

Catalogue of life China: Towards an index of known species present in China

Congtian Lin,^{1,4} Bing Liu,² Mingjun Zhao,³ Keping Ma,^{2,*} and Liqiang Ji^{1,*}

¹Scientific Data Center, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

²State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

³State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

⁴National Basic Science Data Center, Beijing 100190, China

*Correspondence: ji@ioz.ac.cn (L.J.); kpma@ibcas.ac.cn (K.M.)

Received: February 4, 2025; Accepted: May 24, 2025; Published Online: May 30, 2025; <https://doi.org/10.59717/j.xinn-life.2025.100141>

© 2025 The Author(s). This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

Citation: Lin C., Liu B., Zhao M., et al. (2025). Catalogue of life China: Towards an index of known species present in China. *The Innovation Life* 3:100140.

THE NEED FOR A COMPREHENSIVE CHINESE SPECIES CHECKLIST

The system of species nomenclature and classification introduced by Linnaeus since 270 years ago remains central to all efforts to describe, document and inventory the world's biodiversity.¹ In China, scientists have spent more than 100 years to inventory species, and have made remarkable achievements in this field to answer what and how many species we have. The most important achievements are the serial publications of Fauna of China, Flora of China and Cryptogamic Flora of China, which started since about 60 years ago.² But, As we know, taxonomy is a science advancing with

the times, especially when the DNA technology has been used in this field and speed up its progress. Many classification systems of certain taxonomic groups of China have been updated according to the results of new studies. On the other hand, China is one of the countries with the richest biodiversity in the world, and a large number of new species have been discovered every year which are needed to be integrated together for usage with many efforts. So it is an urgent and major task to build a comprehensive, accurate, dynamic and national wide species checklist for China with inbuilt mechanisms for maintenance and updating by using information technology. This work is of



Figure 1. Overview of CoL-China 2025 (A) Cover image of CoL-China 2025. (B) Species number of different taxon groups in CoL-China 2025. (C) Species number in annual CoL-China from 2008 to 2025. (D) species richness in different provinces based on distribution data in CoL-China 2025.

great significance in the following aspects: (1) An authoritative species checklist provides an index for accessing to Chinese species knowledge, and plays a significant role throughout the biological sciences and biodiversity research for giving the accurate context and meaning for those organisms. (2) In such a big data and artificial intelligent era, a comprehensive species checklist is usually adopted as backbone to index species data from various sources for building high quality biodiversity databases, which are critical for training state-of-the-art AI models. (3) Comprehensive species checklists are also utilized as tools for communication about species internationally for trade, conservation, or bio-security. They can help governments and inter-governmental bodies to draw up treaties, legislation, and regulations, like the CITES listings for controlling trade endangered species. (4) Furthermore, an inventory of species is required by The Convention on Biological Diversity for each member states to fulfill their obligations. Biodiversity Conservation Strategy and Action Plan (2023-2030) (BCSAP) of China also clearly proposed the target of updating national species checklist annually.

These needs and challenges are the driving forces that led governments and scientific communities of China to carry out the plan for building a comprehensive and authoritative species Checklist: Catalogue of Life, China (CoL-China).

Progress of CoL-China

Toward building a global comprehensive species checklist, Species 2000 was initiated in 1996 by several international organizations, and sought partnerships from all over the world to create global Catalogue of Life.³ China joined in Species 2000 as one of its regional nodes in 2006 to start compiling CoL-China. The goal of CoL-China is to collate a uniform and validated index for Chinese known species, and provide the checklist annually for global access to biodiversity knowledge of China.

In 2008, the first annual version of CoL-China was released, which started a new stage for biodiversity informatics in China. The number of species documented in CoL-China has been growing steadily from 44,159 to 148,341

between 2008 and 2025. In CoL-China 2025 version, there are 170 classes, 706 orders, 3,473 families, 148,341 species and 14,376 infra-species, including 145,162 synonyms, 40,303 common names, 149,310 distribution items and 207,355 references. It has covered seven kingdom level taxa (Animalia, Bacteria, Chromista, Fungi, Plantae, Protozoa and Viruses) and encompassed all known higher plant species (39,660) and vertebrate species (8,778).

CoL-China has a strict taxonomic data quality control process to ensure its authority. It complies the data standard delivered by Species 2000, and extends it from 8 items to 12 items for accommodating to real practices in China.⁴ The data items of each species includes: accepted scientific name with references, Chinese formal name formally used in government documents and publications, Pinyin for Chinese formal name, classification in Chinese and English, synonyms with references which imply relationship between taxa in different taxonomic context, common names with references, distribution data in Chinese and English which are normalized into province list, latest taxonomic scrutiny information of when and by whom the species was last reviewed, source database information attached to each species record for tracking where it is from, confidence score an indicator of data quality, Chinese Science and Technology Resource Identifier (CSTR) as universal unique identifier of each species for data connection and exchange, and comment field for additional information.

Building CoL-China is a tough task that requires a broad network of taxonomists and IT experts to work together rather than single effort. More than 300 specialists from 120 institutions are involved in this project. They need to follow the new progress of taxonomy on the taxa groups which they manage, and to track the nomenclatural or taxonomic changes for revising and updating species checklist annually. A taxonomic tree tool for comparing and visualizing different classifications was developed to help experts to reconcile older nomenclatural records with more recent taxonomic revisions. To Ensure all fields are filled correctly, a validation system with 28 steps is developed for producing CoL-China from checking spelling errors of various names, classification accuracy to filed completeness. Each species page is reviewed by taxonomists, and confidence score - a scoring mechanism is introduced into CoL-China with 1-5 stars indicating different ranks of data quality.

Application of CoL-China

CoL-China provides three ways for access. Users can visit the website (www.sp2000.org.cn) to browse and search species data (Taxonomic Tree for Searching: http://sp2000.org.cn/browse/browse_taxa, Species Page: http://sp2000.org.cn/species/show_species_details/c9c43b00833c412884d060f98bfd0cbd). It provides a function for matching names of users' checklists with itself in batch (Species Name Match Tool: http://sp2000.org.cn/names_match_services). There are rich application programming interfaces for users to get species information by program automatically (Application Programming Interfaces for Accession to CoL-China: <http://sp2000.org.cn/api/document>). In case the internet is unavailable, a single standalone application was developed, which has a full copy of database and can run in Windows and Linux system locally.

CoL-China has been widely adopted because of its authority and high quality. It has received over 20 million visits since 2018, and has been cited over 1200 times in publications relative to biodiversity. CoL-China plays a key role in many government sectors relative to biodiversity in China. As we know, China is the only country in the world which publishes species Checklist annually. That supports the government to fulfill its obligations of CBD. It is the main data source of the China Ecological Environment Status Bulletin and listed as reference database in the guide of biodiversity survey, both of which are published by Ministry of Ecology and Environment (MEE). It was used as tool for species assessment, and helps the government to publish the national species red list, and to update the checklist of the state key wild

animals and plants of China. It also pushed the development of an AI-based biodiversity monitoring system which has been applied by China Meteorological Administration and MEE for long-term monitoring spatial and temporal dynamics of biodiversity required by BCSAP. Not only for the government, CoL-China also supports many companies and federations like Tencent and Tiktok to carry out their public welfare career to push biodiversity mainstreaming. CoL-China is playing an important role in global biodiversity knowledge communication. It is one of the supplementary data sources of CoL from regional nodes. It was also published in GBIF and The Linking Open Data Cloud for broader access and usage from all over the world.

Future plan of CoL-China

Although CoL-China has released 18 versions and included more than 162,000 species and infra-species, there are still many gaps, especially for the insect and marine invertebrate species. To fill these gaps, more taxonomists and data sources should be mobilized. An action to build species checklist repository is carrying out, which is intended to collect different checklists from taxonomists and various databases as candidate data sources. On the other hand, we should strengthen the connection between CoL-China and other global databases (e.g. GBIF, CoL) to mobilize more data sources and experts to involved into this project, since many taxon groups are lack of taxonomists in China.

As we know, the meaning of a species scientific name is changed when taxonomic actions (like merging multiple species into one or splitting one species to many) revise the taxon concept.⁵ That always makes people confused in real practices like species conservation, species trade and so on, especially when multiple classifications coexist. So, it is necessary to develop a system based on CoL-China to connect taxa from different species checklists to create taxonomic knowledge graph for tracking taxon concepts in next step.

Building CoL-China each year is an arduous task and involves many efforts. The emergence of Large Language Model (LLM) provides a good opportunity to optimize the workflow. We are trying to create a serial of agents based on LLM for correcting the miss-spelled scientific names, extracting distribution data from references, and comparing taxon concepts with the same name and so on. It will accelerate the process of producing CoL-China and achieve the goal of building a uniform checklist as an index of known Chinese species.

REFERENCES

1. Hobern D., Barik S.K., Christidis L. et al. (2021). Towards a global list of accepted species VI: The catalogue of life checklist. *Org. Divers. Evol.* **21**:677-690. DOI:10.1007/s13127-021-00516-w
2. Keping Ma. (2015). Species Catalogue of China: a remarkable achievement in the field of biodiversity science in China. *Biodiv. Sci.* **23**:137-138. DOI:10.17520/biods.2015077
3. Bisby F. A. (2000). The quiet revolution: biodiversity informatics and the internet. *Science* **289**:2309-2312. DOI:10.1126/science.289.5488.2309
4. Cachuela-Palacio M. (2006). Towards an index of all known species: the catalogue of life, its rationale, design and use. *Integr. Zool.* **1**:18-21. DOI:10.1111/j.1749-4877.2006.00007.x
5. Jones A. C., White R. J. and Orme E. R. (2011). Identifying and relating biological concepts in the Catalogue of Life. *J. Biomed. Semant.* **2**:7. DOI:10.1186/2041-1480-2-7

FUNDING AND ACKNOWLEDGMENTS

This work was financially supported by Biological Resources Programme of Chinese Academy of Sciences and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA19050202). The funder had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We appreciate all the experts contributing their data and efforts to this project.

DECLARATION OF INTERESTS

The authors declare no competing interests.