989. Species-area relationship for stream fishes. *Ecology* 70: 1450–1462.

Bánki O, Roskov Y, Döring M, Ower G, Hernández Robles DR, Plata Corredor CA, Stjernegaard Jeppesen T, Örn A, Pape T, Hobern D, Garnett S, Little H, DeWalt RE, Ma K, Miller J, Orrell T, Aalbu R, Abbott J, Adlard R, *et al.* 2025. Catalogue of Life (Version 2025-02-13). Catalogue of Life, Amsterdam, Netherlands.

Burgman MA, Fox JC. 2003. Bias in species range estimates from minimum convex polygons: implications for conservation and options for improved planning. *Anim Conserv* 6: 19–28.

Boettiger C, Lang DT, Wainwright PC. 2012. Rfishbase: exploring, manipulating and visualizing FishBase data from R. *J Fish Biology* 81: 2030–2039.

Collen B, Whitton F, Dyer EE, Baillie JE, Cumberlidge N, Darwall WR, Pollock C, Richman N, Soulsby A, Böhm M. 2014. Global patterns of freshwater species diversity, threat and endemism. *Global Ecol Biogeogr* 23: 40–51.

Costa MJ, Duarte G, Segurado P, Branco P. 2021. Major threats to European freshwater fish species. *Sci Total Environ* 797: 149105.

de Jager A, Vogt J. 2007. Rivers and Catchments of Europe – Catchment Characterisation Model (CCM). *121–144. European Commission, Joint Research Centre (JRC)*, [Dataset] PID: http://data.europa.eu/89h/fe1878e8-7541-4c66-8453-afdae7469221

Doadrio I, Sousa-Santos C, Robalo J, Perea S. 2024. Description of three new species of the genus Cobitis L., 1758. *Limnetica* 43: 177–198.

Dudgeon D, Arthington AH, Gessner MO, Kawabata Z-I, Knowler DJ, Lévêque C, Naiman RJ, Prieur-Richard AH, Soto D, Stiassny MLJ, Sullivan CA. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. *Biol Rev* 81: 163–182.

Freyhof J, Brooks E. 2011. European Red List of Freshwater Fishes. *European Commission: Directorate-General for Environment, Brooks, Publications Office*. URL: https://data.europa.eu/doi/10.2779/85903

Froese R, Pauly D. 2024a. *FishBase*. URL: https://www.fishbase.se/

Froese R, Pauly D. 2024b. *FishBase: The SPECIES Table*. URL: https://www.fishbase.se/

IUCN - International Union for Conservation of Nature. 2024. IUCN Red List of Threatened Species, Version 2024-1. URL: https://www.iucnredlist.org/

Lehner B, Grill G. 2013. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrol Processes* 27: 2171–2186.

Lomolino. 2001. The species-area relationship: new challenges for an old pattern. *Prog Phys Geogr: Earth Environ* 25: 1–21.

Matthews TJ, Triantis KA, Whittaker RJ, Guilhaumon F. 2019. sars: an R package for fitting, evaluating and comparing. *Ecography* 42: 1446–1455.

Mustafic P, Marcic Z, Duplic A, Mrakovcic M, Caleta M, Zanella D, Buj I, Podnar M, Dolenec Z. 2008. A new loach species of the genus *Cobitis* in Croatia. *Folia Zoolog* 57: 4–9.

Oberdorff T, Tedesco PA, Hugueny B, Leprieur F, Beauchard O, Brosse S, Dürr HH. 2011. Global and regional patterns in riverine fish species richness: a review. *Int J Ecol* 2011:967631.

R Core Team. 2021. *The R Foundation for Statistical Computing Platform*.

Reid AJ, Carlson AK, Creed IF, Eliason EJ, Gell PA, Johnson PTJ, Kidd KA, Maccormack TJ, Olden JD, Ormerod SJ, Smol JP, Taylor WW, Tockner K, Vermaire JC, Dudgeon D, Cooke S. 2019. Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biol Rev* 94: 849–873.

Schmidt-Kloiber A, Bremerich V, De Wever A, Jähnig S, Martens K, Strackbein J, Tockner K, Hering D. 2019. The Freshwater Information Platform: a global online network providing data, tools and resources for science and policy support. *Hydrobiologia* 838: 1–11.

Tedesco PA, Beauchard O, Bigorne R, Blanchet S, Buisson L, Conti L, Cornu JF, Dias MS, Grenouillet G, Hugueny B, Jézéquel C, Leprieur F, Brosse S, Oberdorff T. 2017. A global database on freshwater fish species occurrence in drainage basins. *Sci Data* 4: 170141.

Tedesco PA, Leprieur F, Hugueny B, Brosse S, Dürr HH, Beauchard O, Busson F, Oberdorff T. 2012. Patterns and processes of global riverine fish endemism. *Glob Ecol Biogeogr* 21: 977–987.

Vorste RV, McElmurray P, Bell S, Eliason K, Brown BL. 2017. Does stream size really explain biodiversity patterns in lotic systems? A call for mechanistic explanations. *Diversity* 9: 26.

Cite this article as: Mameri D, Duarte G, Cabo J, Figueira R, Segurado P, Santos JM, Ferreira MT, Branco P. 2025. Fishspotting: freshwater fish species presence in European river basins - RivFISH database. *Knowl. Manag. Aquat. Ecosyst.*, 426, 7. https://doi.org/10.1051/kmae/2025002.