

cnfishbase: A cyber Chinese fish database

Fish constitute the largest group among vertebrates, accounting for more than half of all vertebrate species (Nelson et al., 2016), with new fish described each year. In the first edition of *Fishes of the World*, Nelson (1976) recorded 18 818 fish species from 450 families, which increased to 21 723 fish species from 445 families in the second edition (Nelson, 1984), 24 618 fish species from 482 families in the third edition (Nelson, 1994), 27 977 fish species from 515 families in the fourth edition (Nelson, 2006), and 32 000+ fish species from 536 families in the fifth edition (Nelson et al., 2016). According to the latest data from the FishBase database (March 2023), there are currently over 35 000 fish species belonging to 623 families worldwide.

The substantial diversity of fish species presents a significant obstacle for domestic Chinese researchers in the field of ichthyology in acquiring both the scientific and Chinese nomenclature of fish species. Although various texts on fish taxonomy have been published in China over the last century (Zhang & Cao, 2021), many are now outdated or incomplete and lack adequate mechanisms for timely updates, thus hampering their application in contemporary fish biology research. Furthermore, the early publication times and reliance on traditional morphological classification methods in these previous publications fail to accurately reflect the current diversity and phylogenetic position of many fish species.

Books serve as a simple and intuitive medium for the storage and preservation of small-scale data that do not require frequent modifications or updates. However, in biological directories where the classification status of organisms is subject to frequent changes, books present challenges for timely updates due to the high cost of reprinting or republishing. Some biological directories, which include numerous species and large-scale data, may go many years without timely updates or the incorporation of classification system revisions. Thus, the emerging trend is toward the digitization of biological directories. Utilizing databases for data storage offers several distinct advantages over conventional books: (1) Expedited access and retrieval: Databases enable rapid information retrieval through the application of query languages, which is vastly more efficient than browsing books. (2) Simultaneous multi-user support: Databases enable access for multiple users at the same time without data conflicts or errors. (3) Data persistence: Databases allow for ongoing storage without loss. (4) Data dissemination and collaboration: Databases facilitate the sharing of data and collaborative scientific inquiry. In the

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2023 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

context of biological directories, which require frequent updates and changes, the costs of modifying information are reduced, and data retrieval can be achieved rapidly.

The sharp increase in biological information has led to greater demand for database construction (Ball-Damerow et al., 2019). A major consideration in biodiversity science is the need to convert information accumulated over many years into an electronic format. Several international biodiversity databases have been created, each with varying purposes, content, and features. Although biodiversity is a global phenomenon, most databases dedicated to biodiversity have been established in developed countries (Shanmughavel, 2007). As such, many Chinese scientists have been actively engaged in the construction of domestic biodiversity databases. For example, the first iteration of the *Catalogue of Life: China* (CoL China) was released in 2008, with annual updates thereafter (Jiang et al., 2015). These catalogues, providing essential information about species names, taxonomic relationships, and distributions within China, serve as critical resources for advancing our understanding of the origins, evolution, and biodiversity conservation of such species (Mi et al., 2021).

Considering the above, we developed and released the first version of a Chinese fish database (cnfishbase, <https://cnfishbase.cn/>, Figure 1). The database was developed using the Python Django framework for web back-end functionality and the Bootstrap framework for web front-end display. MySQL was used for data storage and management, while the pyecharts package was used for fish data visualization. Notably, cnfishbase features the following sub-domains: (1) World Fish Database: This sub-domain incorporates information on 35 137 fish species belonging to 623 families and 93 orders obtained from the FishBase database (current as of March 2023; <http://www.fishbase.org/>), resulting in the inclusion of 30 463 Chinese names and 16 940 commonly used English names of fish species. Users can enter the name of the fish of interest in the search bar located at the top of the website to access relevant information. Upon selecting the FishBase link corresponding to a specific fish, users can retrieve more detailed information regarding the morphology, distribution, taxonomy, ecology, fishery, and aquaculture of that species. (2) Information by Country: This sub-domain encompasses a summary of fish statistics relating to 215 countries and regions around the globe. It offers a convenient means for users to ascertain the fish composition of each individual country or region. (3) Distribution Map: This sub-domain contains the fish distribution maps of 33 641 species derived from the fish data relating to 215 countries and regions, thus covering more than 95% of fish species. The

Received: 24 April 2023; Accepted: 17 May 2023; Online: 12 September 2023

Foundation items: This work was supported by the National Natural Science Foundation of China (91731301, 32170438)

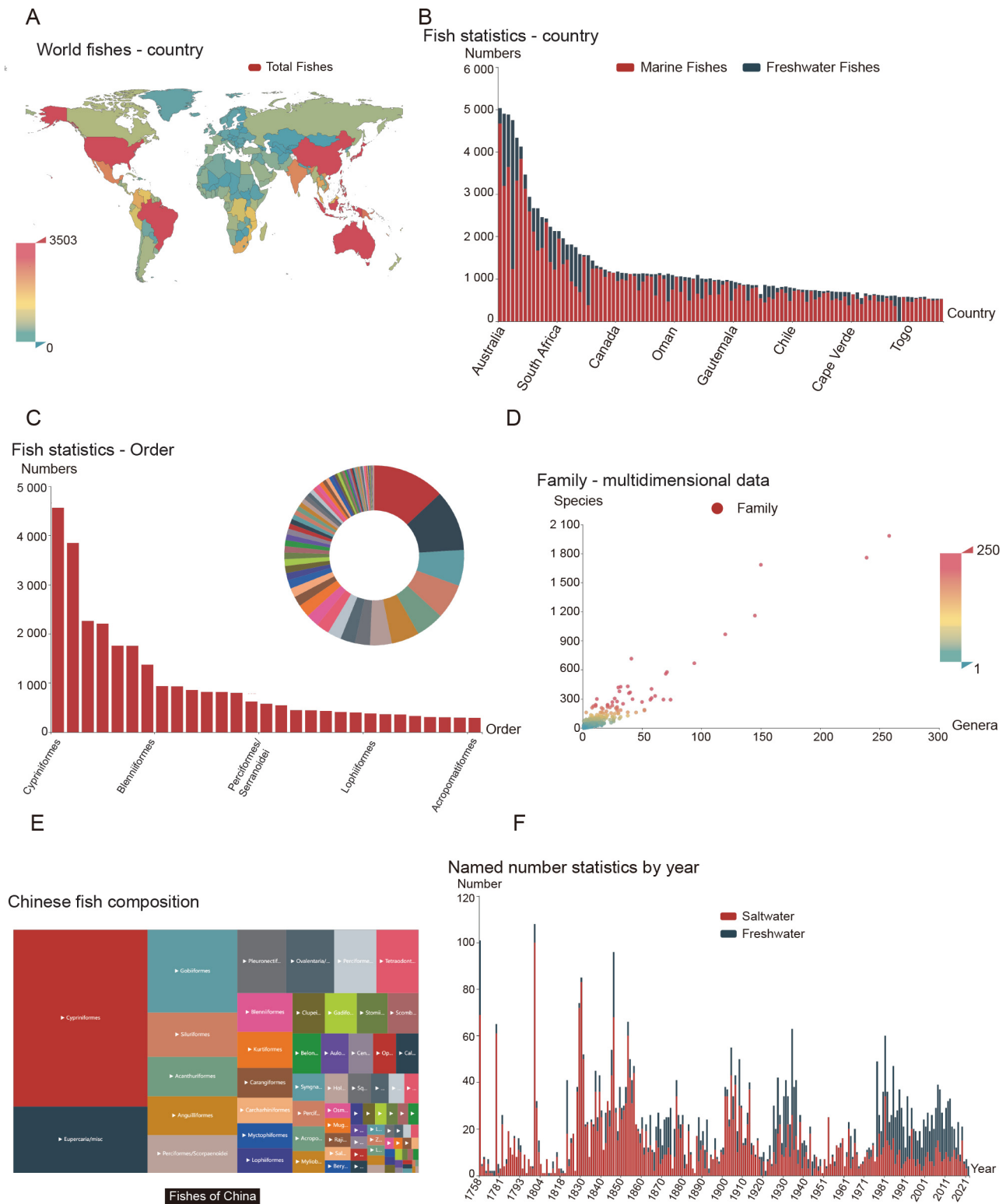


Figure 1 Partial image display of the cnfishbase Chinese fish database

A: Country-level statistics of world fish map. B: Stacked column chart of fish statistics by country. C: Order-level statistics of world fish. D: Family-level statistics of world fish. E: Treemap of Chinese fish composition. F: Number of named Chinese fish by year. Above images only show certain chart types from the cnfishbase website. Charts provided by the online cnfishbase database are dynamic and interactive. Users can visit the website to view more information.

database provides global distribution maps of 93 order-level fish and 622 family-level fish. Users can readily observe the geographical distribution patterns of different fish species, an important feature in the field of biogeography. These dynamic, interactive charts represent original contributions by our

research team. (4) Chinese Fish Database: This sub-domain includes data on fish recorded in mainland China, Taiwan, Hong Kong, and Macao based on FishBase records. It uses the latest international classification system to correct the limitations of previous traditional morphological classifications

in China and denotes the habitat types of fish distributed in China and whether they are an introduced species. (5) Interactive Data Visualization: This sub-domain uses pyecharts (<https://github.com/pyecharts>) for fish data visualization under different scenarios, including maps, bar charts, scatter plots, and pie charts. Pyecharts supports various interaction methods, such as zooming, scrolling, and hovering, to enhance data exploration and analysis. The fish distribution map includes a link function, where users can click on the title to access a new page with detailed fish information. Dynamic images can also be conveniently saved locally through two methods: the first is to right-click the mouse to save static images in PNG format, the second is to press “Ctrl + S” to save the HTML file, preserving dynamic interaction functionality.

Compared to traditional fish reference materials in China, our database offers the following advantages: (1) Comprehensiveness: FishBase integrates global fish information, including aspects such as taxonomy, distribution, morphology, ecology, biology, and fisheries. However, the content within FishBase is primarily presented in English, which can pose difficulties for users unfamiliar with the language, especially non-professionals. To address this issue, our database provides links to FishBase based on Chinese fish names. This enables users to utilize Chinese names to search for specific fish and subsequently access FishBase for more detailed information. Our database includes nearly all known fish species worldwide, thereby facilitating ichthyological research and teaching. Additionally, it offers a country-specific summary page, simplifying the process for users to access data for individual countries. (2) High-quality data: The information provided in FishBase is derived from diverse sources, including scientific literature, museum specimens, survey data, and expert review, thus ensuring the reliability and accuracy of the data. Similarly, cnfishbase employs the FishBase classification system. Notably, the Chinese names of species adhere to the translation standards found in various resources, including the *Latin-Chinese Dictionary of Fish Names by Classification System* (Wu et al., 2017), *China's Red List of Biodiversity: Vertebrates, Volume V, Freshwater Fishes* (Zhang & Cao, 2021), and *Catalogue of Life China 2022 Annual Checklist* (<http://www.sp2000.org.cn/>). (3) Interface accessibility: The cnfishbase database features a simple and user-friendly interface, facilitating easy access and analysis of fish data for users. The interface dynamically adapts to the different screen sizes of desktop computers, laptops, tablets, and mobile devices. At present, interactive distribution maps of more than 33 000 fish species are available, in both Chinese and English, enabling users can view the global distribution patterns and trends of a vast number of fish. Users can engage dynamic effects via cursor movement, access detailed information by title selection, and locally save images using the right-click function. (4) Novelty: The database uses the latest international fish classification system, better reflecting the phylogenetic relationships among fish compared to traditional morphological classification. Additionally, the database encompasses nearly all extant fish species, maintaining current and comprehensive representation. Notably, to accommodate the annual discovery of new species, the database undergoes yearly updates to ensure data accuracy.

The cnfishbase Chinese fish database is targeted to the following users: (1) Ichthyology researchers: FishBase serves

as an essential resource for ichthyologists, offering comprehensive data and graphical representations pertaining to fish biology, ecology, taxonomy, and distribution. These resources facilitate research and analysis, contributing to an enhanced understanding of fish diversity and adaptability. However, given the extensive number of fish species, it can be difficult to recall their Latin names and corresponding Chinese names. To address this issue, cnfishbase provides the latest information on fish classification across various taxonomic levels, including order, family, genus, and species, complete with scientific and Chinese names. This can assist those working in the field of ichthyology to better understand the classification and evolutionary relationships of fish. Furthermore, integration with FishBase provides quick access to specific fish details, thus enhancing research efficiency. (2) Biogeography researchers: Traditional approaches to visualizing global fish distribution trends have been limited, often lacking intuitive representations. Thus, based on biogeographical research, we developed interactive distribution maps of global fish. We used native species, excluding introduced species, to establish distribution maps and show the original status of a certain fish in various countries. The generated maps cover most fish species, providing a clear and intuitive view of the distribution patterns of fish species on a global scale. These maps should serve as a valuable tool for professionals working in the field of biogeography. (3) Ichthyology students and teachers: The cnfishbase Chinese fish database can serve as a valuable resource for education in ichthyology, ecology, and fishery management. With most domestic textbooks only available in Chinese, the lack of familiarity among students and teachers with Latin nomenclature may present a barrier to learning and teaching. Moreover, existing textbooks may not be current due to the high cost of rewriting and inadequate mechanisms for timely updating, which can hinder learning. In contrast, with its extensive and up-to-date information, the cnfishbase database provides teachers and students with resources for better understanding current fish classification as well as fish diversity and importance. (4) Fish enthusiasts and the public: The cnfishbase database provides comprehensible and readily accessible data and information, serving as an excellent tool for the dissemination of scientific knowledge and environmental awareness. Individuals can use the database to gain insights into fish ecology, diversity, and conservation, thus contributing to the promotion of environmental protection and sustainable development.

Currently in its first version, the cnfishbase database will be continually updated, with the integration of additional functions. Future enhancements will include: (1) introduction of a national search function, allowing users to view fish data for each country; (2) aggregation of fish data for each province in China, facilitating accessibility for domestic users; and (3) development of a fish image upload function for users, allowing the accumulation of image data on Chinese fish species and facilitating species identification and reference.

DATA AVAILABILITY

Data and graphs related to this article are available at cnfishbase (<https://cnfishbase.cn/>).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

AUTHORS' CONTRIBUTIONS

Y.R.L. and S.P.H. conceived the study and wrote the manuscript. C.C.F reviewed the manuscript. Y.R.L. collected and processed the data, wrote all the code, and built the database and website. All authors read and approved the final version of the manuscript.

COMPETING INTERESTS

The authors declare that they have no competing interests.

Yong-Rui Lu^{1,2}, Cheng-Chi Fang¹, Shun-Ping He^{1,3,*}

¹ State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei 430072, China

² University of Chinese Academy of Sciences, Beijing 100049, China

³ Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, Yunnan 650223, China

*Corresponding author, E-mail: clad@ihb.ac.cn

REFERENCES

- Ball-Damerow JE, Brenskelle L, Barve N, et al. 2019. Research applications of primary biodiversity databases in the digital age. *PLoS One*, **14**(9): e0215794.
- Jiang ZG, Qin HN, Liu YN, et al. 2015. Protecting biodiversity and promoting sustainable development: in memory of the releasing of catalogue of life China 2015 and China biodiversity red list on the international day for biological diversity 2015. *Biodiversity Science*, **23**(3): 433–434. (in Chinese)
- Mi XC, Feng G, Hu YB, et al. 2021. The global significance of biodiversity science in China: an overview. *National Science Review*, **8**(7): nwab032.
- Nelson JS. 1976. Fishes of the World. New York: John Wiley & Sons.
- Nelson JS. 1984. Fishes of the World. 2nd ed. New York: John Wiley & Sons.
- Nelson JS. 1994. Fishes of the World. 3rd ed. New York: John Wiley & Sons.
- Nelson JS. 2006. Fishes of the World. 4th ed. New York: John Wiley & Sons.
- Nelson JS, Grande TC, Wilson MVH. 2016. Fishes of the World. 5th ed. New York: John Wiley & Sons.
- Shanmughavel P. 2007. An overview on biodiversity information in databases. *Bioinformation*, **1**(9): 367–369.
- Wu HL, Shao GZ, Lai CF, et al. 2017. Latin-Chinese Dictionary of Fish Names by Classification System. Qingdao: China Ocean University Press. (in Chinese)
- Zhang E, Cao WX. 2021. China's Red List of Biodiversity: Vertebrates, Volume V, Freshwater Fishes. Beijing: Science Press. (in Chinese)