Data in Brief 54 (2024) 110525

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Freshwater fish scale database

Bálint Bánó[a,b,c], Aleksey Bolotovskiy[d], Boris Levin[d,e], George M.T. Mattox[f], Mauricio Cetra[g], István Czeglédi[b,c], Péter Takács[c,*]

[a] *Department of Applied Fish Biology, Institute of Aquaculture and Environmental Safety, Hungarian University of Agriculture and Life Sciences, Kaposvár, Hungary*
[b] *National Laboratory for Water Science and Water Security, HUN-REN Balaton Limnological Research Institute, Tihany, Hungary*
[c] *HUN-REN Balaton Limnological Research Institute, Tihany, Hungary*
[d] *Papanin Institute for Biology of Inland Waters, Russian Academy of Sciences, Borok, Russia*
[e] *A.N. Severtsov Institute of Ecology and Evolution of the Russian Academy of Sciences, Moscow, Russia*
[f] *Laboratório de Ictiologia de Sorocaba, Departamento de Biologia, Universidade Federal de São Carlos, Sorocaba, Brazil*
[g] *Departamento de Ciências Ambientais, Universidade Federal de São Carlos, Sorocaba, Brazil*

ARTICLE INFO

Article history:
Received 14 March 2024
Revised 16 April 2024
Accepted 7 May 2024
Available online 11 May 2024

Dataset link: Scale photo database of freshwater fishes (Original data)

Keywords:
Photo
Morphometrics
Shape
Size, artificial intelligence
Environmental characteristics

ABSTRACT

Scales of the ray finned fishes can be used for multiple purposes. Beside others by their specific structure and elemental composition they are usable for age determination and food-web researches. Additionally, just their presence, absence, shape, location, or numbers could provide a reliable taxonomic information. The fish scales show remarkable size variation also, which characteristics provide reliable information about the environmental needs of freshwater fish. But till now this information was not interpretable and comparable in the absence of a sufficiently large and detailed database. In this study we provide a database which can facilitate the further complex comparative studies.
Our dataset consists of 2954 scale photos characterising 193 freshwater fish species. All photos have a uniquely identification code showing exactly which species the scale belongs to. In addition to the photos, our database includes a table showing the exact taxonomic classification of the studied species, the average body profile-index, and the relative

* Corresponding author.
E-mail address: takacs.peter@blki.hu (P. Takács).

https://doi.org/10.1016/j.dib.2024.110525

scale size, as well as the ecological (flow and feeding habitat preferences) and life strategic characteristics (feeding and breeding guild memberships) of each species.
The sampled species have diverse origin, covering five biogeographical regions. An average of five adults from each species were selected for sampling. Three-to-four scales were collected from the anterior part of the body between the dorsal fin and the lateral line. The removed scales were cleaned and prepared to microscope slides, and after this process they were photographed or scanned. Our database can be used for both research and educational purposes. A large number of scale photos assigned of species can be an opportunity to create an automatic species identification system. Moreover, the subjectivity in morphometric measurements can be eliminated by analysing the database with artificial intelligence.

Specifications Table

Subject	Biodiversity, Zoology, Computer Vision and Pattern Recognition
Specific subject area	Photos of freshwater fish scales and species-specific information of their ecological characteristics.
Data format	Raw, Analysed
Type of data	Image, Table
Data collection	In case of each sampled individuals 3–4 scales from the right anterior trunk area (between the dorsal fin and the lateral line of each sampled individual) were removed and cleared. The fish scales than placed between microscope slides and scanned or photographed. Altogether 2954 scale photos of 193 freshwater fish species were compiled for the database. Beside the scale photos a table is provided containing the following supplementary data of the studied species: relative scale size, body profile index (body fineness), taxonomic position, native realm, Genbank code - if the Citochrome-Oxidase I (COI) subunit is available, flow and feeding habitat preferences and life strategic characteristics (feeding and breeding guild memberships).
Data source location	The fish scales of 193 taxa originated from five biogeographic regions. 110 species from Palearctic, 56 from Neotropical, 12 from Afrotropical, 9 from Nearctic and 6 from Indomalayan regions. The donor individuals were collected from various habitats and their scales are deposited at the coauthors working place's collections. (HUN-REN Balaton Limnological Research Institute, Tihany, Hungary; Laboratório de Ictiologia de Sorocaba, Departamento de Biologia, Universidade Federal de São Carlos, Sorocaba, Brazil; A.N. Severtsov Institute of Ecology and Evolution of the Russian Academy of Sciences - Moscow, Russia).
Data accessibility	Repository name: CONCORDA Data identification number: 1NTPHJ_2024 Direct URL to data: https://hdl.handle.net/21.15109/CONCORDA/1NTPHJ
Related research article	[1] B. Bánó, A. Bolotovskiy, B. Levin, G.M.T. Mattox, M. Cetra, I. Czeglédi, P. Takács, Scale morphology is a promising, additional tool for exploring the taxonomy and ecology of freshwater fishes., Fish Fish. (2024) 1–20. 10.1111/faf.12826.

1. Value of the Data

- This dataset originally was used to reveal how the shape and relative size of fish scales depend on the ecological needs and taxonomic position of the freshwater fish [1].
- The presented database we share the original photos analysed in the above mentioned work. In this case anyone can check the results of our statistical analyses.

- This database can be used for educational purposes, especially in the field of zoology or biostatistics.
- Moreover, scale photos are suitable for training artificial intelligence as well as computer vision and pattern recognition software [2], thus it facilitates the creation of automatic species recognition software [3]. This kind of data recordings (photo analyses) allows artificial intelligence to eliminate the bias caused by the human subjectivity during morphometric measurements.
- To the scale database a supplementary table added to the photo database provides essential taxonomical and ecological information about the studied species. This table summarizes the results of an extensive data collection effort to gather as much information as possible on the fish species studied. The database contains both our own results (e.g. relative scales size, profile index data) and those reported in the literature (e.g. flow and feeding habitat preferences, spawning and feeding guilds).
- These data and images can help other researchers to complement or even compare their morphometric and or ecological datasets of other fish species, other populations, or species with other environmental needs also.

2. Background

Morphological data measurements - especially the conversion of shape into numerical data - is not a simple process. In doing so, we are still usually relying on the human eye and practical experiences, therefore our measurements are often subjected to the measurers' effects [4]. For example, when comparing the results of morphometric measurements made by two independent measurers, the differences detected are maybe artefacts and due only to the different measurement modes, precision so called the subjectivity of the measurers [5,6]. There are several ways to reduce this impact (e.g., choose easily measurable variables, change in the applied methods, etc.). But the most effective one is to compare measurements taken by the same person, although measurement errors cannot be excluded in this case either. The above-mentioned problems can be further moderated if the human factor is as largely excluded from the analysis when possible. For example, this subjectivity can be eliminated by using machine learning and artificial intelligence.

In this present work a database was made publicly available containing 2954 scale photos taken from 193 freshwater fish species. This database originally was used to reveal the usability the scale shape and relative sizes of freshwater fish for taxonomical and ecological purposes [1].

By publishing the analysed photos, we give you full access to the raw data, thus our morphometric measurements are now fully reproducible. Additionally other uses of this database are also possible. Beside others it can be used for research and educational purposes. Moreover, this database could be used to advance machine learning, automating the identification of species and eliminating human influence entirely. Beside the raw data - the scale images itself – we publish some other characteristics, such as the average body profile index, the relative scale sizes and several ecological characteristics (flow and feeding habitat preferences, reproductive and feeding guilds) for each freshwater fish species used in our original research article. Thus, the publicly available knowledge on the analysed species is further extended, and we will also provide information on a number of data deficient species.

3. Data Description

The database contains 2954 photos of 193 freshwater fish species scales. Each photograph has a unique code that can be used to infer the name of the species, the individual within the species, and the sampled scale within the individual (Fig. 1). The table in the database contains the unique code, scientific name, order, family and native realm of each species. Additionally the number of sampled scales, the body profile index, the average scale profile index, and the

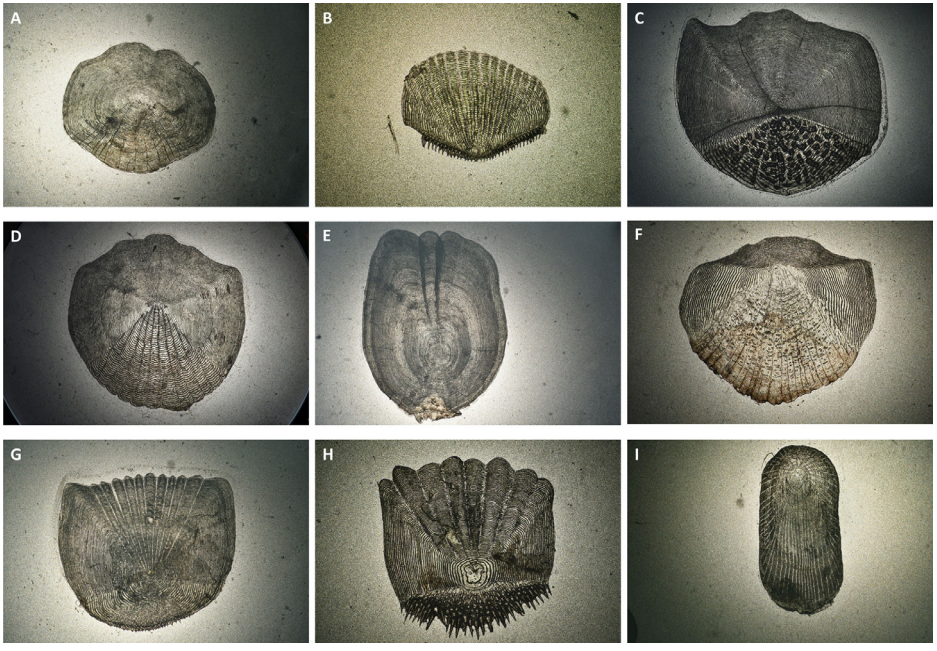

Fig. 1. Sample images of fish scales. A: Abrbra_03_02, B: Babgym_01_04, C: Cargib_01_02, D: Chonas_03_02, E: Esoluc_01_02, F: Gobdéli_01_04, G: Lepgib_03_01, H: Sanvol_01_03, I: Tintin_05_03. The unique code of the images refers to the species itself, the sampled individual within the species, and the certain scale's number sampled from the same within the individual. (e.g. Abrbra_03_02 image belongs to *Abramis brama* species, second scale of the third individual).

relative height and width of scales of each species can be found here. These latter values were measured on the scales, and these raw values were divided by the Standard Length (SL) or the height (H) of the certain donor individual. The relative scale size can be found not only as average by species, but also individually by scale on the another worksheet of the table. Additionally the Genbank Accession number (mtCOI) and ecological characteristics of each species such as flow preference, feeding habitat, reproductive guild, feeding group were also included in the table in case of species where these features were available.

4. Experimental Design, Materials and Methods

The experiment was designed to create a database of freshwater fish scales as large as possible. The database was built using the scales sampled from conserved fish individuals deposited in museum collections (Natural History Museum, Vienna, Natural History Museum Budapest). Permission numbers: 748/1/2023/FŐIG, 748–2/2023/FŐIG. After data imaging the sampled scales were returned to the museum collections.

Additionally several samples are originated from the fish collections of the authors' workplaces. Our own collections were made using electric fishing gears. The permisson numbers for fish collections: HAGF/69/2022 and HAGF/79/2022. The extensive collection work has resulted in data on a total of 193 freshwater fish species inhabiting five biogeographical regions. Although most of the sampled taxa are valid species, but in the case of gudgeons (Cyprinidae, *Gobio*), some cryptic species [7–9] were also included in the database. For Neotropical taxa where the species identification was not clear, we only genus level was indicated.

An average of five adults from each species/taxon were selected for sampling. To compute the relative scale size and the body profile index we measured the standard length (SL) and

the body height (H) for every sampled individual. It used three-to-four scales collected from the anterior part of the body between the dorsal fin and the lateral line. We chose this area because the it is commonly accepted in ichthyology for purposes of age identification by reason; herewith, scales stored in archival scale books can be used for comparative purposes worldwide [10]. The fish scales were put between microscope slides then we photographed or scanned. On the microscope slides, we also recorded the unique codes of the species, as well as the body parameters belong to the sampled individuals (H, SL).

Since the people involved in compiling the database had different tools at their disposal, the scale photos were taken according to three methods. The scales were enlarged by Mitutuyo JP300 profile projector then photographed by a NIKON D500 camera with a 40 mm NIKON macro objective. In other cases, an Epson Perfection V600 black light scanner and a stereomicroscope (Zeiss Discovery V20, Leica M165C) mounted cameras were used to scanned/photographed the scales.

Every photo has a unique code (indicated in the file name), with which it can be accurately identified. The height and width of the scale were measured from the images by using the ImageJ program [11]. Since as the body parameters (H, SL) were recorded, we were able to calculate a body profile index (body fineness) and the relative scale sizes also. Literature data eg. [12] and online databases [13,14] were used to determine the exact taxonomic classification and ecological characteristics of each species.

Limitations

None.

Ethics Statement

The authors have read and follow the ethical requirements for publication in Data in Brief and confirming that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms. No ethic permission was needed for the sampling of conserved fish individuals deposited in fish collections.

Data Availability

Scale photo database of freshwater fishes (Original data) (CONCORDA).

CRediT Author Statement

Bálint Bánó: Conceptualization, Investigation, Formal analysis, Writing – original draft, Visualization; **Aleksey Bolotovskiy:** Investigation, Writing – review & editing; **Boris Levin:** Investigation, Writing – review & editing; **George M.T. Mattox:** Investigation, Writing – review & editing; **Mauricio Cetra:** Investigation, Writing – review & editing; **István Czeglédi:** Formal analysis, Software, Writing – review & editing; **Péter Takács:** Conceptualization, Investigation, Formal analysis, Writing – original draft, Visualization, Supervision.

Acknowledgements

The research presented in the article was carried out within the framework of the Széchenyi Plan Plus program with the support of the RRF 2.3.1 21 2022 00008 project. Bálint Bánó was

supported by the ÚNKP-22–3 New National Excellence Programme of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund. Péter Takács supported by the OTKA FK 140902 project. István Czeglédi was supported by the OTKA PD 138296 grant. The research was funded by the Sustainable Development and Technologies National Programme of the Hungarian Academy of Sciences (FFT NP FTA, NP2022-II3/2022) and the OTKA K145933. Boris Levin and Aleksey Bolotovskiy were supported by the Russian Science Foundation (grant no. 24–44-20019). George M.T. Mattox was supported by FAPESP 2017/01970-4. We are very grateful to the following colleagues for their help with the sample collections: Ádám Staszny, Márk Liziczai (Kossuth Lajos Gimnázium, Mosonmagyaróvár), Előd Márton (Natural History Museum, Budapest), Frederic Marchand (Biological Resource Centre, Colisa), Anja Palandacic (Natural History Museum, Vienna), Gyula Pasaréti (Akvárium Magazin), Pavel Vrána (Czech Anglers Union).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] B. Bánó, A. Bolotovskiy, B. Levin, G.M.T. Mattox, M. Cetra, I. Czeglédi, P. Takács, Scale morphology is a promising, additional tool for exploring the taxonomy and ecology of freshwater fishes, Fish Fish (2024) 1–20, doi:10.1111/faf.12826.
[2] L. Shamir, J.D. Delaney, N. Orlov, D.M. Eckley, I.G. Goldberg, Pattern recognition software and techniques for biological image analysis, PLoS Comput. Biol. 6 (2010), doi:10.1371/journal.pcbi.1000974.
[3] J.G.A. Barbedo, A review on the use of computer vision and artificial intelligence for fish recognition, monitoring, and management, Fishes 7 (2022), doi:10.3390/fishes7060335.
[4] S.M. Yezerinac, S.C. Lougheed, P. Handford, Measurement error and morphometric studies: statistical power and observer experience, Syst. Biol. 41 (1992) 471–482.
[5] M.V. Mina, B.A. Levin, A.N. Mironovsky, On the possibility of using charachter estimates obtained by different operators in morphometric studies of fish, J. Ichthyol. 45 (2005) 284–294.
[6] P. Takács, Z. Vitál, Á. Ferincz, Á. Staszny, Repeatability, reproducibility, separative power and subjectivity of different fish morphometric analysis methods, PLoS ONE 11 (2016) 1–16, doi:10.1371/journal.pone.0157890.
[7] J. Mendel, S. Lusk, E.D. Vasil'eva, V.P. Vasil'ev, V. Lusková, F.G. Ekmekci, F. Erk'akan, A. Ruchin, J. Koščo, L. Vetešník, K. Halačka, R. Šanda, A.N. Pashkov, S.I. Reshetnikov, Molecular phylogeny of the genus Gobio Cuvier, 1816 (Teleostei: Cyprinidae) and its contribution to taxonomy, Mol. Phylogenet. Evol. 47 (2008) 1061–1075, doi:10.1016/j.ympev.2008.03.005.
[8] P. Takács, G. Maasz, Z. Zrínyi, N. Boross, Z. Vitál, D.I. Kánainé Sipos, B. Bánó, Á. Staszny, P. Sály, K. Balázs, Infirm effect of phylogeny on morphometric features in a cryptic Gobio species complex, Contrib. Zool. 16 (2022) 1–18, doi:10.1163/18759866-bja10026.
[9] L. Zangl, D. Daill, W. Gessl, T. Friedrich, S. Koblmüller, Austrian gudgeons of the genus Gobio (Teleostei: gobionidae): a mixture of divergent lineages, J. Zool. Syst. Evol. Res. 58 (2020) 327–340, doi:10.1111/jzs.12340.
[10] A.L. Ibáñez, P. O'Higgins, Identifying fish scales: the influence of allometry on scale shape and classification, Fish. Res. 109 (2011) 54–60, doi:10.1016/j.fishres.2011.01.016.
[11] W.S. Rasband, ImageJ: Image Processing and analysis in Java, 2012 ascl, ascl-1206.
[12] P. Sály, T. Erős, Vízfolyások ökológiai állapotminősítése halakkal: minősítési indexek kidolgozása= Ecological assessment of running waters in Hungary: compilation of biotic indices based on fish, Pisces Hungarici 10 (2016) 15–45.
[13] URL1, Eschmeyer's Catalog of Fishes, (n.d.). https://www.calacademy.org/scientists/projects/eschmeyers-catalog-of-fishes Data (Accessed 1 February 2024).
[14] URL2, seriouslyfish.com, (n.d.). https://www.seriouslyfish.com/Data (Accessed 1 February 2024).