Ecological Indicators 137 (2022) 108732

Contents lists available at ScienceDirect

Ecological Indicators

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

Original Articles

Gap analysis for DNA-based biomonitoring of aquatic ecosystems in China

Feilong Li [a,b], Yuan Zhang [a,*], Florian Altermatt [c,d], Xiaowei Zhang [e,*], Yanpeng Cai [a,b], Zhifeng Yang [a,b]

[a] Guangdong Provincial Key Laboratory of Water Quality Improvement and Ecological Restoration for Watersheds, School of Ecology, Environment and Resources, Guangdong University of Technology, Guangzhou 510006, China

[b] Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, China

[c] Department of Aquatic Ecology, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland

[d] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

[e] State Key Laboratory of Pollution Control & Resource Reuse, School of the Environment, Nanjing University, Nanjing 210023, China

ARTICLE INFO

Keywords:
DNA barcoding
Environmental DNA
Freshwater fish
Aquatic insects
Molluscs

ABSTRACT

DNA-based taxon identification is improving the assessment and management of biodiversity in rivers. However, the lack of comprehensive DNA barcode reference libraries and globally highly unequal coverage are still hindering the application prospects of this method worldwide. Here, we analyzed the COI barcode gap in two reference libraries, Barcode of Life Data Systems (BOLD) and NCBI GenBank, with a focus on three aquatic animal groups (freshwater fish, aquatic insects and molluscs) in Chinese rivers. Our data show gaps in barcode coverage (e.g., organisms without barcodes) of *ca.* 40–70% of taxa in these groups in the BOLD or NCBI GenBank database, respectively. These gaps can rise even further if the barcode thresholds are set to contain at least five reference sequences per taxon. Furthermore, most barcodes are from non-local samples, and only 14.4% (BOLD) and 28.8% (NCBI GenBank) of reference sequences were from organisms sampled in China, respectively. The pairwise genetic distance of local barcodes is 3 to 5 times lower than non-local barcodes, indicating that the latter may not be a good substitute. When looking at individual catchments, *ca.* 60% of the potentially occurring aquatic species have one or more barcodes, yet the barcode coverage varies slightly across ten major river catchments, ranging from 54.3% (Liao River basin) to 68.2% (Huai River basin). The taxa Salmoniformes and Perciformes in freshwater fish, Odonata and Diptera in aquatic insects, and Bivalvia in molluscs have the best barcode coverage in most catchments (mean coverage >70%). This study gives the first overview and current status of barcode reference libraries of three major aquatic animal groups in Chinese rivers. Our results will help to better interpret current metabarcoding studies from China, and also provide a basis to develop a strategy of filling the gaps in the reference libraries of aquatic species in China.

1. Introduction

A high biodiversity is the foundation of ecosystem services and functions in rivers and other ecosystems (Jetz et al., 2019; Lynch et al., 2020). Unfortunately, rivers are globally seriously affected by humans, which directly threatens water security and species survival (Su et al., 2021; Vorosmarty et al., 2010). Monitoring aquatic biodiversity and its change are relevant to document and mitigate biodiversity loss. DNA-based taxon identification (e.g., DNA barcoding or metabarcoding) provides an unprecedented opportunity to explore life on our planet (Altermatt et al., 2020; Baird and Hajibabaei, 2012; Cristescu and Hebert, 2018; Deiner et al., 2017). This method not only increases the number of taxa that can be studied in a unifying protocol, but also allows them to be studied in immature or larval stages (Hering et al., 2018; Sagova-Mareckova et al., 2021). In particular, the non-invasive biomonitoring attributes of environmental DNA (eDNA) metabarcoding provide a promising way to assess biodiversity without harming organisms (Deiner et al., 2017; Pawlowski et al., 2020), and the collection, processing and metabarcoding of bulk samples allow higher throughput of samples. Recent advances in High-Throughput Sequencing (HTS) technology are accelerating access and the ability for rapid, high-frequency and big data biomonitoring (Beng and Corlett, 2020; Deiner

* Corresponding authors at: Institute of Environmental and Ecological Engineering, Guangdong University of Technology, Guangzhou 510006, China (Y. Zhang). School of the Environment, Nanjing University, Nanjing, 210089, China (X. Zhang).

E-mail addresses: zhang.yuan@gdut.edu.cn (Y. Zhang), zhangxw@nju.edu.cn (X. Zhang).

https://doi.org/10.1016/j.ecolind.2022.108732
Received 19 January 2022; Received in revised form 22 February 2022; Accepted 24 February 2022
Available online 3 March 2022

et al., 2020). The availability and accuracy of DNA-based taxon identification largely depend on the coverage of taxa in the DNA barcode reference libraries and the quality of available barcode records (Leite et al., 2020; Weigand et al., 2019). However, due to the lack of comprehensive DNA barcode reference libraries, sequences cannot be annotated, limiting their use in biomonitoring and seriously affecting the accurate understanding of species occurrence.

Mitochondrial cytochrome *c* oxidase subunit 1 (COI) has been widely used for the barcoding of animal species, and is thus also used in monitoring and biodiversity assessment (Andujar et al., 2018; Leray et al., 2019; Porter and Hajibabaei, 2018). Importantly, also other barcode regions (e.g., 12S or 16S genes) have been postulated, and may be for some taxa even better, to be used for metabarcoding. However, due to its early use as a barcode region, COI is best-covered in the databases, and thus suitable for an analysis of "best" coverage. Currently, more than 8.3 million and 3.7 million COI reference sequences have been recorded in either the BOLD or NCBI GenBank databases, respectively (up to March 2021) (Benson et al., 2013; Ratnasingham and Hebert, 2007). Arthropoda (Diptera, Insecta) contain most records, followed by Actinopterygii. Although the number of COI records in both the BOLD and NCBI GenBank databases has increased significantly since 2003, a highly disproportionate global distribution of these records has seriously hindered the prospect of DNA-based taxon identification worldwide (Porter and Hajibabaei, 2018). Canada, Costa Rica and the United States alone account for more than 60% of the COI records of Insecta in the BOLD database, while another 175 countries and regions have less than 1% contributions each, of which 135 countries and regions together even account for only about 1.5% of the total. A recent pan European continental study suggests that the gap in DNA barcode reference libraries of monitored aquatic species is significantly different between taxonomic groups and different countries of the European Union (EU), and *ca.* 60% of taxa have fewer than 5 records/species in the reference libraries (Weigand et al., 2019). It has been noted that this gap must be even more extensive and problematic especially in developing countries and regions worldwide, where we still lack a clear and detailed picture of the coverage of the record, and China may be one of those.

China covers an area of 9.6 million km^2. It has a high biodiversity as well as a high variability of biodiversity across the country, containing many major ecosystem regions (Liu et al., 2003; Lu et al., 2020; Mi et al., 2021). It is estimated that China contains *ca.*1400–1600 species of freshwater fish, accounting for more than 10% of all globally known freshwater fish (He et al., 2020; Zhang et al., 2020), of which >60% species are endemic, such as the enigmatic Chinese sturgeon (*Acipenser sinensis*). In addition, more than 2000 species of aquatic insects of the orders Ephemeroptera, Plecoptera and Trichoptera and 1000 species of molluscs (i.e., Bivalvia and Gastropoda) are listed in the Catalogue of Life China 2020 Annual Checklist (http://www.sp2000.org.cn/) and the Global Biodiversity Information Facility (GBIF) database (https://www.gbif.org/occurrence/search). China not only has a rich biodiversity, but also has exceptional threats to biodiversity due to damming, water pollution and environmental change (Huang et al., 2021; Lu et al., 2020; Mi et al., 2021). Timely and accurate data on the occurrence of these species are thus needed for species protection and conservation. However, the lack of DNA barcode reference libraries undoubtedly restricts the application of DNA-based taxon identification in China. While the use of COI is broadly agreed on for invertebrates, we note that for fish other barcode regions (especially 12S genes) may be more appropriate, yet for a comparison of barcode coverage we can still rely on COI, as a case example of "best-scenario". Currently, there may be two strategies considered to fill the shortcomings and gaps of these databases (McGee et al., 2019; Paz and Rinkevich, 2021; Weigand et al., 2019). One is to retrieve and collate the species' sequences from the current reference libraries (short-term benefits, with the caveat that records of species or taxa may come from other countries); the other is to establish and improve the barcode database with local, indigenous barcodes (long-term benefits). The prerequisite of these two strategies is to have a comprehensive gap-analysis of the sequence records of different aquatic taxonomic groups in Chinese rivers, which hitherto has not been available.

Here, we conducted the first gap analysis of all COI records in two reference libraries, Barcode of Life Data Systems (BOLD) and NCBI GenBank, with a focus on three aquatic animal groups (freshwater fish, aquatic insects and molluscs) in Chinese rivers. We also provided comprehensive and annotated reference libraries for these three groups. The purpose of this study is: 1) to analyze the current status of COI records of freshwater fish, aquatic insects and molluscs in Chinese rivers; 2) to compare the barcode coverage of different aquatic animal groups in two databases (BOLD and NCBI GenBank) and data sources (contrasting local and indigenous with non-local and non-indigenous barcodes); and 3) to reveal the record differences of barcodes in ten major river catchments of China.

2. Material and methods

2.1. Species checklists

Comprehensive species-occurrence records on freshwater fish (Actinopterygii), aquatic insects (with a focus on Ephemeroptera, Plecoptera, Trichoptera, Odonata and parts of Diptera, Coleoptera, Hemiptera and Megaloptera that are commonly used in aquatic biomonitoring) and molluscs (Bivalvia and Gastropoda) in China were downloaded from GBIF, published papers and available local faunistic compilations (e.g., on the freshwater fishes of China) (Chen, 1998; He et al., 2020; Kang et al., 2014; Xie, 2007; Zhang et al., 2016). Initially, a total of 47,236 Actinopterygii, 190,985 Insecta and 48,063 Mollusca records were retrieved. Then, scientific names of species were standardized by comparing them with Fishbase, MolluscaBase (https://fishbase.cn/search.php, http://www.molluscabase.org), Catalogue of Life China 2020 Annual Checklist, and Wikipedia (e.g., removing synonyms and checking the validity of names). The species records not in the natural range of the Chinese mainland were removed from the retrieved database. Finally, 16,418 freshwater fish (1,201 species), 16,288 aquatic insects (2,256 species) and 3,709 mollusc records (505 species) were kept for further analysis (Table S1).

We focused on ten major river catchments of China covering *ca.* 9.6 million km^2 territory (Fig. S1), for which we delimited the species checklists to a geographical selection by compiling only the species with China occurrences that include the following regions: Yangtze River basin (YanZ), Yellow River basin (YelR), Songhua River basin (SonH), Liao River basin (LiaR), Pearl River basin (PerR), Huai River basin (HuaR), Hai River basin (HaiR), Southeast River basin (SouE), Southwest River basin (SouW) and Continental River basin (ConR). The distribution map of the ten major river catchments was derived from the Resource and Environment Science and Data Center of the Chinese Academy of Sciences (RESDC, http://www.resdc.cn/). We only kept records with coordinate data (latitude and longitude) or county names, and records with country names were georeferenced using the Baidu Picking Coordinate System (http://api.map.baidu.com/lbsapi/getpoint/) (Fig. S2). Binary datasets (presence/absence) indicate the presence of a species in a catchment as '1' and the absence as '0', 1,167 freshwater fish, 772 aquatic insects and 294 molluscs were finally used to compare the data gaps in different catchments (Tables S2-4).

2.2. Reference libraries

All the available COI sequences matching the species' names in the checklist of the freshwater fish, aquatic insects and molluscs were mined separately in both the BOLD and NCBI GenBank databases in March 2021. Only COI records longer than 500 base pairs and with a clear taxonomic assignment (e.g., matching a species name) were retained. We kept the georeferenced data and uploaded the organization of the records for comparison of subsequent data sources (from Chinese institutions or other countries). To this dataset, we have also added new COI sequences from indigenous species, including records from our recent publication (Li et al., 2020). Finally, we collected a total of 20,578 and 17,365 COI sequences from 1,201 freshwater fish, 36,950 and 22,712 sequences from 2,256 aquatic insects, 11,807 and 9,407 sequences from 505 molluscs species in either the BOLD or NCBI GenBank databases, respectively (Table 1).

2.3. Gap analysis of databases

To conduct a detailed gap analysis of the barcoded species in Chinese rivers, we compared the species checklists with all the published COI records in both the BOLD and NCBI GenBank databases. Referring to the threshold of the minimum number of DNA barcodes (i.e., 1 and 5) set by published study (Weigand et al., 2019), the coverages of barcoded species in each checklist from each database were calculated, respectively. To analyze the data gaps in the reference libraries, we considered the species records from China, including geographical samples and uploaded from Chinese institutions ("Indigenous Barcodes"), and compared the differences between species from geographical records outside China ("Non-indigenous Barcodes"). To further verify the differences of the barcode sequences in Indigenous Barcodes and Non-indigenous Barcodes, we chose species with more than 5 sequences to calculate pairwise genetic distances for the same species but different sources in each group, separately. Sequences were aligned using the ClustalW algorithm with default parameters in MEGA 10.2 software, and then the pairwise genetic distances were calculated using Kimura's 2–parameter model with 1000 bootstrap replicates (Kumar et al., 2018). The frequency density of these distances for each group was plotted to visualize the barcode gaps between indigenous and non-indigenous barcodes. To reveal the record differences of barcodes in Chinses rivers, catchment-based summaries were generated to provide a clear picture of the coverage of barcoded species for each taxonomic group in both the BOLD and NCBI GenBank databases.

3. Results

3.1. Gap-analysis of the reference barcodes

Our data revealed a barcode coverage of only *ca.* 30–60% of all species in the three considered aquatic animal groups in either the BOLD or NCBI GenBank databases (Fig. 1, Table 1). Specifically, freshwater fish had the highest coverage, with 60.5% (BOLD) and 57.2% (NCBI GenBank) of the species having one or more barcodes available, respectively. However, the coverage drops to less than 40% if the barcode threshold is set to include minimally five reference sequences (36.9% for BOLD; 34.4% for NCBI GenBank). Molluscs come second, with a coverage of 56.4% and 51.1%, respectively, for one or more barcodes, and 31.7% and 28.7% of species, respectively, having at least five barcodes in the BOLD or NCBI GenBank databases. The lowest barcode coverage was found in aquatic insects, namely 30.2% (BOLD) and 26.8% (NCBI GenBank) for at least one barcode available, 15.7% (BOLD) and 12.1% (NCBI GenBank) for five or more barcodes, respectively.

A total of 1201 freshwater fish were compiled, including Cypriniformes (68.2% of the species checklists), Siluriformes (12.0%) and Perciformes (10.3%) (Table S1). Clear differences in the coverage of DNA barcodes among taxonomic groups were observed (Fig. 1, Table 1). For example, considering all species with at least one barcode, the five orders Acipenseriformes, Tetraodontiformes, Salmoniformes, Beloniformes and Clupeiformes are best covered, with a mean of 80% of barcoded species in both reference libraries, followed by Perciformes (77.0%), Osmeriformes (63.9%) and Siluriformes (60.1%), Cypriniformes had the lowest coverage, 52.9% for the 819 listed species. If the stricter criterion of more than five barcodes per species is considered, only in Acipenseriformes >80% of species are still represented by at least five barcodes, the coverage of the two largest orders, Cypriniformes and Siluriformes, falls below 30%.

Aquatic insect checklists on the taxonomic groups assessed here contain 2,256 species that were partitioned into eight prominent groups

Table 1
Overall barcode coverage for three aquatic animal groups at the order (freshwater fish and aquatic insects) or class (molluscs) level in either BOLD or NCBI GenBank databases, respectively. Taxonomic groups with less than eight known species from China are not listed. Total number of records (Total Records) as well as the proportion of these records from China ("From China %") are given. "≥1 barcode %" refers to the percentage of records with minimally one record; "≥5 barcode %" refers to the percentage of records with minimally five records.

Taxonomic groups	NO. Species	BOLD				NCBI GenBank			
		Total Records	From China %	≥ 1 barcode %	≥ 5 barcodes %	Total Records	From China %	≥ 1 barcode %	≥ 5 barcodes %
Freshwater Fish	**1201**	**20,578**	**9.7**	**60.5**	**36.9**	**17,365**	**32.73**	**57.2**	**34.4**
Acipenseriformes	8	294	31.6	100.0	87.5	290	49.66	100.0	75.0
Beloniformes	13	340	3.5	76.9	61.5	325	54.46	84.6	61.5
Clupeiformes	10	345	13.6	70.0	60.0	291	58.08	90.0	60.0
Cypriniformes	819	9372	13.8	54.5	29.9	8323	40.61	51.3	27.1
Osmeriformes	18	282	14.9	61.1	44.4	211	25.59	66.7	38.9
Perciformes	124	4962	3.6	80.7	66.9	3713	17.88	73.4	64.5
Salmoniformes	19	1304	4.1	89.5	79.0	942	6.05	79.0	68.4
Siluriformes	144	1676	11.0	59.7	28.5	1659	39.54	60.4	30.6
Tetraodontiformes	11	168	35.7	90.9	54.6	127	36.22	100.0	54.6
Aquatic Insects	**2256**	**36,950**	**7.9**	**30.2**	**15.7**	**22,712**	**12.35**	**26.8**	**12.1**
Coleoptera	35	268	0.4	17.1	8.6	186	0.54	17.1	5.7
Hemiptera	11	4	0.0	9.1	0.0	0	0.00	0.0	0.0
Megaloptera	12	132	1.5	83.3	33.3	120	95.83	75.0	25.0
Diptera	1206	31,019	6.5	27.3	17.0	18,198	13.51	23.6	12.9
Ephemeroptera	61	857	31.6	37.7	32.8	551	2.00	21.3	14.8
Odonata	453	2813	2.0	37.0	14.5	2038	7.21	35.5	11.8
Plecoptera	181	262	19.5	10.5	4.4	125	0.00	4.4	2.8
Trichoptera	297	1595	32.4	42.4	16.5	1494	4.89	41.4	15.5
Molluscs	**505**	**11,807**	**25.1**	**56.4**	**31.7**	**9407**	**43.26**	**51.1**	**28.7**
Bivalvia	193	4370	18.2	54.4	36.3	3290	57.87	44.6	30.1
Gastropoda	312	7437	29.1	57.7	28.9	6117	35.39	55.1	27.9

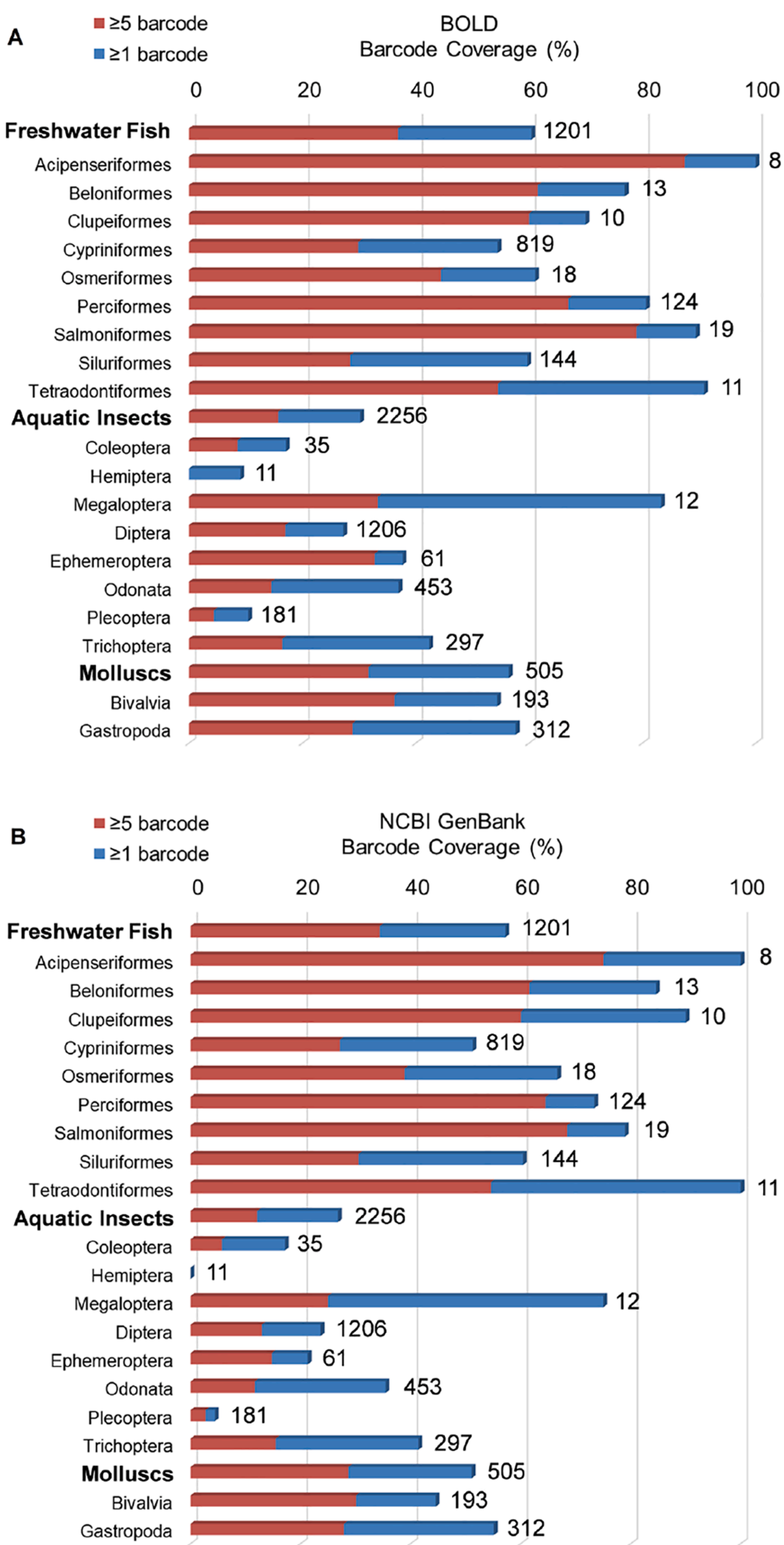

Fig. 1. Cumulative barcode coverage for three aquatic animal groups occurring in Chinese rivers at the order (freshwater fish and aquatic insects) or class (molluscs) level in either the BOLD (A) or NCBI GenBank databases (B), respectively. Barcode coverage was assessed at the level of at least one or five reference sequences per taxon, respectively. Taxonomic groups with less than eight known species from China are not listed. Numbers above bars indicate the total number of species reported in checklists.

(Table S1). Our data found considerable gaps (up to 80%) in barcode coverage among these groups (Fig. 1, Table 1). Specifically, Megaloptera displayed the highest proportion of species with at least one barcode (83.3% for BOLD; 75.0% for NCBI GenBank). The largest order (1206 species, Diptera) had just 27.3% (BOLD) and 23.6% (NCBI GenBank) coverage of one or more barcodes. Odonata had at least one barcode for 37.0% (BOLD) and 35.5% (NCBI GenBank) of the 453 species, respectively. The coverages of the remaining taxa were between 9 and 42%

(BOLD) and 0–41% (NCBI GenBank), including Ephemeroptera, Plecoptera and Trichoptera (EPT insects). The lowest proportion of species with barcodes was Hemiptera, with 0% in BOLD and 9.1% in NCBI GenBank. A similar pattern also appeared in the five barcode thresholds, but the barcode coverage of each taxonomic group decreased by *ca.* 60%.

Taxonomic checklists of molluscs included 505 species (Table S1), a gap analysis showed moderate levels of barcode coverage (≥1 barcode) in Bivalvia (193 species, 54.5% for BOLD and 44.6% for NCBI GenBank) and Gastropoda (312 species, 57.7% for BOLD and 55.1% for NCBI GenBank) (Fig. 1, Table 1). Similarly, the barcode coverage of species declined by *ca.* 30% (Bivalvia) and 50% (Gastropoda) in either the BOLD or NCBI GenBank databases when a minimum coverage of five barcodes was used.

3.2. Comparation of barcodes in different data sources

The overall proportion of barcoded species for the three aquatic animal groups deposited was slightly higher in BOLD (mean 54.9%) than in NCBI GenBank (mean 52.7%). For the different taxonomic groups, this was for freshwater fish 60.5% in BOLD and 57.2% in NCBI GenBank, for aquatic insects 30.2% vs. 26.8%, and for molluscs 56.4%

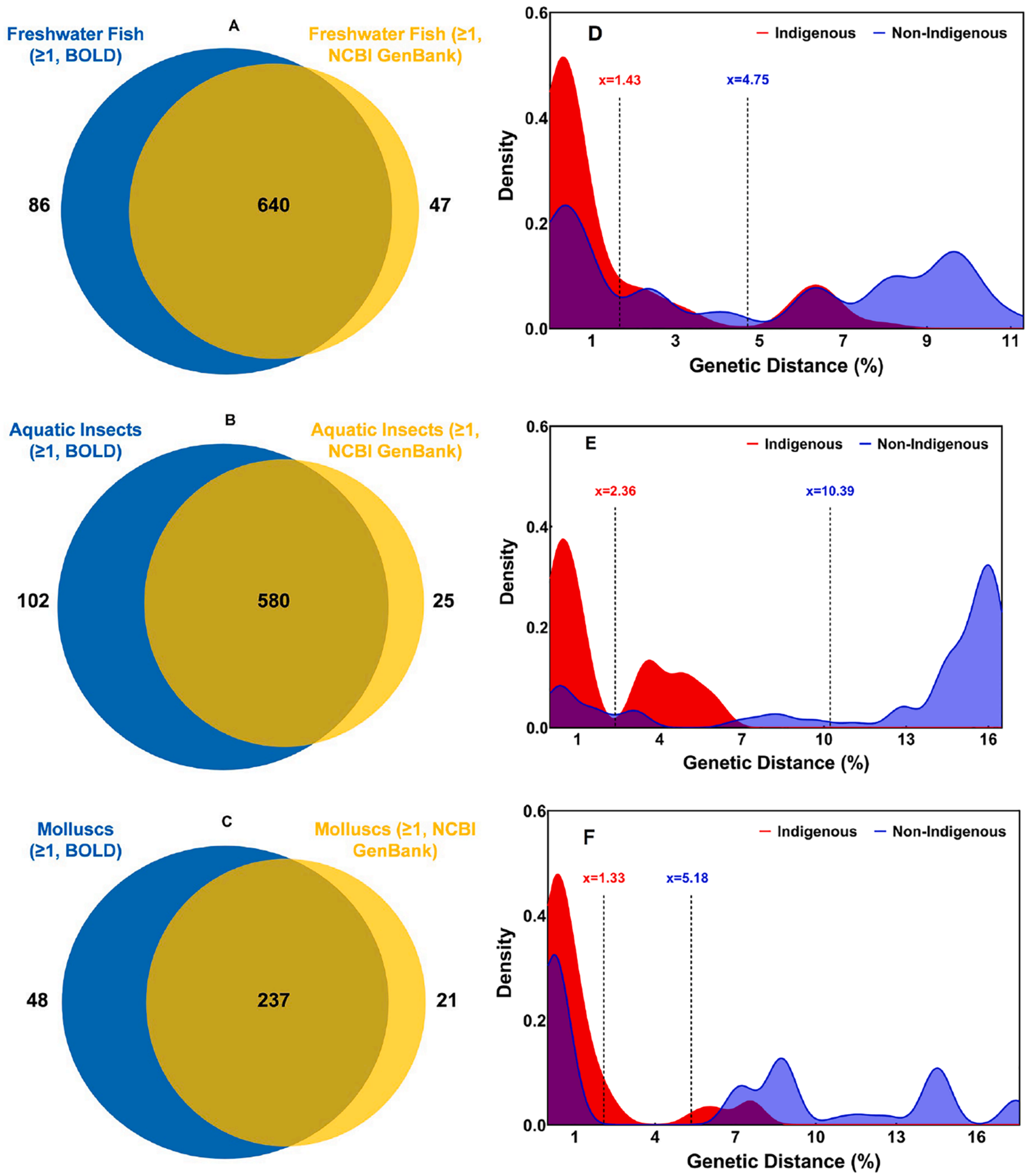

Fig. 2. Comparisons of common and unique barcodes for three aquatic animal groups occurring in Chinese rivers in both the BOLD and NCBI GenBank databases (≥1 barcode, A-C), and pairwise differences in the genetic distances in indigenous barcodes and non-indigenous barcodes for each group calculated using Kimura's 2-parameter model (D-E). Vertical dashed lines represent the average value of pairwise genetic distances.

vs. 51.1%, respectively (see also Table 1). Among the barcoded species in these three groups, irrespective of 1 or 5 barcodes, more than 70% of the reference sequences were shared in both databases (Fig. 2A-C, Fig. S3). Our data revealed that only 14.4% and 28.8% of the barcoded species in both the BOLD and NCBI GenBank databases were derived from organisms sampled in China, respectively. Among them, molluscs had the highest proportion of sequences belonging to the indigenous barcodes (mean 24.2% for BOLD; 45.5% for NCBI GenBank), followed

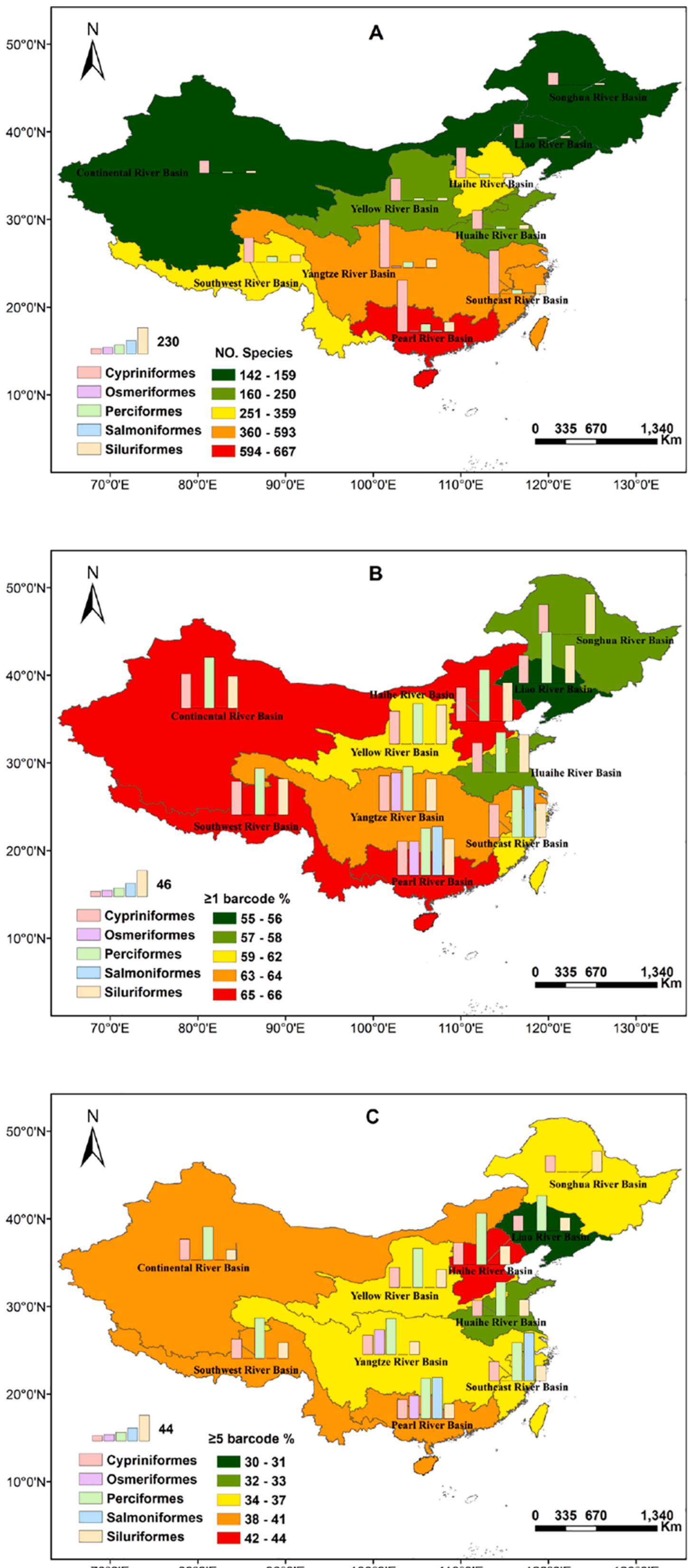

Fig. 3. Spatial variation in species number and barcode coverage of freshwater fish across ten major river catchments in China. Number of species per catchment (A), and barcode coverage per catchment for different thresholds (≥1 barcode %, percentage of minimum one record; ≥5 barcode %, percentage of minimum five records) in the BOLD database (B and C). The shaded plots reflect the overall coverage of freshwater fish, and the bar plots represent the coverage at the order level. Taxonomic groups with fewer than eight known species from China are not shown.

by freshwater fish (mean 14.2% vs. 36.1%) and aquatic insects (mean 11.3% vs. 15.2%).

Analysis of pairwise genetic distances for the same species suggested that indigenous barcodes had lower pairwise genetic distances than non-indigenous barcodes (3–5 times) as a whole (Fig. 2D-E). Specifically, the genetic distances of the barcode sequences for aquatic insects were the highest, the indigenous barcodes (mean 2.0%) were fivefold lower than non-indigenous barcodes (mean 10.4%); molluscs were the second-

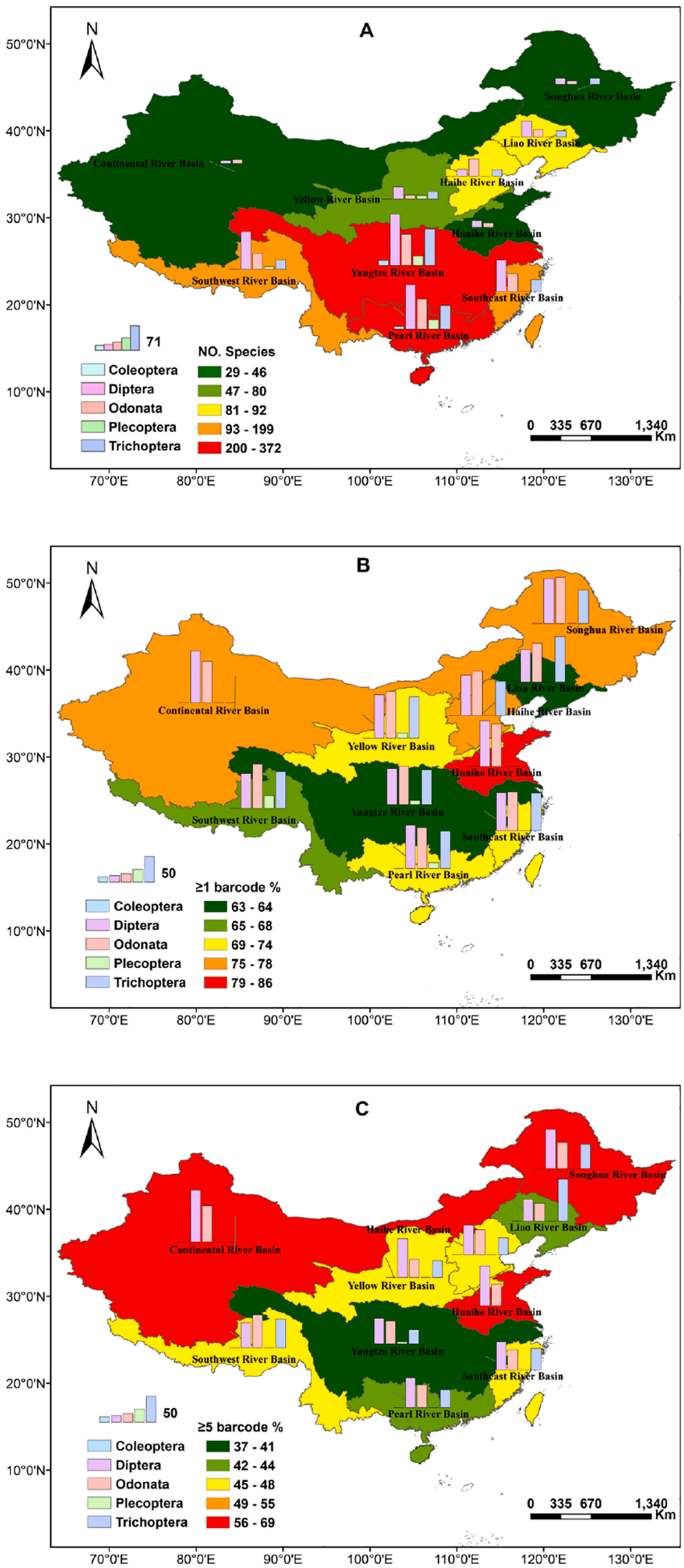

Fig. 4. Spatial variation in species number and barcode coverage of aquatic insects across ten major river catchments in China. Number of species per catchment (A), and barcode coverage per catchment for different thresholds (≥1 barcode %, percentage of minimum one record; ≥5 barcode %, percentage of minimum five records) in the BOLD database (B and C). The shaded plots reflect the overall coverage of aquatic insects, and the bar plots represent the coverage at the order level. Taxonomic groups with fewer than eight known species from China are not shown.

highest *ca.* 4 times (1.3% vs. 5.2%), and the lowest was freshwater fish *ca.* 3 times (1.4% vs. 4.8%).

3.3. Record differences of barcodes in Chinese rivers

Results revealed that *ca.* 60% of aquatic species looked at per catchment in China were covered with at least one barcode in both BOLD and NCBI GenBank. In terms of these three groups as a whole, the barcode coverage had a slight variation across different catchments, ranging from 54.3% (LiaR) to 68.2% (HuaR). The taxa Salmoniformes and Perciformes in freshwater fish, Odonata and Diptera in aquatic insects, and Bivalvia in molluscs were best barcoded in most catchments (mean >70%), and except for Coleoptera and Plecoptera in aquatic insects (<10%), the coverage of most groups in three aquatic animals per catchment was more than 50% (Figs. 3–5, Fig. S4-6, Table S5).

The recorded species of freshwater fish across ten major river catchments varied greatly, ranging from 142 (SonH) to 667 (PerR), and Cypriniformes, Siluriformes and Perciformes were the most abundant and representative orders (Fig. 3A, Table S2). We found that the barcode coverages in both BOLD and NCBI GenBank were relatively evenly distributed, LiaR had the lowest proportion of species with COI barcodes (54.7% ≥1 barcode, 29.6% ≥5 barcodes for BOLD; and 47.2% ≥1 barcode, 28.3% ≥5 barcodes for NCBI GenBank), HaiR reached the maximum coverage of a mean 64.4% (≥1 barcode) and 42.9% (≥5 barcodes) (Fig. 3B-C, Fig. S4). In addition, nearly 60% of freshwater fish per catchment were barcoded (61.9% ≥1 barcode for BOLD; 57.8% ≥1 barcode for NCBI GenBank), ranging from orders with 62.1% (Cypriniformes) and 63.0% (Siluriformes) barcoded species to highly covered orders (Perciformes, 73.2% and Salmoniformes, 82.5%) in both reference libraries. The overall coverage dropped by nearly 30% when the threshold was set to 5 barcodes (Table S5).

Assessed taxa of aquatic insects per catchment ranged from 29 (ConR) to 372 (YanZ), including mainly Diptera, Odonata and Trichoptera (Fig. 4A, Table S1). The barcode coverage in both BOLD and NCBI GenBank was relatively unevenly distributed, ranging between 62.1% (YanZ) and 86.5% (HuaR) (Fig. 4B-C, Fig. S5). We found that more than 70% of assessed aquatic insects per catchment were barcoded (≥1 barcode), ranging from orders with low coverage, such as Coleoptera (0%) and Plecoptera (3.6%), to highly covered orders, such as Trichoptera (58.1%), Diptera (74.6%) or Odonata (82.1%). Similarly, the overall coverages dropped by nearly 40% when the threshold was set to 5 barcodes (Table S5).

The spatial distribution of retrieved molluscs species varied strongly, with a minimum of 30 (ConR) to 118 (SouE) (Fig. 5A, Table S4). The barcode coverage of molluscs ranged from 36.5% (SouH) and 48.6% (LiaR) to a high coverage (SouE, 63.6%; YelR, 64.45%; YanZ, 66.7%) (Fig. 5B-C, Fig. S6). As in the other two groups, the barcode coverage dropped strongly (mean 30.8%) when a stricter criterion of ≥5 barcodes was applied. More specifically, the barcode coverage of Bivalvia (59.9% ≥1 barcode, 45.3% ≥5 barcodes) was higher than that of Gastropoda (52.1% ≥1 barcode, 33.6% ≥5 barcodes). The highest coverage of Bivalvia was found in YelR (mean 84.4%) in either the BOLD or NCBI GenBank databases, followed by YanZ (73.6%) and ConR (72.7%), and SonH had the lowest coverage (37.5%). Furthermore, the highest coverage of Gastropoda occurred in HuaR (mean 64.8%), followed by SouE (64.6%) and PerR (60%), and the lowest coverage occurred (36.2%, Table S5).

4. Discussion

Our barcode gap analyses of the BOLD and NCBI GenBank databases showed that *ca.* 40–70% of freshwater fish, aquatic insects and molluscs species reported from China are still missing the barcode coverage of COI reference sequences (e.g., organisms without barcodes), directly limiting the use of DNA-based taxon identification for assessing and detecting aquatic biodiversity. While COI may not be the ideal barcode region for all of these taxa (and other regions, such as 12S becoming more popular for fish) (Zhang et al., 2020), it is still one of the most commonly used, and most commonly covered region. Thus, all results can be interpreted as a "best case" scenario with respect to coverage, and even lower coverage may be expected for other barcode regions. Overall, the results are consistent with other studies that revealed large gaps in DNA barcodes for specific taxa (Leite et al., 2020; Paz and Rinkevich, 2021; Trebitz et al., 2015; Weigand et al., 2019). For example, the aquatic insects considered in our analysis (2,256 species), one of the most species-rich taxa retrieved in this study, have a gap of 70% of the species not included in the reference libraries. A similar gap analysis of aquatic biota in Europe found that the gaps of COI barcodes for aquatic insects were lower, namely around 40%, yet still twice that of marine and freshwater fish (<20%) (Weigand et al., 2019). We still found that those highly enigmatic or species-poor taxonomic groups had relatively high coverage of up to 80%, just as Acipenseriformes and Tetraodontiformes, dominated by rare and endangered species (e.g., *Acipenser sinensis* and *Psephurus gladius*), had the highest coverage (>80%), despite their small species number (<20 species, Table 1). Given the non-invasive and highly sensitive advantages of DNA-based taxon identification (e.g., eDNA metabarcoding or metabarcoding of bulk samples) to assess species abundance and distribution (Deiner et al., 2017; Deiner et al., 2020), DNA-based assessments will be crucial for the monitoring and protection of highly endangered or potentially even extinct species, such as *Psephurus gladius* (Zhang et al., 2020).

The best-covered group is freshwater fish, with *ca.* 60% of all species having at least one DNA barcode in both the BOLD and NCBI GenBank databases. While many fish species can be reliably identified based on COI markers (Balasingham et al., 2018; Ivanova et al., 2007; Shen et al., 2019), recent research suggests that 12S is a more appropriate marker than COI for fish monitoring due to better metabarcoding-primers specific for fish (Collins et al., 2019; Zhang et al., 2020). However, it is still fair to perform a gap analysis for COI reference sequences for fish, as COI is still better covered in the databases (thus giving a "best possible scenario") and for some species groups also better resolving species (Porter and Hajibabaei, 2018; Shen et al., 2019). Even for such relatively well-represented groups, the coverage drops by less than 40% if the barcode threshold is set to at least five reference sequences available per species (Table 1). This indicates that the three freshwater animal groups looked at (freshwater fish, aquatic insects and molluscs) are far from representatively covered in China. Especially Cyprinidae and Diptera, the largest and most widespread families of freshwater fishes and orders of aquatic insects in this study, had fairly low barcode coverage in both databases (mean 52.9% for 819 listed Cyprinidae species, mean 25.5% for 1206 listed Diptera species, respectively). These data give rise to another critical consideration, namely that the baseline species lists are likely also not complete (Bogan, 2008; Morse, 2017; Strong et al., 2008). Consequently, the observed gaps may be the tip of an iceberg: If the reference species lists against which the COI barcodes are mapped are incomplete, then the coverage of the true diversity is even lower. Resolving this will also require complementing the species inventories and checklists of species, requiring major taxonomic and fieldwork, and may uncover a large undiscovered diversity. Although our data are derived from the GBIF and Catalogue of Life China 2020 Annual Checklist, which is among the most complete databases at present, it is estimated that more than 80% of the species in most taxonomic groups (e.g., insects) on our planet are yet unknown (Eisenhauer et al., 2019; Scheffers et al., 2012). Additionally, for freshwater fish, where relatively good coverage can be assumed, the 1201 species reported from China may still contain dozens if not hundreds of species, as has been shown from other areas where radiations have only recently been uncovered (Chen, 1998; He et al., 2020; Kang et al., 2014; Xie, 2007; Zhang et al., 2016).

A remarkable finding for three aquatic groups recorded in China is that more indigenous barcodes are found in NCBI GenBank than BOLD, although the coverage of species is still very low (mean less than 30%).

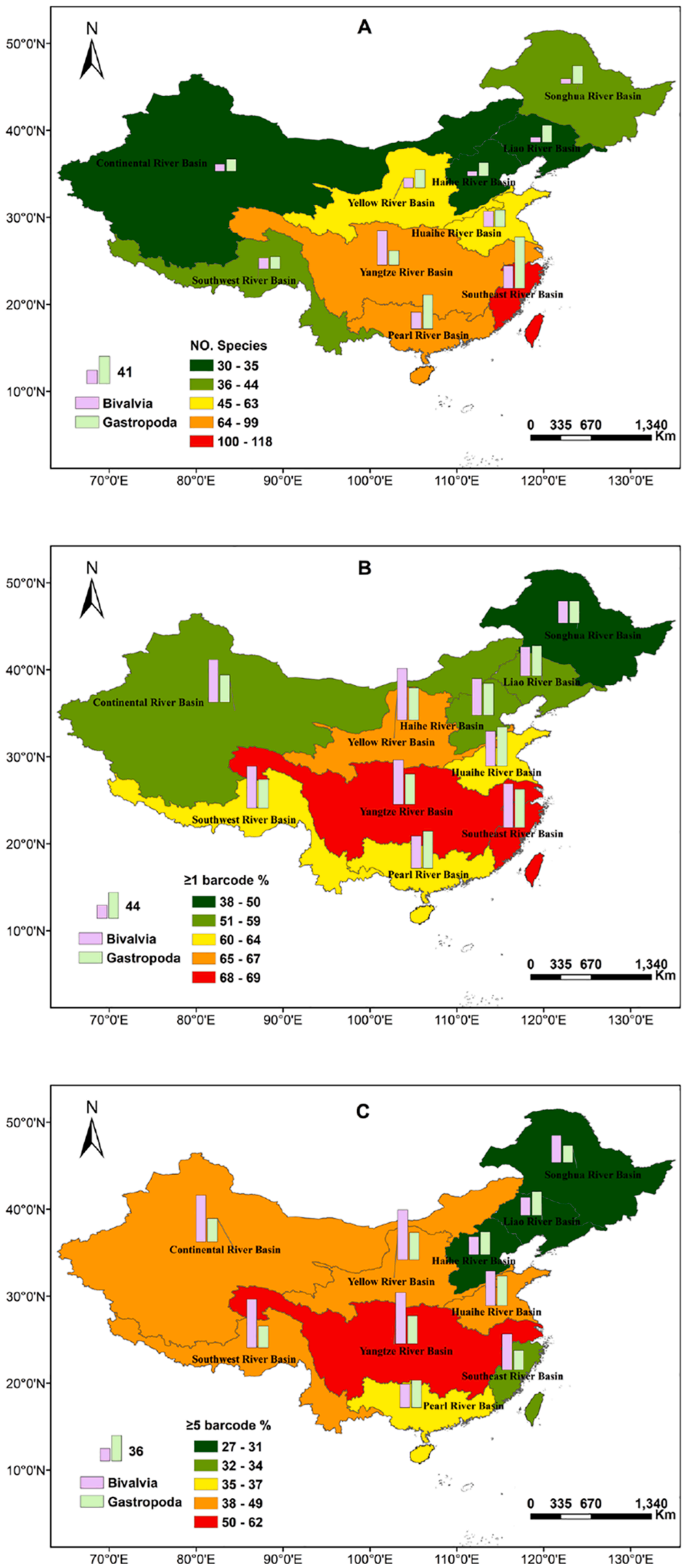

Fig. 5. Spatial variation in species number and barcode coverage of molluscs across ten major river catchments in China. Number of species per catchment (A), and barcode coverage per catchment for different thresholds (≥1 barcode %, percentage of minimum one record; ≥5 barcode %, percentage of minimum five records) in the BOLD database (B and C). The shade plots reflect the overall coverage of molluscs, and the bar plots represent the coverage at the class level. Taxonomic groups with fewer than eight known species from China are not shown.

We suggested that two factors can explain the results: on the one hand, scholars preferred to upload their sequences in the NCBI GenBank database in the pre-BOLD era (Chen et al., 2015; Shen et al., 2019; Zhang and Hanner, 2012), which can be regarded as an objective factor; On the other hand, a possible subjective bias was derived from the process of data collection, that is, compared with the barcode sources determined by longitude and latitude data in BOLD, NCBI GenBank mainly relies on the location of scholars and their institutions to determine the source of barcodes, which inevitably integrate the international cooperation samples participated by Chinese scholars into the indigenous barcodes (Song et al., 2018; Yang et al., 2016). As a direct consequence of only a few indigenous barcodes, we found that the pairwise genetic distance of non-indigenous barcodes is 3 to 5 times that of indigenous barcodes. This directly reduces the ability of DNA-based taxon identification to assess real biodiversity in ecosystems (Dinca et al., 2015; Jackson et al., 2014; Sweeney et al., 2011). Moreover, spatial scale coverage of barcodes (also called barcode maps) is not only important to explain the genetic diversity related to geographical distance (Bergsten et al., 2012; Manel et al., 2020; Millette et al., 2020), but can also at least provide scholars and managers with priority areas to fill the gaps in species barcodes.

Our analysis further emphasizes the urgency of filling the gaps in the barcode libraries used for biodiversity assessments based on metabarcoding. The taxonomic span of detection and accuracy of DNA-based taxon identification could be seriously limited by the incompleteness of reference sequence databases. Providing a greater number of reference sequences for diverse and distinct populations is undoubtedly a key step to enhance the credibility of DNA-based taxon identification (Hering et al., 2018; Meiklejohn et al., 2019; Pawlowski et al., 2018; Paz and Rinkevich, 2021; Weigand et al., 2019). This is also a critical point for the next biomonitoring of Chinese rivers, for which compilation and filing of the reference library are of high importance and where researchers and government administrators start to incorporate DNA-based methods into their daily biomonitoring routines for the future investigation of aquatic organisms (Qu et al., 2020; Shen et al., 2019). Similarly, DNA-based taxon identification will be applied to other river catchments in succession, because the new methods have been compiled into the 14th five-year plan for aquatic biodiversity monitoring and protection of China's central and local governments. All of these activities will profit and depend on an accurate understanding and interpretation of the DNA-based data, which themselves depend on an accurate and as complete as a possible assignment to taxonomic databases.

5. Conclusions

In this study we conducted a comprehensive gap analysis of COI barcodes for freshwater fish, aquatic insects and molluscs in Chinese rivers. We found gaps of 40–70% of species in these three groups not covered, meaning that China is far from having a representative and complete reference library for these taxonomic groups. Given that freshwater fish, aquatic insects and molluscs looked at are among the best researched and also most important taxa for biomonitoring, this is likely severely hindering effective implementation, but also shows that these gaps may be only the tip of an iceberg, with less-well studied groups likely having even larger gaps. Furthermore, nearly 90% of the COI sequences corresponding to aquatic species in China are from other countries. Although these barcodes can be substitutes for short-term strategies, there is an urgent need to supplement local reference libraries with barcodes from the biogeographic region studied itself for adequate long-term use. Because the pairwise genetic distances of indigenous barcodes are 3–5 times lower than that of non-indigenous ones, a mismatch (over- or underestimation of diversity) is likely. In addition, given that at least 70% of species have less than 5 barcodes, increasing the number of reference sequences for less barcoded species is the key goal of future biomonitoring projects. With the continuous development of DNA barcode reference libraries, DNA-based taxon identification will enable us to better understand intraspecific and interspecific diversity, as well as unknown organisms in Chinese rivers.

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.

Acknowledgements

This study was supported by the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (2019ZT08L213), Guangdong Provincial Key Laboratory Project (2019B121203011), and National Natural Science Foundation of China (52100216). F.A. was supported by the University of Zurich Research Priority Program "URPP Global Change and Biodiversity".

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2022.108732.

References

Altermatt, F., Little, C.J., Mächler, E., Wang, S., Zhang, X., Blackman, R.C., 2020. Uncovering the complete biodiversity structure in spatial networks: the example of riverine systems. Oikos 129, 607–618.
Andujar, C., Arribas, P., Yu, D.W., Vogler, A.P., Emerson, B.C., 2018. Why the COI barcode should be the community DNA metabarcode for the metazoa. Mol. Ecol. 27, 3968–3975.
Baird, D.J., Hajibabaei, M., 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. Mol. Ecol. 21, 2039–2044.
Balasingham, K.D., Walter, R.P., Mandrak, N.E., Heath, D.D., 2018. Environmental DNA detection of rare and invasive fish species in two Great Lakes tributaries. Mol. Ecol. 27, 112–127.
Beng, K.C., Corlett, R.T., 2020. Applications of environmental DNA (eDNA) in ecology and conservation: opportunities, challenges and prospects. Biodivers. Conserv. 29, 2089–2121.
Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2013. GenBank. Nucleic Acids Res. 41, D36–42.
Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M.T., Balke, M., Hendrich, L., Geijer, J., Herrmann, J., Foster, G.N., Ribera, I., Nilsson, A.N., Barraclough, T.G., Vogler, A.P., 2012. The effect of geographical scale of sampling on DNA barcoding. Syst. Biol. 61, 851–869.
Bogan, A.E., 2008. Global diversity of freshwater mussels (Mollusca, Bivalvia) in freshwater. Hydrobiologia 595, 139–147.
Chen, W., Ma, X., Shen, Y., Mao, Y., He, S., 2015. The fish diversity in the upper reaches of the Salween River, Nujiang River, revealed by DNA barcoding. Sci. Rep. 5, 17437.
Chen, Y., 1998. Fauna Sinica: Osteichthyes, Cypriniformes II. Science Press.
Collins, R.A., Bakker, J., Wangensteen, O.S., Soto, A.Z., Corrigan, L., Sims, D.W., Genner, M.J., Mariani, S., 2019. Non-specific amplification compromises environmental DNA metabarcoding with COI. Methods Ecol. Evol. 10, 1985–2001.
Cristescu, M.E., Hebert, P.D.N., 2018. Uses and Misuses of Environmental DNA in Biodiversity Science and Conservation. Annu. Rev. Ecol. Evol. Syst. 49, 209–230.
Deiner, K., Bik, H.M., Machler, E., Seymour, M., Lacoursiere-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D.M., de Vere, N., Pfrender, M.E., Bernatchez, L., 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Mol. Ecol. 26, 5872–5895.
Deiner, K., Yamanaka, H., Bernatchez, L., 2020. The future of biodiversity monitoring and conservation utilizing environmental DNA. Environ. DNA 3, 3–7.
Dinca, V., Montagud, S., Talavera, G., Hernandez-Roldan, J., Munguira, M.L., Garcia-Barros, E., Hebert, P.D., Vila, R., 2015. DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. Sci. Rep. 5, 12395.
Eisenhauer, N., Bonn, A., A. Guerra, C., 2019. Recognizing the quiet extinction of invertebrates. Nat. Commun. 10, 50.
He, D., Sui, X., Sun, H., Tao, J., Ding, C., Chen, Y., Chen, Y., 2020. Diversity, pattern and ecological drivers of freshwater fish in China and adjacent areas. Rev. Fish Biol. Fish. 30, 387–404.
Hering, D., Borja, A., Jones, J.I., Pont, D., Boets, P., Bouchez, A., Bruce, K., Drakare, S., Hanfling, B., Kahlert, M., Leese, F., Meissner, K., Mergen, P., Reyjol, Y., Segurado, P., Vogler, A., Kelly, M., 2018. Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. Water Res. 138, 192–205.

Huang, G., Ping, X., Xu, W., Hu, Y., Chang, J., Swaisgood, R.R., Zhou, J., Zhan, X., Zhang, Z., Nie, Y., Cui, J., Bruford, M., Zhang, Z., Li, B., Zhang, L., Lv, Z., Wei, F., 2021. Wildlife conservation and management in China: achievements, challenges and perspectives. Nat. Sci. Rev.

Ivanova, N.V., Zemlak, T.S., Hanner, R.H., Hebert, P.D., 2007. Universal primer cocktails for fish DNA barcoding. Mol. Ecol. Notes 7, 544–548.

Jackson, J.K., Battle, J.M., White, B.P., Pilgrim, E.M., Stein, E.D., Miller, P.E., Sweeney, B.W., 2014. Cryptic biodiversity in streams: a comparison of macroinvertebrate communities based on morphological and DNA barcode identifications. Freshw. Sci. 33, 312–324.

Jetz, W., McGeoch, M.A., Guralnick, R., Ferrier, S., Beck, J., Costello, M.J., Fernandez, M., Geller, G.N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F.E., Pereira, H.M., Regan, E.C., Schmeller, D.S., Turak, E., 2019. Essential biodiversity variables for mapping and monitoring species populations. Nat. Ecol. Evol. 3, 539–551.

Kang, B., Deng, J., Wu, Y., Chen, L., Zhang, J., Qiu, H., Lu, Y., He, D., 2014. Mapping China's freshwater fishes: diversity and biogeography. Fish Fish. 15, 209–230.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549.

Leite, B.R., Vieira, P.E., Teixeira, M.A.L., Lobo-Arteaga, J., Hollatz, C., Borges, L.M.S., Duarte, S., Troncoso, J.S., Costa, F.O., 2020. Gap-analysis and annotated reference library for supporting macroinvertebrate metabarcoding in Atlantic Iberia. Reg. Stud. Marine Sci. 36, 101307.

Leray, M., Knowlton, N., Ho, S.L., Nguyen, B.N., Machida, R.J., 2019. GenBank is a reliable resource for 21st century biodiversity research. Proc. Natl. Acad. Sci. U. S. A. 116, 22651–22656.

Li, F., Altermatt, F., Yang, J., An, S., Li, A., Zhang, X., 2020. Human activities' fingerprint on multitrophic biodiversity and ecosystem functions across a major river catchment in China. Glob. Chang. Biol. 26, 6867–6879.

Liu, J., Ouyang, Z., Pimm, S.L., Raven, P.H., Wang, X., Miao, H., Han, N., 2003. Ecology. Protecting China's biodiversity. Science 300, 1240–1241.

Lu, Y., Yang, Y., Sun, B., Yuan, J., Yu, M., Stenseth, N.C., Bullock, J.M., Obersteiner, M., 2020. Spatial variation in biodiversity loss across China under multiple environmental stressors. Sci. Adv. 6, eabd0952.

Lynch, A.J., Elliott, V., Phang, S.C., Claussen, J.E., Harrison, I., Murchie, K.J., Steel, E.A., Stokes, G.L., 2020. Inland fish and fisheries integral to achieving the Sustainable Development Goals. Nat. Sustainability 3, 579–587.

Manel, S., Guerin, P.E., Mouillot, D., Blanchet, S., Velez, L., Albouy, C., Pellissier, L., 2020. Global determinants of freshwater and marine fish genetic diversity. Nat. Commun. 11, 692.

McGee, K.M., Robinson, C.V., Hajibabaei, M., 2019. Gaps in DNA-based biomonitoring across the globe. Front. Ecol. Evol. 7.

Meiklejohn, K.A., Damaso, N., Robertson, J.M., 2019. Assessment of BOLD and GenBank - Their accuracy and reliability for the identification of biological materials. PLoS ONE 14, e0217084.

Mi, X., Feng, G., Hu, Y., Zhang, J., Chen, L., Corlett, R.T., Hughes, A.C., Pimm, S., Schmid, B., Shi, S., Svenning, J.-C., Ma, K., 2021. The global significance of biodiversity science in China: an overview. National Science. Review.

Millette, K.L., Fugere, V., Debyser, C., Greiner, A., Chain, F.J.J., Gonzalez, A., 2020. No consistent effects of humans on animal genetic diversity worldwide. Ecol. Lett. 23, 55–67.

Morse, J.C., 2017. Biodiversity of aquatic insects. Insect Biodivers. 205–227.

Pawlowski, J., Apotheloz-Perret-Gentil, L., Altermatt, F., 2020. Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. Mol. Ecol. 29, 4258–4264.

Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apotheloz-Perret-Gentil, L., Beja, P., Boggero, A., Borja, A., Bouchez, A., Cordier, T., Domaizon, I., Feio, M.J., Filipe, A.F., Fornaroli, R., Graf, W., Herder, J., van der Hoorn, B., Iwan Jones, J., Sagova-Mareckova, M., Moritz, C., Barquin, J., Piggott, J.J., Pinna, M., Rimet, F., Rinkevich, B., Sousa-Santos, C., Specchia, V., Trobajo, R., Vasselon, V., Vitecek, S., Zimmerman, J., Weigand, A., Leese, F., Kahlert, M., 2018. The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. Sci. Total Environ. 637–638, 1295–1310.

Paz, G., Rinkevich, B., 2021. Gap analysis of DNA barcoding in ERMS reference libraries for ascidians and cnidarians. Environ. Sci. Eur. 33, 4.

Porter, T.M., Hajibabaei, M., 2018. Over 2.5 million COI sequences in GenBank and growing. PLoS ONE 13, e0200177.

Qu, C., Stewart, K.A., Clemente-Carvalho, R., Zheng, J., Wang, Y., Gong, C., Ma, L., Zhao, J., Lougheed, S.C., 2020. Comparing fish prey diversity for a critically endangered aquatic mammal in a reserve and the wild using eDNA metabarcoding. Sci. Rep. 10, 16715.

Ratnasingham, S., Hebert, P.D., 2007. BOLD: The Barcode of Life Data System. Mol. Ecol. Notes 7, 355–364.

Sagova-Mareckova, M., Boenigk, J., Bouchez, A., Cermakova, K., Chonova, T., Cordier, T., Eisendle, U., Elersek, T., Fazi, S., Fleituch, T., Fruhe, L., Gajdosova, M., Graupner, N., Haegerbaeumer, A., Kelly, A.M., Kopecky, J., Leese, F., Noges, P., Orlic, S., Panksep, K., Pawlowski, J., Petrusek, A., Piggott, J.J., Rusch, J.C., Salis, R., Schenk, J., Simek, K., Stovicek, A., Strand, D.A., Vasquez, M.I., Vralstad, T., Zlatkovic, S., Zupancic, M., Stoeck, T., 2021. Expanding ecological assessment by integrating microorganisms into routine freshwater biomonitoring. Water Res. 191, 116767.

Scheffers, B.R., Joppa, L.N., Pimm, S.L., Laurance, W.F., 2012. What we know and don't know about Earth's missing biodiversity. Trends Ecol. Evol. 27, 501–510.

Shen, Y., Hubert, N., Huang, Y., Wang, X., Gan, X., Peng, Z., He, S., 2019. DNA barcoding the ichthyofauna of the Yangtze River: Insights from the molecular inventory of a mega-diverse temperate fauna. Mol. Ecol. Resour. 19, 1278–1291.

Song, C., Lin, X.-L., Wang, Q., Wang, X.-H., 2018. DNA barcodes successfully delimit morphospecies in a superdiverse insect genus. Zoolog. Scr. 47, 311–324.

Strong, E.E., Gargominy, O., Ponder, W.F., Bouchet, P., 2008. Global diversity of gastropods (Gastropoda; Mollusca) in freshwater. Hydrobiologia 595, 149–166.

Su, G., Logez, M., Xu, J., Tao, S., Villéger, S., Brosse, S., 2021. Human impacts on global freshwater fish biodiversity. Science 371, 835–838.

Sweeney, B.W., Battle, J.M., Jackson, J.K., Dapkey, T., 2011. Can DNA barcodes of stream macroinvertebrates improve descriptions of community structure and water quality? J. North Am. Benthol. Soc. 30, 195–216.

Trebitz, A.S., Hoffman, J.C., Grant, G.W., Billehus, T.M., Pilgrim, E.M., 2015. Potential for DNA-based identification of Great Lakes fauna: match and mismatch between taxa inventories and DNA barcode libraries. Sci. Rep. 5, 12162.

Vorosmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., Davies, P.M., 2010. Global threats to human water security and river biodiversity. Nature 468, 334-334.

Weigand, H., Beermann, A.J., Ciampor, F., Costa, F.O., Csabai, Z., Duarte, S., Geiger, M.F., Grabowski, M., Rimet, F., Rulik, B., Strand, M., Szucsich, N., Weigand, A.M., Willassen, E., Wyler, S.A., Bouchez, A., Borja, A., Ciamporova-Zatovicova, Z., Ferreira, S., Dijkstra, K.B., Eisendle, U., Freyhof, J., Gadawski, P., Graf, W., Haegerbaeumer, A., van der Hoorn, B.B., Japoshvili, B., Keresztes, L., Keskin, E., Leese, F., Macher, J.N., Mamos, T., Paz, G., Pesic, V., Pfannkuchen, D.M., Pfannkuchen, M.A., Price, B.W., Rinkevich, B., Teixeira, M.A.L., Varbiro, G., Ekrem, T., 2019. DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. Sci. Total Environ. 678, 499–524.

Xie, Y., 2007. Freshwater Fishes in Northeast Region of China. Liaoning Science and Technology Press, Shenyang, China (in Chinese).

Yang, Z., Landry, J.F., Hebert, P.D., 2016. A DNA Barcode Library for North American Pyraustinae (Lepidoptera: Pyraloidea: Crambidae). PLoS ONE 11, e0161449.

Zhang, C., Zhao, Y., Xing, Y., Zhou, W., Tang, W., 2016. Species Diversity and Distribution of Inland Fishes in China. Science, Beijing, pp. 14–43.

Zhang, H., Jaric, I., Roberts, D.L., He, Y., Du, H., Wu, J., Wang, C., Wei, Q., 2020. Extinction of one of the world's largest freshwater fishes: Lessons for conserving the endangered Yangtze fauna. Sci. Total Environ. 710, 136242.

Zhang, H., Wu, J., Gorfine, H., Shan, X., Shen, L., Yang, H., Du, H., Li, J., Wang, C., Zhou, Q., Liu, Z., Kang, M., Wei, Q., 2020. Inland fisheries development versus aquatic biodiversity conservation in China and its global implications. Rev. Fish Biol. Fish. 30, 637–655.

Zhang, J., Hanner, R., 2012. Molecular approach to the identification of fish in the South China Sea. PLoS ONE 7, e30621.

Zhang, S., Zhao, J., Yao, M., Gilbert, M., 2020. A comprehensive and comparative evaluation of primers for metabarcoding eDNA from fish. Methods Ecol. Evol. 11, 1609–1625.