

RESEARCH

Open Access



Cryptic diversity, phenotypic congruence, and evolutionary history of the *Leptobotia citraurata* complex (Pisces: Botiidae) within subtropical eastern China

Dongming Guo^{1,2}, Xiong Gong¹, Wenjing Yi¹, Liang Cao¹ and E Zhang^{1*}

Abstract

Elucidating the emergence and maintenance of cryptic diversity is a major focus of evolutionary biology. Integrative taxonomy is widely considered as the best practice for delimiting cryptic species and exploring cryptic speciation. This approach is used here to study the *Leptobotia citraurata* complex, a group of small-sized loaches so far found in subtropical floodplains and hills of eastern China. A total 170 specimens were collected from 24 sampling sites, encompassing geographical variations and divergent habitats. Six putative species, out of which two are cryptic, were delineated by integrating molecular (two mtDNA and three nuDNA genes) and morphological analyses. These species constituted three ecotypes, exhibiting phenotypic disparities concordant with a habitat transition from high- to low-flow environments. Phenotypic similarities among them were shown to not align with their phylogenetic relationships but closely correlate with habitat utilization. Convergent evolution, driven by similar selective pressure associated with habitat-specific use, likely accounts for the cryptic diversity unveiled in the recently diverging species complex. The diversification of this species complex began in the late Pliocene, coinciding with tectonic activities in the subtropical region of eastern China. Subsequent rapid differentiation during the Pleistocene was possibly driven by regional climate fluctuations. This evolutionary trajectory highlights the crucial roles of geological, climate and ecological factors in shaping biodiversity in this region.

Keywords Loaches, Integrative taxonomy, Morphological cypsis, Cryptic speciation

Introduction

Cryptic species are biological entities which cannot readily be morphologically discernible [1]. The appreciation of cryptic diversity has remarkably increased with the use of molecular techniques in taxonomy [2]. While the incorporation of genetic data into taxonomic research poses a great challenge to traditional taxonomy relying on morphological evidence [3], this data used alone to delimit species may run a risk of misidentification of population-level structure as putative species, thus leading to over-splitting of species in taxonomic works [4, 5]. The integration of multiple lines of evidence, such as morphology, genetic, biogeography, behaviors, and ecology is

*Correspondence:

E Zhang
zhange@ihb.ac.cn

¹ Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei Province 430072, China

² Key Laboratory of Ecological Impacts of Hydraulic Projects and Restoration of Aquatic Ecosystem of Ministry of Water Resources, Institute of Hydroecology, Ministry of Water Resources and Chinese Academy of Sciences, Wuhan, Hubei Province 430079, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

recognized as a preferred practice for delimiting morphologically indiscernible or cryptic species [5–7].

Elucidating the processes responsible for cryptic speciation remains a key focus of evolutionary biology, as it provides unique opportunities to understand the origin of species and study phenotypic similarity, genetic differentiation and divergence times [8, 9]. Cryptic speciation may occur via diverse evolutionary trajectories [1], among which convergence is ubiquitously viewed as one of potential explanation [2, 10]. It involves the emergence of similar phenotypes in distantly related species in response to similar selective pressure [11, 12]. Comparisons between morphological characters and molecular phylogenetic analyses have demonstrated many cases of convergent evolution in fish groups, including sticklebacks in northern Europe, North America and Northeast Asia [13, 14], as well as cyprinids, cichlids, mormyrids, and spiny eels in tropical Africa [15–19]. Recently, phenotypic convergence in distantly related evolutionary lineages was unravelled in an alpine stone loach from western China [20]. Research on freshwater fishes in the subtropical region of eastern China have until now provided no empirical example of phenotypic convergent evolution. Convergence, however, is believed to be common, occurring at all biological levels from DNA sequences to communities [21, 22].

Subtropical China extends from the eastern Tibetan Plateau to the vast expanse of the Pacific Ocean, spanning a latitudinal range from the Qingling Mountains-Huai-He line (34°N) to the southern tropical region (20°N) [23]. This region experienced geological tectonic uplift during the late Miocene to the Pliocene, leading to diverse topography and various habitats that promoted biotic diversification [24, 25]. Besides, climatic oscillations related to glacial cycles during the Pleistocene played a crucial role in shaping biodiversity and leaving genetic imprints on rapid splitting lineages [26, 27]. The subtropical region of China is thus considered as a center of speciation and renowned for a high species diversity and endemism [24]. It may have served as a floristic ‘museum’ and an evolutionary ‘cradle’ for woody plants [28]. For animals, the influence of Pleistocene geological and climatic changes on contemporary genetic structuring within and among extant species has been well studied [29]. However, the role of ecological factors in species diversification remains under-explored. Habitat transition has been identified as one of the driving forces for species diversification in freshwater fishes [30], for instance, the European whitefish, threespine stickleback, Arctic char, and *Triplocheilichthys* characins in Amazon [31], as well as gudgeons in Japan [32]. Obviously, most of these cases, if not all, targeted at the temperate and tropical, but not subtropical region.

The botiid genus *Leptobotia* Bleeker, including 20 currently-identified species (excluding *S. zebra* so far referred to *Sinibotia*; [33]), has a concentrated distribution in subtropical China, with *L. pellegrini* into the Red River (in northern Vietnam) and *L. flavolineata* extended into the Hai-He (in northern China) [34–37]. It was resolved as a monophyletic group in phylogenetic analyses of the Cobitoidea [38] and the Botiidae [39]. Within *Leptobotia*, the *L. citraurata* complex (hereafter LCC) includes four species spanning subtropical floodplains and hills of eastern China, namely *L. brachycephala* Guo & Zhang, *L. citraurata* (Nichols), *L. hansuiensis* Fang & Hsu, and *L. micra* Bohlen & Šlechtová [37]. Recent taxonomic works of the LCC [36, 37] reinforce that re-examination of genetically distinct lineages can discern previously overlooked but subtle diagnostic characters, leading to the release of involved species from phenotypic cryptic species. It is thus anticipated that more cryptic species of the LCC will be detected with more careful examination on morphology of genetically distinct lineages and more sampling in its known range.

The LCC is a candidate group for exploring cryptic diversity and deciphering processes underlying morphological cryptic species. Phenotypic convergence may pass unnoticed among closely related species, and the key role of selection in generating phenotypic and species diversity may be underestimated for morphologically similar congeneric species of a single origin [16]. This is the case for the LCC that has been resolved as a monophyletic group within *Leptobotia* [37]. Since 2000, the number of *Leptobotia* species has increased by around one third (from 14 to 20), with half of higher species contributed by the LCC. The distribution of LCC, covering subtropical floodplains and hills of China (Fig. 1), falls within the East Asian monsoon climate region characterized by spatially heterogeneous and temporally dynamic landscapes. This complex topography is accountable for cryptic diversity in freshwater fishes of this region [40, 41].

Here, we explore the species diversity of the LCC utilizing molecular species delimitation analysis with a specific focus to recover genetically divergent lineages. Then, we examine morphological variations among these lineages using multivariate statistical analysis, aiming to assess whether they are phenotypically cryptic. These results incorporate geographic disparity and ecological divergence to delimit species of the LCC, and explore their phylogeographic pattern and underlying processes of species diversification. Our study can provide an empirical example for phenotypic convergent evolution in subtropical floodplains and hills of eastern China, and also highlights the role of habitat transition in generating cryptic diversity. This study will extend our knowledge

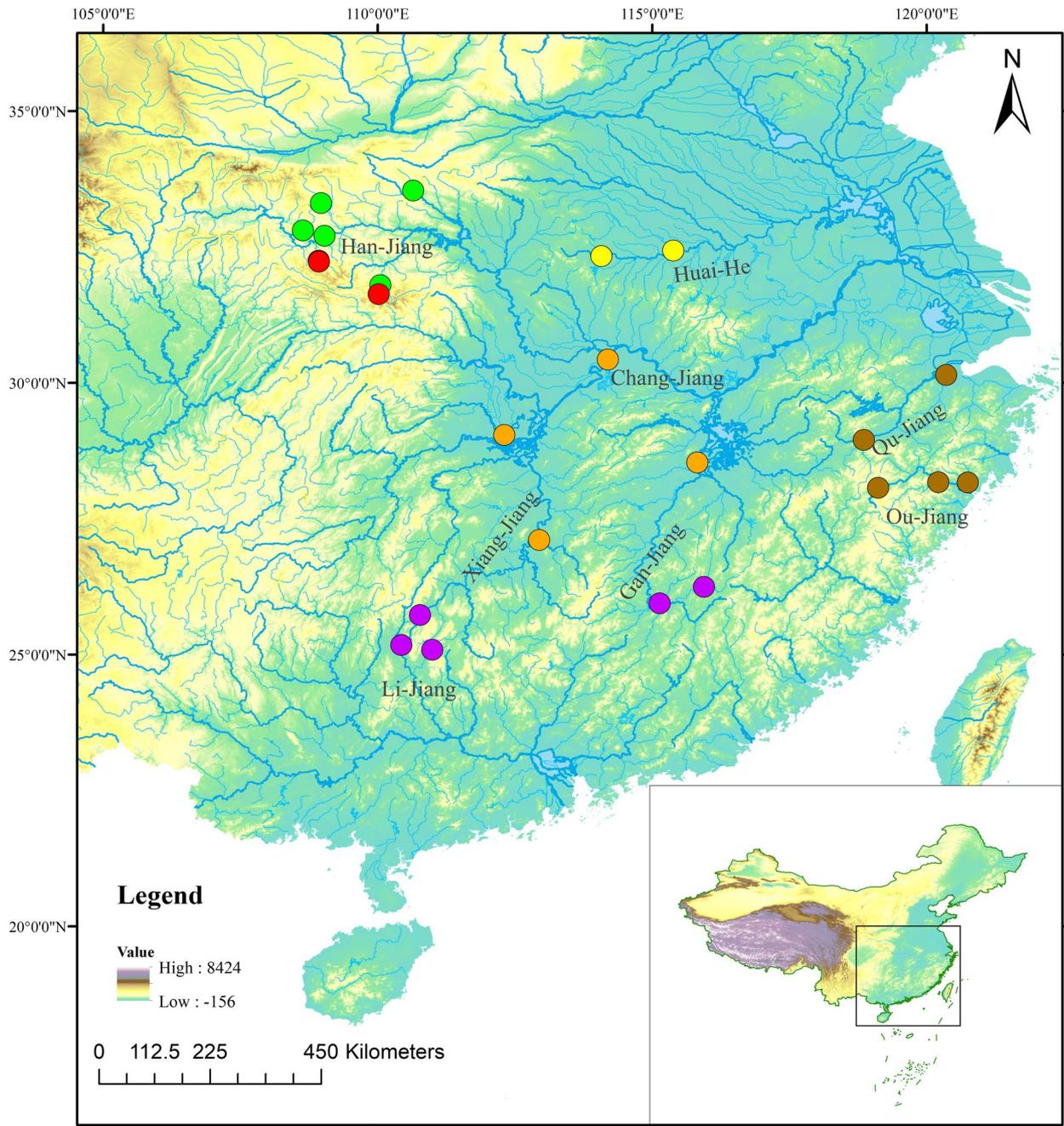


Fig. 1 Sampling localities of the six clades or putative species of the LCC. Distinct colors indicate different clades (brown: Coastal river Clade; green: Han-Jiang Clade I; orange: MCJ lowland Clade; purple: MCJ montane Clade; yellow: Huai-He Clade; and red: Han-Jiang Clade II; see details in the text)

about the species diversification of the LCC, and provide basic information for its diversity conservation.

Material and methods

Taxon sampling and laboratory procedures

With the permission of the fisheries departments of the Chinese government, we carried out fish sampling in

field surveys. A total of 170 specimens of the LCC (Tables S1-S2) were captured from 24 sampling locations across their known range in the subtropical region of China (Fig. 1). Detail information of the GPS coordinates of sampling locations and the date of sampling were given in Table S3. All fish were first totally immersed and deeply anesthetized with tricaine methane sulfonate (MS-222,

0.15 g/L) until losing all rhythmic opercular movements for a minimum of 5 min, then they were either preserved in 95% ethanol for DNA extraction or initially fixed in 10% formalin and then transferred to 70% ethanol for morphological examination and permanent curation. Among them, 111 and 132 specimens were respectively used for molecular and morphological analysis (Tables S1–S2). All voucher specimens are deposited in the ichthyological collection of the Institute of Hydrobiology (IHB), Chinese Academy of Sciences, Wuhan City, Hubei Province, China.

DNA extraction, amplification, and sequence analyses

Genomic DNA was extracted from fin clips stored in ethanol using the TIANamp Genomic DNA Kit (Tiangen Biotech, Beijing) with the recommended protocol. Two mitochondrial (mtDNA) genes (cytochrome b, *cyt b*; cytochrome c oxidase subunit I, *COI*) and three nuclear (nuDNA) genes (rhodopsin, *RH1*; early growth response 2b, *EGR2B*; myosin heavy polypeptide 6, *myh6*) were amplified, and sequenced. The PCR reaction solution (25 μ l) contained 1 μ l of each primer, 1 μ l template DNA, 12.5 μ l Master mix Taq (Beijing TsingKe Biotech Co. Ltd.) and 9.5 μ l double distilled water (dd H₂O). The information about primer pairs and thermocycling conditions are given in Table S4. Sequences were assembled in SeqMan v.7.1 implemented in the DNASTar software package and aligned using MAFFT v.7.4 [42]. Heterozygous positions in nuclear genes were resolved by PHASE algorithm [43] in DnaSP v.6.0 [44] (Table S5).

Phylogenetic analyses

PhyloSuite v.1.2 [45], which is a desktop platform integrating several phylogenetic and bioinformatic tools and therefore streamlining the entire analytical procedure, was used for phylogenetic analysis. Bayesian Inference (BI) and Maximum likelihood (ML) were used in our phylogenetic analyses based on separate mtDNA and nuDNA dataset. ModelFinder [46] was used to select the best-fit partition model (Edge-linked) applying BIC criterion [47] for every gene by codon position separately. The result of the best-fit partition model for every gene is provided in Table S6.

MrBayes v.3.2.6 [48] was used for Bayesian analysis following the best partitioning strategy assessed by ModelFinder, applying the Markov Chain Monte Carlo (MCMC) method with four chains (three hot chains and one cold chain) running simultaneously for 20,000,000 generations to calculate posterior probability. Trees were sampled for every 1000 generations. The initial 25% of sampled data were discarded as burn-in. Sufficient mixing of the chains was viewed to be reached when the mean standard deviation of split frequencies was below 0.01. The ML

phylogenies were inferred utilizing IQ-TREE v.1.6.8 [49] under Edge-linked partition model assessed by ModelFinder for 20,000 ultrafast [50] bootstraps. The Shimodaira–Hasegawa–like approximate likelihood-ratio test (SH-aLRT) [51] for LCC in ML tree was performed 1000 replicates using IQ-TREE.

The genetic distances for mtDNA genes were calculated with MEGA v.7.0 [52] between and within each clade of the LCC, based on the uncorrected p-distance model.

Population genetic analyses

To visualize the relationships among putative species, haplotype network was constructed for the combined mtDNA dataset using the TCS method [53] in PopART v.1.7 [54]. For the concatenated nuDNA dataset, population genetic structure was conducted in STRUCTURE v.2.3 [55]. The number of genetic clusters (K) was set from 1 to 6 with 10 independent runs using 1,200,000 MCMC generations after a burn-in of 200,000 repeats. Structure Harvester [56] program was used to identify the most likely value of K. Then, average coefficients of individuals across replicate runs were calculated using CLUMPP v.1.1.2 [57]. Finally, Distruct v.1.1 [58] was used to produce the genetic structure plots.

Species delimitation and species tree

Three species delimitation methods were used to delineate putative species: the assemble species by automatic partitioning (ASAP) [59], bayesian implementation of the PTP model (bPTP) [60] and multi-rate Poisson tree processes (mPTP) [61]. ASAP is the implementation of a hierarchical clustering algorithm which utilizes pairwise genetic distances (Jukes-Cantor, Kimura 2-parameter and simple p-distance). Species delineation analysis was performed through uploading aligned fasta file to the online server website of ASAP (<https://bioinfo.mnhn.fr/abi/public/asap>). mPTP incorporates different levels of intraspecific genetic diversity which were derived from differences in the evolutionary history or sampling of each species. Unrooted ML trees created by IQtree was analyzed in the online server website of mPTP (<http://mptp.h-its.org>). bPTP is an updated version of the original maximum likelihood PTP. It adds Bayesian support (BS) values to delimit species on the input tree. Unrooted BI trees created by MrBayes was analyzed in the online server website of bPTP (<https://species.h-its.org/ptp/>).

A multilocus coalescent-based species tree analysis was run via StarBeast [62] as implemented in BEAST v.2.6 [63]. Alignments of individual markers *cyt b*, *COI*, *RH1*, *EGR2B*, and *myh6* were used. An initial species assignment was carried out based on mtDNA phylogenetic analysis. The substitution models of each gene marker

were calculated by ModelFinder [46]. All clock models were set as uncorrelated lognormal relaxed clock. Two independent replicates were performed for 50,000,000 MCMC generations with sampling every 1000 generations in BEAST. Convergence and effective sample size (ESS) were checked utilizing TRACER v.1.7 (<http://beast.bio.ed.ac.uk/Tracer>) [64] to ascertain that all parameters of ESS values are above 200. A cladogram of the species tree was visualized in DensiTree v.2.2 [65].

Morphological analyses

Twenty-four morphometric measurements from 106 individuals of the LCC were selected. Following the methods of Kottelat [66], measurements taken on the left side of each individual were taken point to point with digital calipers directly linked to a data-recording computer and data recorded to the nearest 0.01 mm. The lateral head length (HL) and measurements of other parts of the body were given as percentages of standard length (SL), and measurements of parts of the head were expressed as proportions of HL. Principal components analysis (PCA) implemented in Past v.3.2 [67] with a variance–covariance matrix was used to analyze morphological variations and visualize the bidimensional distribution of the putative species. All measurements were log transformed prior to the analysis. To explore the effects of habitat types on specific morphological traits, ratios were calculated as follows: HL to SL, dorsal-fin length (DFL) to SL, upper caudal-fin lobe length (UCL) to SL, anal-fin length (AFL) to SL, and caudal-peduncle depth (CPD) to its length (CPL). Analysis of variance (ANOVA) and Tukey's HSD post hoc tests were used to evaluate variations of these traits in the LCC.

Vertebral numbers were counted from photographs of Micro-CT (X-ray-based micro-computed tomography) to compare morphological variations among 94 individuals of the LCC. The fishes were scanned with a Skyscan 1276 micro-CT instrument (Bruker microCT, Kontich, Belgium) using the following settings: source voltage, 55 kV; source current, 200 μ A; AI 0.25 mm filter; pixel size 8 μ m; rotation step, 0.3 degree. The images were then reconstructed with NRecon software (Bruker microCT, Kontich, Belgium).

Divergence time estimation

The age of major divergence events of the LCC was estimated in BEAST v.2.7 [63] based on the combined mtDNA and nuDNA dataset (cyt b, COI, RH1, EGR2B, and myh6). Two fossil records are used as calibration points to infer the time: the fossil of †*Cobitis longiptoralis* from the late early Miocene (set in analysis with normal. Mean=19.5 Mya; standard deviation=1.0) of East China, which is the best-preserved and earliest

cobitid fossil known so far [68], and the origin of the Cyprinini (sensu [69]) dating back at least to the late Eocene (set in analysis with normal prior. Mean=39.0 Mya; standard deviation=1.0) based on the fossil of †*Eoprocypris maomingensis* [70]. The Yule speciation model and the strict clock were selected as priors. The best substitution models of each gene were calculated by ModelFinder [46]. Three independent replicates were run for 50,000,000 MCMC generations with sampling every 1000 generations. Log and tree files were combined by LogCombiner v.2.7 (BEAST 2 package). ESS were again checked in TRACER v.1.7 (<http://beast.bio.ed.ac.uk/Tracer>) [64]. A maximum clade credibility (MCC) tree was built in TreeAnnotator v.2.7 (BEAST 2 package) with the first 10% of the trees discarded as burnin.

Reconstruction of habitat types

Since no available information regarding habitat requirements of species of the LCC, preferred habitat of each species is determined based on our observations during fieldworks and habitat-mediated morphological characters detected in this study. Three habitat types (low flow, medium flow and high flow) were selected for ancestral state reconstruction. Three habitat types were mapped onto the ML tree yielded from the concatenated dataset of five genes, with maximum likelihood approach and the model Mk1 of trait evolution (which allows multiple characters and equal probability of change among the habitat types scored) as implemented in Mesquite v.3.81 [71].

Results

Molecular species delimitation

Within the LCC, 4433 bp nucleotides were obtained from two mtDNA and three nuDNA genes: cyt b (1098 bp), COI (1026 bp), RH1 (764 bp), myh6 (735 bp), and EGR2B (810 bp). Their detailed information is given in Table 1. The monophyly of the LCC was strongly confirmed in our molecular phylogenetic analyses based on the separate mtDNA, nuDNA, and combined mtDNA and nuDNA dataset (Figs. 2, 3 and 4). The discordance between the mtDNA and nuDNA tree was unveiled.

For the mtDNA dataset, the BI or ML tree, inferred based on combined genes, were concordant in the revelation of six major well-resolved clades (Fig. 2B). The six clades of the LCC were significantly supported by the SH-aLRT ($p > 0.05$; Fig. 2B). The six clades corresponded to the six major haplotype groups yielded in the TCS analysis (Fig. 2A), without any haplotype shared among them. These clades also corresponded to five geographical populations (Fig. 1); two clades were detected for the Han-Jiang population. To facilitate result presentation and interpretation, they are hereafter referred to as: Coastal river, Han-Jiang I, Han-Jiang II, MCJ (middle

Table 1 Sequence characteristics of two mtDNA and three nuDNA genes in the LCC

Gene	Length	Variable sites	Parsimony informative sites	Mean frequency (%)			
				A	C	G	T
cyt b	1098bp	123	108	28.6	29.6	13.9	27.8
COI	1026bp	265	241	24.6	27.2	19.9	28.3
RH1	764bp	13	13	19.1	27.7	23.6	29.5
myh6	735bp	11	10	32.0	20.6	23.1	24.3
EGR2B	810bp	5	4	22.0	37.5	21.4	19.0

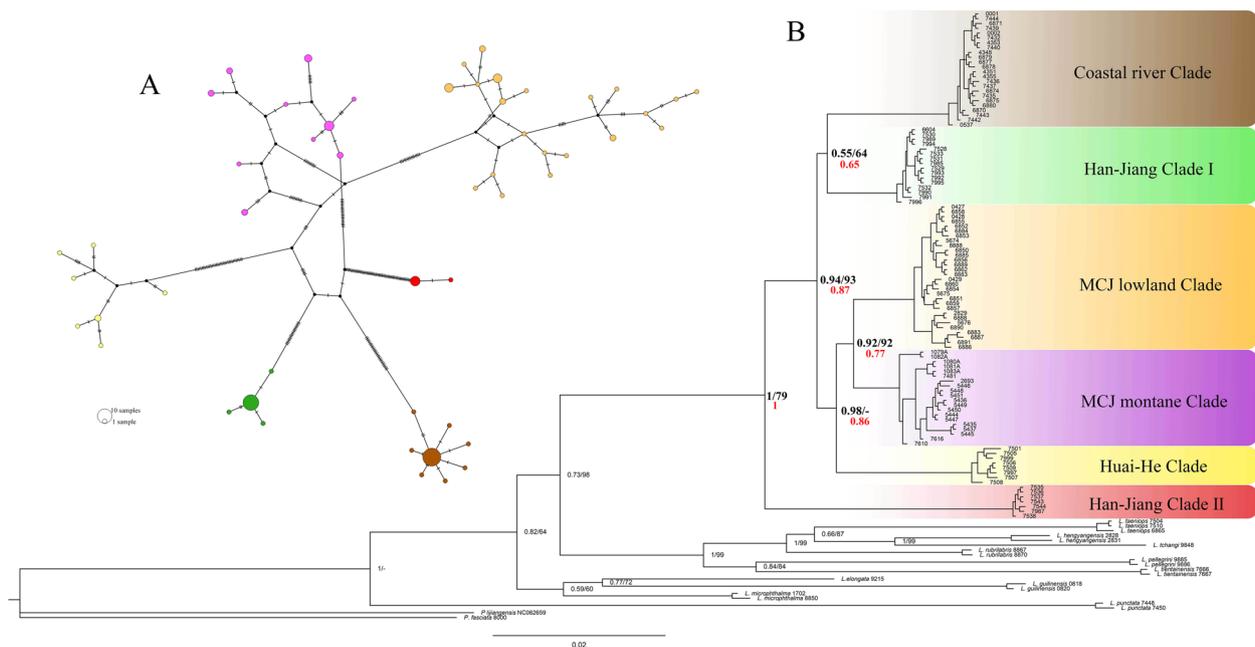


Fig. 2 Bayesian inference (BI) tree (A) and TCS network (B) inferred from the combined mtDNA dataset for the LCC. Only the BI tree is presented, given that its topology is generally congruent with that of the maximum likelihood (ML) tree. Black nodal numbers indicate posterior probabilities and bootstrap values obtained in the BI/ML analyses; only values ≥ 0.5 and $\geq 50\%$ are represented. Red nodal numbers indicate SH-aLRT support. Distinct colors represent the corresponding species. The circle sizes in TCS network represent the number of individuals

Chang-Jiang) lowland, MCJ montane (including the Li-Jiang), and Huai-He Clade (see Fig. 4 and the discussion section for details on species delineation for each clade).

The nuDNA dataset-based BI or ML tree generated a vast different phylogenetic structure (Fig. 3C). The Han-Jiang clade II (red) was strongly supported (PP=1; BS=95; SH-aLRT support=0.92) to nest with the remaining five clades. The Han-Jiang clade I (green) and Coastal river clade (brown) remained stable, but with poorly supports and low SH-aLRT support (PP<0.5; BS<50; $p < 0.05$), and the other three clades were intertwined to constitute a paraphyletic entity. The structure analysis of the nuDNA dataset indicated that the most likely numbers of clusters (K) were 2 and 5 (Fig. 3B). When K=2 was accepted, one cluster corresponded to

the Han-Jiang clade II, and the other was represented by the remaining clades (Fig. 3A), consistent with the phylogenetic analyses based on the nuDNA dataset. At K=5, the bar plots suggested that each of the Coastal river clade, and Han-Jiang clades I and II formed its own cluster, but the Huai-He, MCJ lowland and montane clades shared admixed ancestry in the remaining two clusters (yellow and purple) (Fig. 3A).

The genetic distance of two mtDNA genes based on the uncorrected p-distance model among the six clades of the LCC are provided in Tables 2 and 3. The genetic distance among them varied from 1.2% to 3.9% for the COI gene, and from 2.0% to 5.1% for the cyt b gene, greater than their intra-clade genetic distance of 0–0.3% and 0–0.4%. The genetic distance of three nuDNA genes

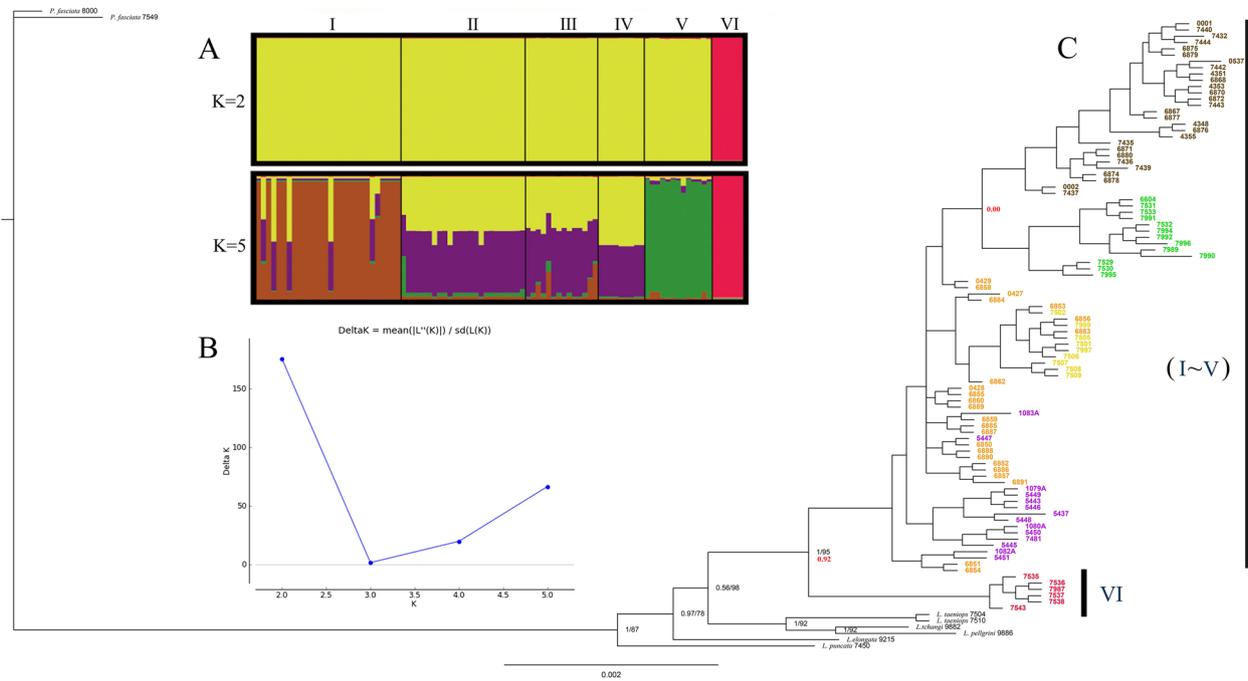


Fig. 3 Genetic structure plots of K=2 and 5 (A); ΔK values of the different number of clusters (B); and the BI tree (C) inferred from the concatenated nuDNA dataset for the LCC. Only the BI tree is presented, given that its topology is generally congruent with that of the ML tree. Black nodal numbers on major branches indicate posterior probabilities and bootstrap values obtained in the BI/ML analyses; only values ≥ 0.5 and $\geq 50\%$ are represented. Red nodal numbers indicate SH-aLRT support. I: Coastal river Clade, II: MCJ lowland Clade, III: MCJ montane Clade, IV: Huai-He Clade, V: Han-Jiang Clade I; and VI: Han-Jiang Clade II

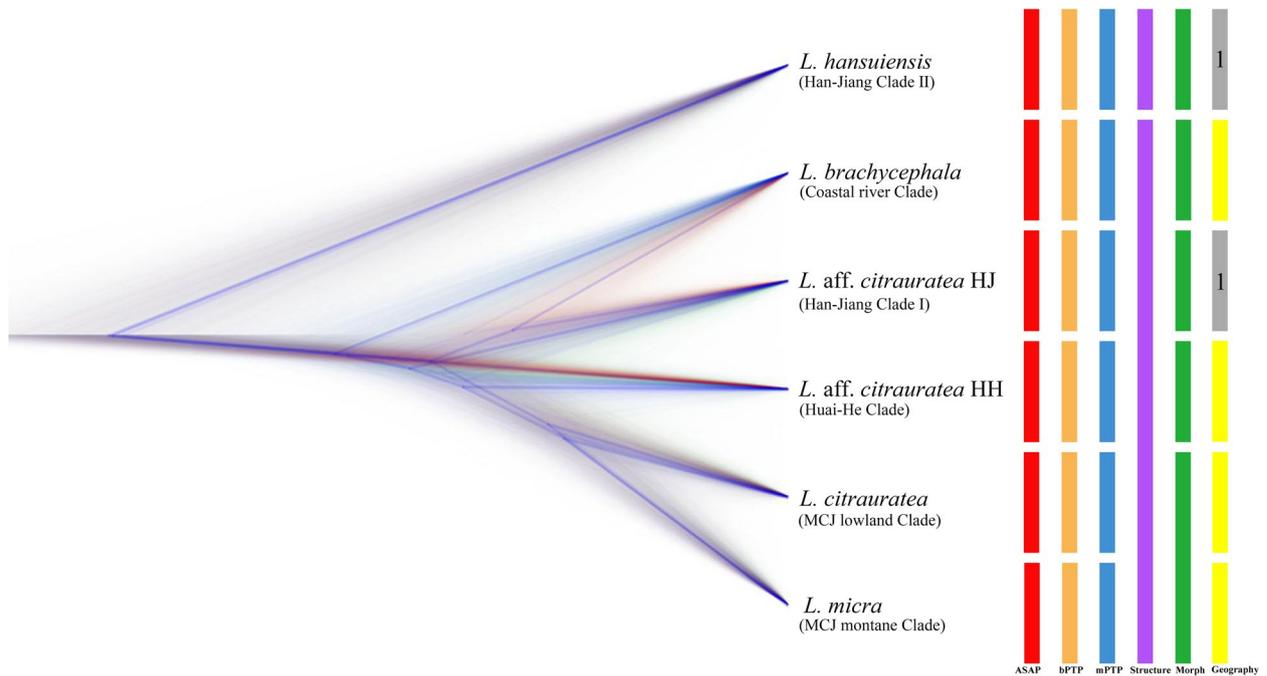


Fig. 4 A cladogram of the species tree with the consensus trees based on the combined mtDNA and nuDNA is visualized on the left. The results of multiple species delineation analysis (ASAP, bPTP, mPTP, structure, morphometry (including PCA and vertebral count) and geography are shown on the right. Colored bars represent the six clades (or putative species) of the LCC. Grey rectangles with the number "1" mean that two taxa should be collapsed into one

Table 2 Uncorrected intra- and interspecific p-distances of the COI gene computed by MEGA 7 among the six clades of the LCC

Species	Within	Between				
		1	2	3	4	5
1. <i>L. brachycephala</i> (Coastal river Clade)	0					
2. <i>L. aff. citrauratea</i> HJ (Han-Jiang Clade I)	0	0.022				
3. <i>L. citrauratea</i> (MCJ lowland Clade)	0.002	0.021	0.020			
4. <i>L. micra</i> (MCJ montane Clade)	0.003	0.018	0.020	0.012		
5. <i>L. aff. citrauratea</i> HH (Huai-He Clade)	0.002	0.031	0.028	0.023	0.020	
6. <i>L. hansuiensis</i> (Han-Jiang Clade II)	0	0.038	0.037	0.039	0.035	0.038

Table 3 Uncorrected intra- and interspecific p-distances of the cyt b gene computed by MEGA 7 among the six clades of the LCC

Species	Within	Between				
		1	2	3	4	5
1. <i>L. brachycephala</i> (Coastal river Clade)	0.000					
2. <i>L. aff. citrauratea</i> HJ (Han-Jiang Clade I)	0.001	0.024				
3. <i>L. citrauratea</i> (MCJ lowland Clade)	0.004	0.028	0.023			
4. <i>L. micra</i> (MCJ montane Clade)	0.003	0.030	0.021	0.020		
5. <i>L. aff. citrauratea</i> HH (Huai-He Clade)	0.003	0.037	0.026	0.032	0.033	
6. <i>L. hansuiensis</i> (Han-Jiang Clade II)	0.000	0.048	0.041	0.042	0.045	0.051

was not calculated due to too few variable sites to provide useful information for species delineation.

Species delineation analyses of the combined mtDNA dataset using ASAP, bPTP and mPTP methods showed very strong support for the recognition of the six clades as putative species; the same result was repeated in the multilocus coalescent-based species-tree analysis using combined mtDNA and nuDNA datasets (Fig. 4). Out of six distinct species here delineated, four are formerly described species: *L. citrauratea* (MCJ lowland Clade), *L. brachycephala* (Coastal river Clade), *L. hansuiensis* (Han-Jiang Clade II) and *L. micra* (MCJ montane Clade). The other two cryptic species are respectively represented by the Han-Jiang clade I and the Huai-He clade, which are temporarily named as *L. aff. citrauratea* HJ and *L. aff. citrauratea* HH.

Morphological distinction

Significant phenotypic disparities among the six clades, here detected within the LCC, were discerned in the PCA

performed for 24 morphometric measurements (Table 4). The first three components accounted for 93.78% of the total variance, of which 81.58%, 8.31% and 3.89% were explained by PC1, PC2 and PC3, respectively (Table 4). The visualization of the PC1 and PC2 showed that these clades except *L. citrauratea* and *L. micra* were separated from each other (Fig. 5A). Apart from *L. micra*, *L. citrauratea* and *L. aff. citrauratea* HJ, the other clades were distinct from each other in the scatter plot of the PC2 and PC3 (Fig. 5B). In other words, each of five geographic populations exclusively grouped together; *L. citrauratea* and *L. micra* are indiscernible in terms of morphometric data, but both had a distinct body coloration in dorsal and caudal fins and contrasting habitat propensities. The former, with hyaline dorsal and caudal fins, exhibits a preference slow-flowing or sluggish deep waters, while the latter has two (basal and median) black bands across the dorsal fin and two or three W-shaped black bands on the caudal fin, and prefer to dwell in fast-running waters. Significant differences among the six clades were found

Table 4 Loadings of morphological measurements on the first three principal components. Variables in bold indicate higher loading values

Variable	PC1	PC2	PC3
Standard length	0.209	-0.166	0.076
Body depth	0.262	0.276	0.019
Head length	0.182	0.051	0.051
Head depth at nape	0.210	0.189	-0.024
Head depth at eye	0.199	0.138	-0.081
Caudal-peduncle length	0.252	-0.509	-0.029
Caudal-peduncle depth	0.214	-0.115	0.047
Caudal-peduncle width	0.225	-0.073	-0.923
Dorsal-fin length	0.224	0.282	0.017
Pectoral-fin length	0.199	0.174	0.011
Pelvic-fin length	0.194	0.125	-0.008
Anal-fin length	0.211	0.278	-0.006
Upper caudal-lobe length	0.185	0.195	0.082
Predorsal length	0.205	-0.138	0.081
Prepectoral length	0.178	0.009	0.082
Prepelvic length	0.212	-0.057	0.103
Preanal length	0.205	-0.091	0.093
Vent to anal-fin origin	0.212	-0.450	0.197
Pelvic-fin insertion to anal-fin origin	0.208	-0.165	0.062
Snout length	0.169	0.046	0.066
Postorbital head length	0.187	-0.015	0.063
Eye diameter	0.191	0.073	0.052
Interorbital width	0.156	0.190	0.181
Median caudal-ray length	0.178	-0.128	-0.058
Cumulative variance (%)	81.575	8.311	3.887

in head length, caudal-peduncle length, upper caudal-fin lobe length, and dorsal- and anal-fin lengths among the PC2 axis (Table 4; Fig. 6) For the visualization of the PC1 and PC3, the six clades cannot be well distinguished (Fig. S1).

The six clades of the LCC also varied in vertebral counts made from Micro-CT photographs (see Fig. S2). *L. aff. citrauratea* HH, *L. citrauratea* and *L. micra* had fewer vertebral counts (32–34) than the other three clades (35–40), while the higher counts (37–40) were found in *L. hansuiensis* and *L. brachycephala* (Fig. 7); *L. aff. citrauratea* HJ had an intermediate vertebral count (35–36) (Fig. 7).

Divergence time estimation

The estimated time, with a 95% highest posterior density (HPD), is provided in Fig. 8. The most recent common ancestor (MRCA) of the LCC dated back to 3.34 Ma (95% HPD: 2.18–5.18). Following the deepest divergence of *L. hansuiensis*, *L. brachycephala* and *L. aff. citrauratea* HJ (a sister species pair) subsequently split from the remaining

three clades 1.95 Ma (95% HPD: 1.14–3.14). The estimated divergence time between *L. brachycephala* and *L. aff. citrauratea* HJ was 1.55 Ma (95% HPD: 0.77–2.62). *L. aff. citrauratea* HH branched off from the sister clade pair (*L. citrauratea* and *L. micra*) 1.82 Ma (95% HPD: 1.03–3.14), and the sister clades diverged from each other 1.17 Ma (95% HPD: 0.57–2.10).

Habitat transition

Three types of habitats (high, medium, and low flow) are identified for species of the LCC based on our observations during fieldworks and phenotypic distinctions (Figs. 5 and 9), corresponding to three ecotypes that are each represented by a pair of species: the high-flow ecotype (*L. brachycephala* and *L. hansuiensis*), the medium-flow ecotype (*L. micra* and *L. aff. citrauratea* HJ), and the low-flow ecotype (*L. citrauratea* and *L. aff. citrauratea* HH) (see the discussion).

The result of habitat reconstruction showed the LCC ancestor had experienced multiple habitat types (Fig. 9). The nearest relatives of the LCC, such as *L. compressicauda* and *L. rotundilobus*, share some morphological traits (e.g., the high vertebral counts and a fusiform body) specialized to the fast-running water lifestyle with *L. hansuiensis* at the base of the topology of the molecular phylogenetic tree. It is rational to speculate that high-flow species of the LCC represent the historical state of habitat; low-flow habitats are more specializations for *L. citrauratea* and *L. aff. citrauratea* HH. It was inferred that the LCC, during its southeastern invasion of the subtropical floodplains and hills of eastern China, experienced one transition from high-flow to medium-flow habitat in the ancestor of *L. aff. citrauratea* HJ, one transition from low-flow to medium-flow habitat in the ancestor of *L. micra*, and one independent transition from high-flow to low-flow habitat in *L. aff. citrauratea* HH and *L. citrauratea* (Fig. 9).

Discussion

Hidden species diversity

Unappreciated species diversity is unveiled in the current taxonomy of the LCC. The six clades, detected in the mtDNA dataset-based phylogenetic analysis (Fig. 2), were delimited utilizing different delineation methods as putative species (Fig. 4), concordant with the six major haplotype groups yielded in the TCS analysis (Fig. 2A). More importantly, the species delineation of the six clades is corroborated by their morphological distinctions. These clades correspond to five geographical populations (Fig. 1), which are discerned by the combination of five morphometric measurements (caudal-peduncle, head, and median fin lengths) and vertebral number (Figs. 5, 6 and 7). Two clades of the mid-lower Chang-Jiang basin

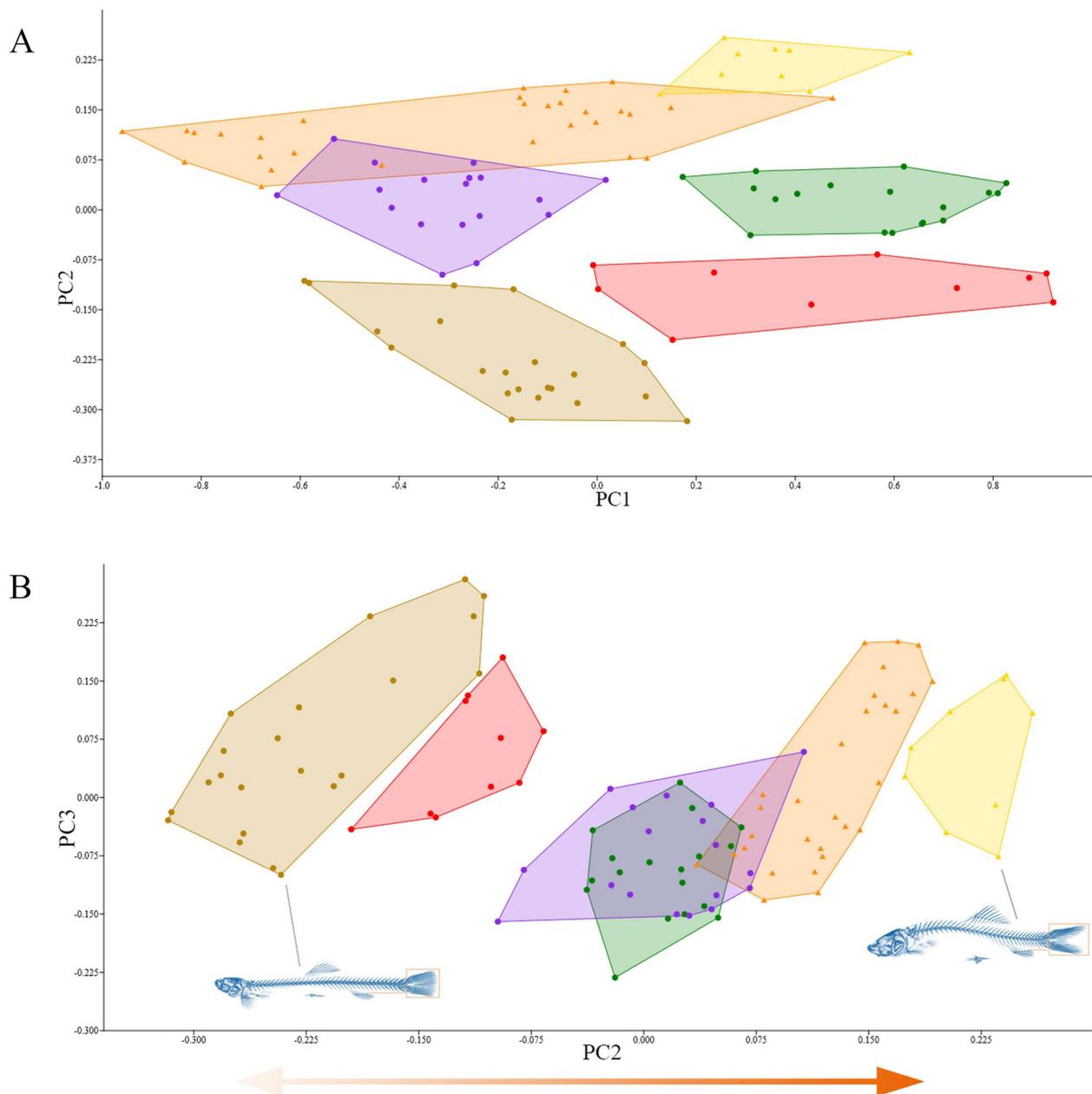


Fig. 5 Scatterplots of the PC1 and PC2 (A) and PC2 and PC3 (B) based on 24 morphometric measurements among the six clades (or putative species) of the LCC. Distinct colors indicate different species or clades, brown: Coastal river Clade (*L. brachycephala*); green: Han-Jiang Clade I (*L. aff. citrauratea* HJ); orange: MCJ lowland Clade (*L. citrauratea*); purple: MCJ montane Clade (*L. micra*); yellow: Huai-He Clade (*L. aff. citrauratea* HH); and red: Han-Jiang Clade II (*L. hansuiensis*). Caudal-peduncle length and caudal-fin shape at extreme values of the PC2 axis are shown in B

were identified, in terms of their dorsal- and caudal-fin colorations (Fig. 9) and contrasting habitat preferences. Multiple lines of evidence are thus integrated to support the delineation of six separate species in the LCC. Out of the six distinct species here delineated, four are previously described species: *L. citrauratea* (MCJ lowland Clade), *L. brachycephala* (Coastal river Clade), *L.*

hansuiensis (Han-Jiang Clade II) and *L. micra* (MCJ montane Clade). The other two cryptic species are respectively represented by the Han-Jiang clade I and the Huai-He clade, which are temporarily named as *L. aff. citrauratea* HJ and *L. aff. citrauratea* HH.

The cyt b gene distance between paired clades of the LCC (Table 3) are equal to or more than the threshold

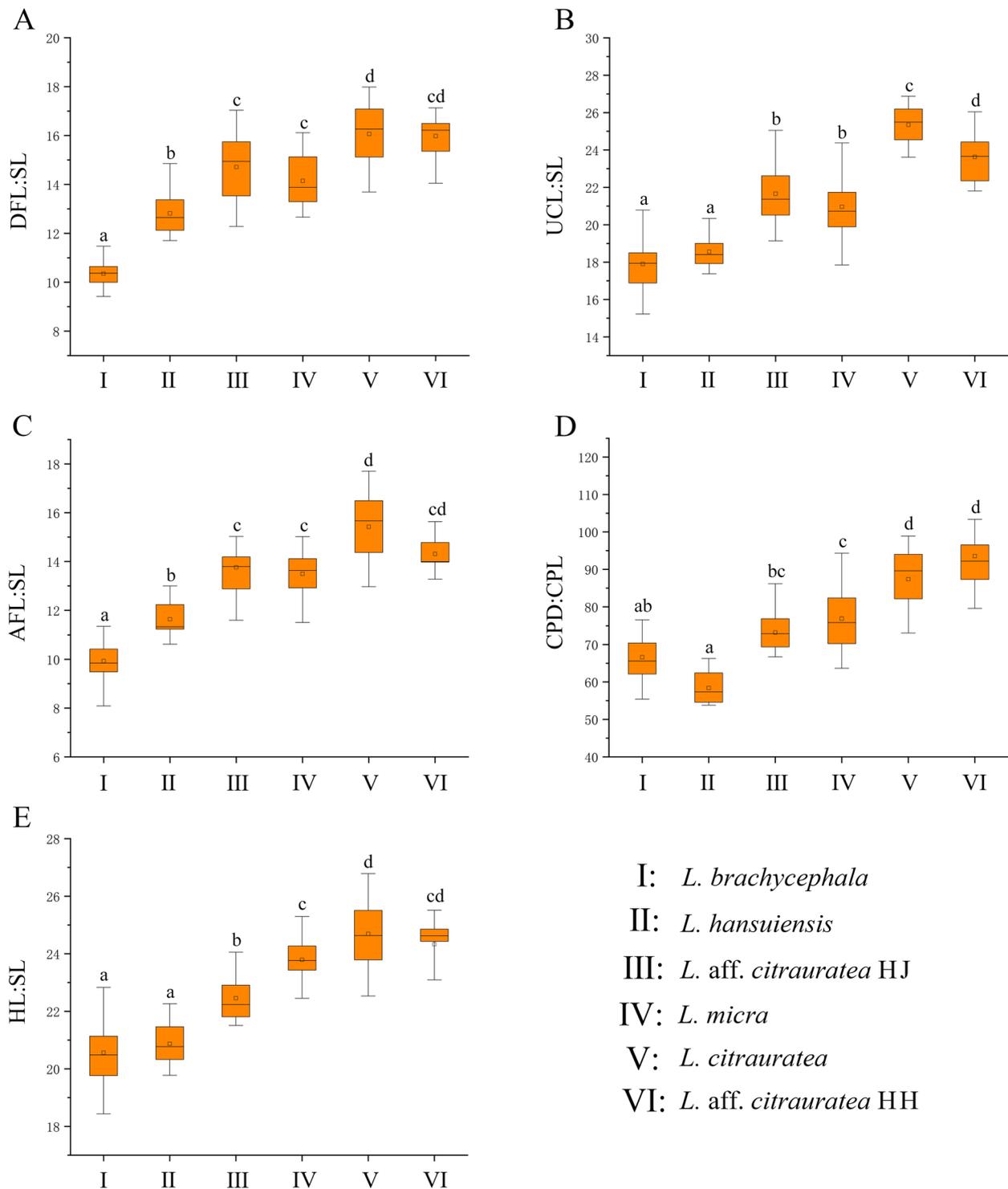


Fig. 6 The ratios of the DFL (dorsal-fin length) to SL (standard length) (**A**), UCL (upper caudal-lobe length) to SL (**B**), AFL (anal-fin length) to SL (**C**), CPD (caudal-peduncle depth) to CPL (caudal-peduncle length) (**D**), and HL (head length) to SL (**E**) for the six clades or putative species of the LCC. Analysis of variance (ANOVA) results are also shown. Different lowercase letters indicate significant differences between populations (Tukey's test: $p < .05$)

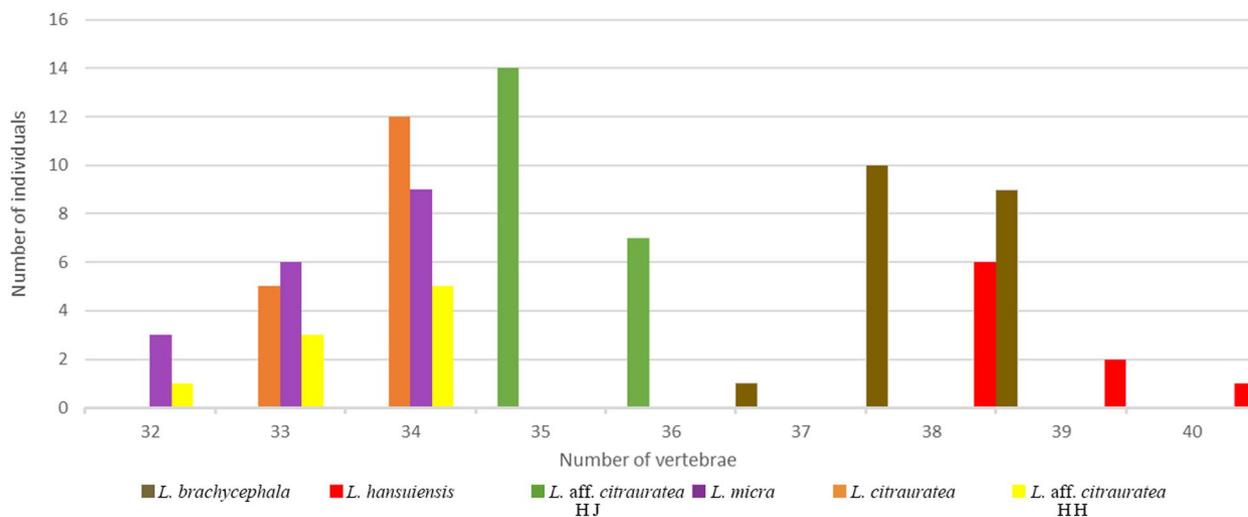


Fig. 7 Differences in vertebrae number among the six clades of the LCC. Colour denotes putative species

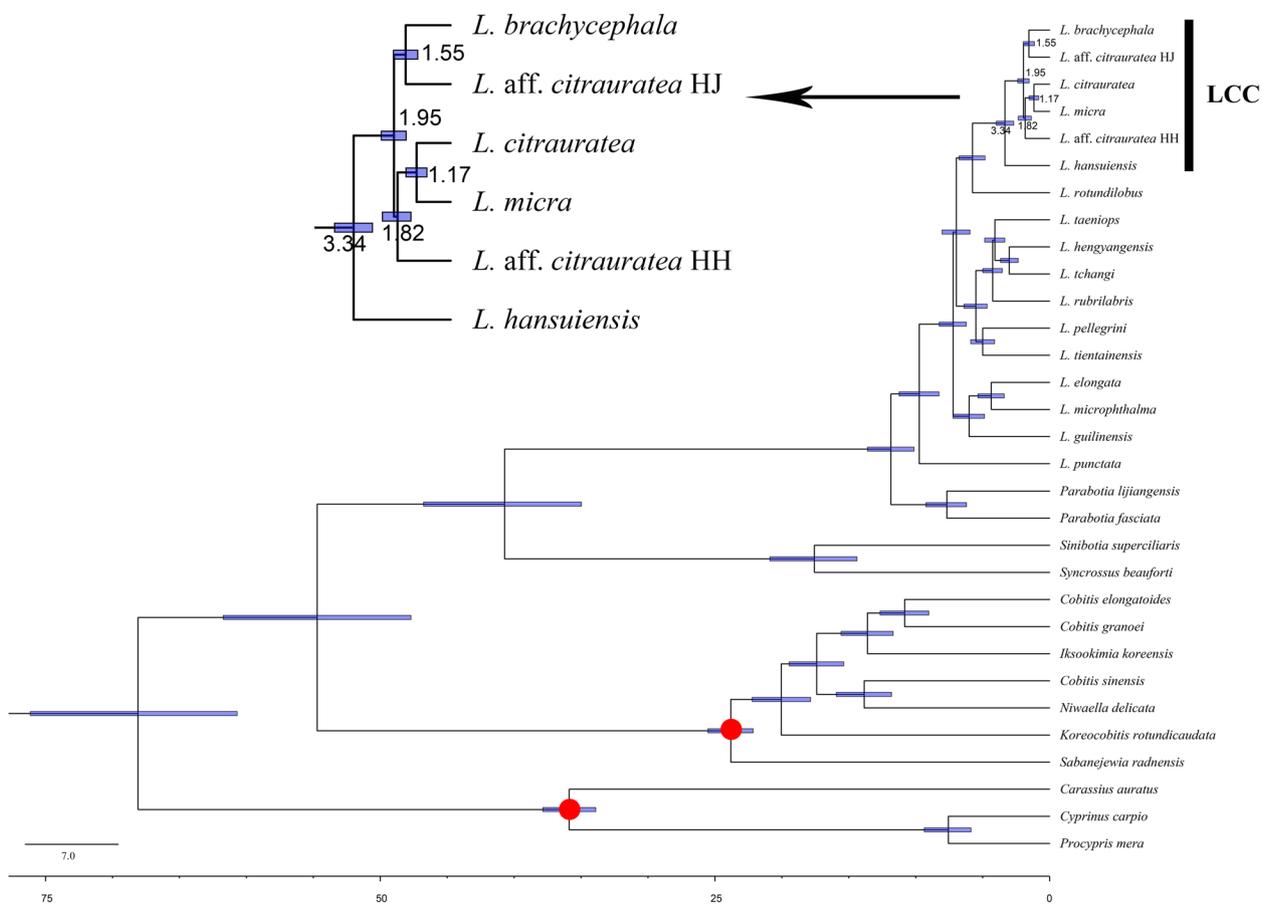


Fig. 8 Divergence time estimation of the main events inferred in BEAST2 for the LCC. Blue bars on the nodes represent 95% highest posterior density (HPD). Two red dots indicate fossil calibrated nodes

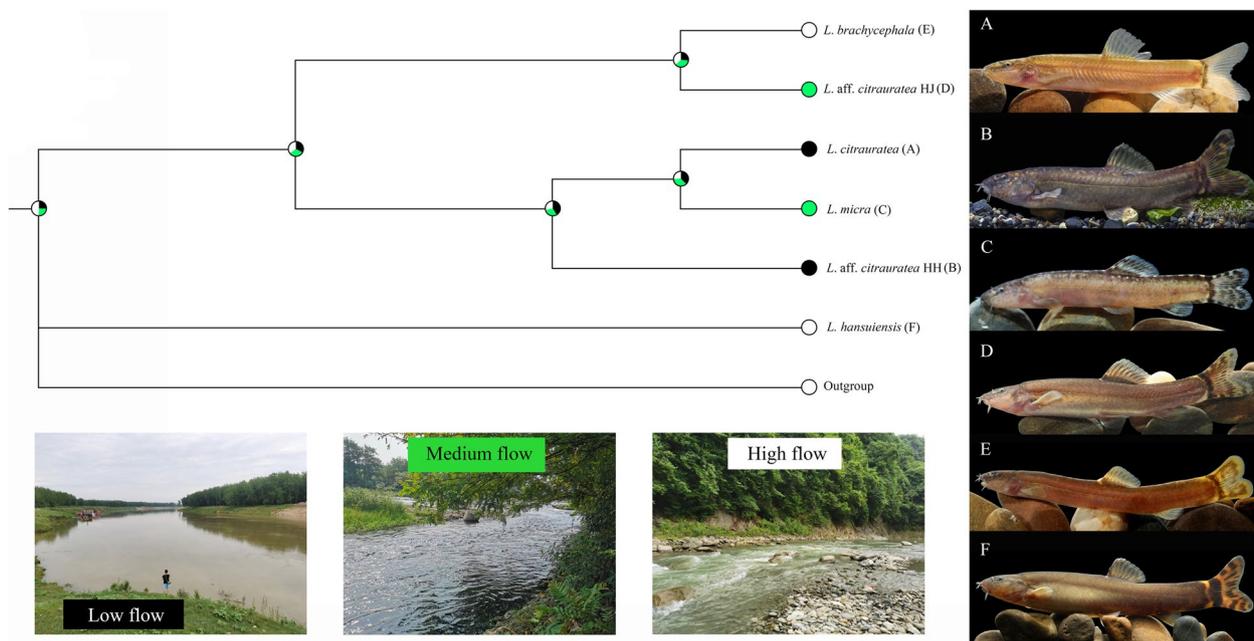


Fig. 9 Reconstruction of habitat types of the LCC using the ML trees obtained from the five combined genes. Color circle at the tip of the phylogenetic tree indicates current habitat types. Pie charts on the nodes represent the probability of the ancestral state for each node. Photographs of the three habitat types are depicted at the bottom. Lateral views of freshly-captured specimens of the LCC are positioned at the right

(2.0%), a cut-off value widely used to denote intraspecific variation [52]. Their COI gene distance ranged from 1.2% to 3.1% (Table 2); the minimum value, here calculated between the Coastal river clade and the MCJ montane or between the MCJ lowland and montane clade, was less than 2%, a phenomenon also encountered in freshwater fishes of subtropical China [72, 73]. The *Tachysurus similis* complex, a recently-diverged bagrid assemblage consisting of four allopatric clades, were recently delineated as four putative species, despite lower inter-clade genetic distances (1.1% to 1.3% for the *cyt b* gene) [41]. This case is exemplified by the LCC except for the Han-Jiang clade II. Given that the five clades constituted a recently splitting assemblage whose MRCA dated back to 1.95 Ma (95% HPD: 1.14–3.14) (Fig. 8), their low inter-clade genetic divergences (Tables 2 and 3) are unsurprising.

The Han-Jiang clade II remained constant in the nuDNA dataset-based phylogenetic and structure analyses (Fig. 3A-C), supporting its delineation as putative species to some extent. Although the remaining five clades cannot be distinguished well by nuclear genes, they were discernible by morphology, divergent habitats, ASAP, bPTP and mPTP methods. The Huai-He calde, Coastal river clade and Han-Jiang clade I were easily identified by multivariate morphometric and vertebral numbers (Figs. 5, 6 and 7). For the

MCJ lowland and montane clade, the ASAP, bPTP and mPTP methods (Fig. 4), coupled with their distinct body colorations and divergent preferred habitats as noted above, verify the delimitation of the two clades as separate species.

The mito-nuclear discordance observed between different datasets is plausibly elaborated by either incomplete lineage sorting or introgression, especially for groups of recent adaptive radiation [10]. It is commonly known that mtDNA locus is a more sensitive indicator of recent divergence compared with nuDNA loci which is more lagging to detect any change in population structure resulting from a lack of clear genetic structure among vertebrate groups [74]. This phenomenon is found in the suckers genus *Moxostoma* (Teleostei: Cypriniformes: Catostomidae), whose interspecific relationships were not well resolved by the conserved nuclear gene [75]. In our study, the mito-nuclear DNA mismatch for the Huai-He, Coastal river, Han-Jiang, MCJ montane and lowland clade is probably a result of incomplete lineage sorting instead of introgression, provided the recent divergence of six putative species within the LCC. Single nucleotide polymorphism (SNP) analysis can be useful to recover more genetic variation for these clades in the future research that is beyond the scope of this study, but underway.

Phenotypic congruence

The phylogenetic pattern of the six species, inferred based on molecular datasets (Figs. 2 and 4), does not reflect morphological similarity among them. Rather, these species exhibited a morphological change towards caudal-peduncle deepening, head elongation, and median (dorsal, anal, and caudal) fin prolongation along the PC2 axis (Figs. 5 and 6), accompanied by a general trend of decreasing vertebral counts (Fig. 7). These patterns coincide with a habitat transition from high- to low-flow environments. Both *L. citrauratea* and *L. aff. citrauratea* HH occur in lowland reaches of rivers, contrasting with the other four species usually found in montane streams (Figs. 1 and 9). Compared with *L. micra* and *L. aff. citrauratea* HJ (Figs. 6 and 7), both *L. brachycephala* and *L. hansuiensis* are well adapted to more fast-running waters as they had higher vertebral counts and shallower caudal peduncles (see explanation below). Given that habitat-mediated morphology is pervasive in fish [76, 77], the overall body plan in the LCC is closely related to habitat utilization.

Based on phenotypic differentiation and habitat divergence, three ecotypes can be identified for the LCC. The high-flow ecotype is represented by *L. hansuiensis* and *L. brachycephala* that constituted a distinct cluster in the PCA (Fig. 5), and had more vertebrae (37–40) than the remaining species (Fig. 7). The phenotypic extremity on the other is the low-flow ecotype shared with *L. aff. citrauratea* HH and *L. citrauratea* that are found in sluggish or slow-running deep waters (Fig. 9A–B). The medium-flow ecotype is exemplified by both *L. aff. citrauratea* HJ and *L. micra*, with a distinct morphology from the high-flow species (Fig. 6); the two hillstream species shared lower vertebral counts (32–36) with the low-flow species (Fig. 7). Morphological differentiations among the three ecotypes likely implies divergent selective pressure between habitats with stark contrasts at least in water flow and substrate structure (see elaboration below).

Each ecotype comprises two morphologically and ecologically similar, yet non-sister species within the LCC. This is putatively a result of phenotypic convergent evolution in response to similar selective pressure of habitat-specific utilization, as observed in West African *Labeo parvus*-like species, which include many distantly related putative species [18]. The mechanisms underlying this convergent pattern within the LCC warrant comprehensive research, emphasizing the need for rigorous quantification of ecotypes and a quantitative assessment of environmental variables. Nevertheless, the interplay of habitat structural complexity and water flow has been shown to prompt fish species in lotic environments to evolve towards greater degrees of convergence [78]. This may also apply to the species pairs gathered in each

ecotype of the LCC, where phenotypically convergent species are actually distantly related.

The most prominent phenotypic differentiation among the three ecotypes is the number of vertebrae, a major functional trait of fish structure closely associated with feeding, swimming, and predator avoidance. Its variations have been reported to be related to taxonomic groups, body shape and size, as well as environmental factors (e.g., temperature) [79, 80]. A positive correlation between body size and vertebral counts, previously reported for galaxiid fishes [81], was not found for the LCC as three relatively larger-sized species (*L. aff. citrauratea* HJ, and *L. aff. citrauratea* HH, and *L. hansuiensis*, with higher scores of the PC1; Fig. 5) differed in their vertebral counts (Fig. 7). The elevated vertebrae of *L. hansuiensis* and *L. brachycephala* did not align with the expectations derived from Jordan's rule, which states that vertebral count in fish increases with latitude (interpreted as a proxy for temperature) [79]. Three species (*L. aff. citrauratea* HH, *L. citraurata*, and *L. micra*) had no significant variation in this count (Fig. 7), despite their different latitudinal distributions (Fig. 1). A covariation of vertebral number with elevation, previously found for the neotropical characid genus *Rhoadsia* [82], was not replicated in *L. micra* and *L. citrauratea* from the highland and lowland reaches of the Xiang-Jiang and Gang-Jiang, respectively. Coexisting *L. hansuiensis* and *L. aff. citrauratea* HJ in the Han-Jiang differed in their vertebral counts (Fig. 7).

Rather, it has been unravelled that higher vertebral counts can give rise to finer segmentation and greater flexibility of the body, which is advantageous for interstitial species [83]. Increased flexibility enable fish to exhibit multiple bends in their bodies [84]. Notably more flexible bodies, resulting from heightened vertebral numbers in *L. hansuiensis* and *L. brachycephala*, facilitate fish to conceal themselves in crevices or beneath stones in high-flow habitats. Additionally, the two high-flow species possess longer caudal peduncles, as expected by the steady-unsteady swimming performance model for fish in lotic environments [76, 85]. The combination of an elongated caudal peduncle and a greater vertebral number results in enhanced body flexibility, which promotes swimming performance in lotic habitats. More likely, *L. brachycephala* and *L. hansuiensis* have convergently evolved a more flexible body form which is well-adapted for more faster-running waters compared to other species within the LCC, thereby facilitating their access to available underwater interstitial habitats.

Evolutionary history

The distribution pattern of the six species, detected for the LCC in the subtropical China, maybe has been

shaped by geological, climatic, and ecological factors (e.g. habitat transition). The first species to split within the LCC is *L. hansuiensis* currently found in the upper Han-Jiang (Figs. 2, 3 and 4). Its diverging time, dated at 3.34 Ma (95% HPD: 2.18–5.18) (Fig. 8), is broadly consistent with the tectonic activities in the Daba Mountains during 3.0–3.4 Ma [86], following the rapid uplift of the Himalaya-Tibetan Plateau and the evolution of Asian monsoons during 2.6–3.6 Ma [87]. Apparently, this species evolved in response to the geological uplift of the Daba Mountains and the subsequent development of the Han-Jiang, given its preference for lotic habitats.

The sister group of *L. hansuiensis* forms a monophyletic lineage that experienced rapid diversification. A significant portion of genetic divergence is observed within the five included species, but with only a minor fraction of this divergence distributed among them (Fig. 2B). This phenomenon is generally interpreted as evidence for rapid diversification [88]. Their most recent common ancestor (MRCA) was dated at 1.95 Ma (95% HPD: 1.14–3.14) (Fig. 8), which broadly aligns with the Pleistocene glacial history of East Asia. Unlike North America and Western Europe, most regions of East Asia did not experience extensive ice coverage during the Pleistocene [89]. However, lineage diversification over this period has been documented for numerous animal species [29]. Climatic fluctuations, leading to range expansion and contraction, have been identified as key drivers of interspecific and intraspecific differentiations in many freshwater fishes [40, 41, 90]. As a result of these repeated climatic events, the ancestor of the LCC (excluding *L. hansuiensis*) was able to extend from the upper Han-Jiang further southeast, eventually occupying the present distribution range of its descendant species and undergoing rapid diversification.

The sister species relationship between *L. aff. citrauratea* HJ (from the Han-Jiang) and *L. brachycephala* (from coastal rivers of southern Zhejiang Province), as revealed by our molecular analysis (Fig. 8), presents an unusual phylogeographical pattern, particularly given the vast geographical gap between them. This pattern becomes even more intriguing when considering geological evidence that suggests the modern East Asian continental shelf was once linked into a single landmass during the Pleistocene [91]. The fish faunistic similarity across the East Asian mainland, Korean Peninsula, and Japanese Archipelago supports the hypothesis of an ancient hydrological network of rivers and lakes in this region [92, 93]. The existence of this ancient alluvial system is further suggested by the shared occurrence of species such as the Southeast China population of the freshwater crab *Sinopotamon yangtsekiense* [94] and the bullhead catfish *Tachysurus albomarginatus* [41] in the Huai-He and

lower Chang-Jiang basins, and Coastal rivers of southern Zhejiang Province. In addition, a potential past linkage between the upper Han-Jiang and Huai-He is implicated in the network geometry that their mainstreams flow in the same direction (Fig. 1). In this context, the ancestor of the allopatric sister pair *L. aff. citrauratea* HJ and *L. brachycephala* is speculated to have a greater southeastern extension of the upper Han-Jiang, through the Huai-He and lower Chang-Jiang basin, into Coastal rivers of southern Zhejiang Province. The descendant population in the Han-Jiang ultimately evolved into *L. aff. citrauratea* HJ and those in the Ou-Jiang and Qu-Jiang into *L. brachycephala*. However, the hypothesis regarding the once east-flowing upper Han-Jiang into the Huai-He requires further verification through the discovery of freshwater fishes or other aquatic species shared with both rivers. The sister species relationship between *L. aff. citrauratea* HJ and *L. brachycephala* remains to be testified in future research utilizing a broader array of molecular markers. Such research will provide deeper insights into the evolutionary history and biogeography of subtropical China.

Leptobotia aff. citrauratea HH (from the Huai-He) displays an allopatry with its sister group made up of *L. citrauratea* and *L. micra*, both mainly known so far from the mid-lower Chang-Jiang basin (Fig. 1). Acting as the water parting between the mid-lower Chang-Jiang and Huai-He basin, Tongbai-Dabie Mountains are commonly known to arise during the early Mesozoic [95], much more earlier than the splitting time between *L. aff. citrauratea* HH and its sister group dated at 1.82 Ma (95% HPD: 1.03–3.14) (Fig. 8). Their allopatric pattern is unlikely attributed to the vicariant event triggered by the uplift of these mountains. Provided that the time to the MRCA of *L. aff. citrauratea* HH, *L. citrauratea*, and *L. micra* (Fig. 8) is consistent with the current consensus on tectonic evolution in East Asia, their widespread ancestor in the Huai-He and mid-lower Chang-Jiang basin was possibly achieved, via the aforementioned ancient alluvial system existing at 3.5 to 0.8 Ma [96], during the period of glaciations. As glaciers retreated, the ancestral population became isolated and rapidly evolved into separate species.

Habitat transition is unveiled here to play a crucial role in species diversification of the LCC. The ancestral state reconstruction showed that at least two independent transitions to intermediate-flow habitat in *L. aff. citrauratea* HJ and *L. micra*, and one transition from high-flow to low-flow habitat in ancestor of *L. aff. citrauratea* HH and *L. citrauratea* (Fig. 9). Habitat divergence may exert strong selection on the body form of this group, thereby leading to significant phenotypic differences between the sister species pair (*L. citrauratea* and *L. micra*; *L. aff. citrauratea* HJ and *L. brachycephala*) (Figs. 5, 6 and 7).

These findings further confirms the key role of divergent selection between discrete habitats in species diversification as shown previously [31]. Given habitat transitions underwent by the LCC during its colonizations into subtropical floodplains and hills of eastern China, invasion of new habitats provides ecological opportunity for the LCC to rapidly diversify. Just as pointed out formerly [32], the ecological adaptation to novel environmental conditions is followed by the derivation of new taxa from ancestral populations.

Conclusion

This study conducts a comprehensive exploration of cryptic diversity within the LCC using integrative taxonomy, so laying a solid foundation for species recognition and subsequent conservation efforts. Our results underscore the critical importance of understanding cryptic species and their habitats, as these species may possess unique habitat requirements and face distinct survival threats. To enhance our understanding of forces driving species diversity and phenotypic convergence, we recommend that future research focus on the specific ecological niches and behaviors of these species. Given the distribution of the LCC in the subtropical zone of eastern China, our study leverages regional diverse habitats and complex geological history to explore cryptic diversity and evolutionary processes underlying phenotypic uniformity. Our findings contribute both theoretically and empirically to biodiversity research in this region.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12862-025-02362-2>.

Supplementary Material 1.

Acknowledgements

Our sincere thanks should be given to Man Wang, Yi Liu, Wei-Han Shao, Peng Shan and Chu-Yi Zhang (IHB) for assistance with field work and laboratory analyses. We are grateful to Martin Reichard (Institute of Vertebrate Biology, Czech Academy of Sciences, Brno, Czech Republic) for his help with polishing the manuscript. We also express our gratitude to Hao-Jun Chen for help in providing photograph of the *L. aff. citrauretea* HH.

Authors' contributions

DM Guo and E Zhang conceived this study and conducted the data analysis. X Gong and L Cao contributed to data collection and phylogenetic analysis. The manuscript was drafted by DM Guo and E Zhang, and revised by WJ Yi. All authors read and approved the final version of the manuscript.

Funding

This research was granted partially by the special foundation for National Science and Technology Basic Research Program of China (2019FY101900), and also by the National Nature Sciences Foundation of China (NSFC no: 31872200).

Data availability

All new gene sequences obtained in this study are deposited in the GenBank and their accession numbers are listed in the Supplementary Material 1 (see Table S1; accession number: PP491479-PP491875, PP492717-PP492816).

Declarations

Ethics approval and consent to participate

All fish specimens used for this study were collected in accordance with the Chinese Laboratory Animal Welfare and Ethics Animal Welfare Laws. All procedures in this study were performed with the approval of Animal Care and Use Committee of Institute of Hydrobiology, Chinese Academy of Sciences (IHB, CAS, Protocol No. 2016–018). This study was completed in compliance with ARRIVE guidelines.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 15 November 2024 Accepted: 5 March 2025

Published online: 17 March 2025

References

- Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. Cryptic species as a window on diversity and conservation. *Trends Ecol Evol.* 2007;22:148–55.
- Fišer C, Robinson CT, Malard F. Cryptic species as a window into the paradigm shift of the species concept. *Mol Ecol.* 2018;27:613–35.
- Vogler A, Monaghan M. Recent advances in DNA taxonomy. *J Zool Syst Evol Res.* 2007;45:1–10.
- Chambers EA, Hillis DM. The multispecies coalescent over-splits species in the case of geographically widespread taxa. *Syst Biol.* 2020;69:184–93.
- Sukumaran J, Knowles LL. Multispecies coalescent delimits structure, not species. *Proc Natl Acad Sci U S A.* 2017;114:1607–12.
- Edwards DL, Knowles LL. Species detection and individual assignment in species delimitation: can integrative data increase efficacy? *Proc Biol Sci.* 2014;281:20132765.
- Hundsdoerfer AK, Lee KM, Kitching IJ, Mutanen M. Genome-wide SNP data reveal an overestimation of species diversity in a group of Hawkmoths. *Genome Biol Evol.* 2019;11:2136–50.
- Ricklefs RE. Evolutionary diversification and the origin of the diversity-environment relationship. *Ecology.* 2006;87:S3–13.
- Schluter D. Evidence for ecological speciation and its alternative. *Science.* 2009;323:737–41.
- Struck TH, Feder JL, Bendiksy M, Birkeland S, Cerca J, Gusarov VI, Kistenich S, Larsson KH, Liow LH, Nowak MD, Stedje B, Bachmann L, Dimitrov D. Finding evolutionary processes hidden in cryptic species. *Trends Ecol Evol.* 2018;33:153–63.
- Esquerre D, Scott KJ. Parallel selective pressures drive convergent diversification of phenotypes in pythons and boas. *Ecol Lett.* 2016;19:800–9.
- Moen DS, Morlon H, Wiens JJ. Testing convergence versus history: Convergence dominates phenotypic evolution for over 150 million years in frogs. *Syst Biol.* 2015;65:146–60.
- Rundle HD, Nagel L, Wenrick Boughman J, Schluter D. Natural selection and parallel speciation in sympatric sticklebacks. *Science.* 2000;287:306–8.
- Wang Y, Wang Y, Cheng X, Ding Y, Wang C, Merilä J, Guo B, von der Heyden S. Prevalent introgression underlies convergent evolution in the diversification of *Pungitius* sticklebacks. *Mol Biol Evol.* 2023;40:msad026.
- Alter SE, Brown B, Stiassny ML. Molecular phylogenetics reveals convergent evolution in lower Congo River spiny eels. *BMC Evol Biol.* 2015;15:224.
- Armbruster JW, Stout CC, Hayes MM. An empirical test for convergence using African barbids (Cypriniformes: Cyprinidae). *Evol Ecol.* 2016;30:435–50.

17. Liyandja TLD, Armbruster JW, Poopola MO, Stiassny MLJ. Evolutionary convergence in body shape obscures taxonomic diversity in species of the African *Labeo forskalii* group: Case study of *L. parvus* Boulenger 1902 and *L. ogunensis* Boulenger 1910. *Journal of Fish Biology*. 2022;101:898–913.
18. Muschick M, Indermaur A, Salzburger W. Convergent evolution within an adaptive radiation of cichlid fishes. *Curr Biol*. 2012;22:2362–8.
19. Peterson RD, Sullivan JP, Hopkins CD, Santaqueria A, Dillman CB, Pirro S, Betancur RR, Arcila D, Hughes LC, Orti G. Phylogenomics of bony-tongue fishes (Osteoglossomorpha) shed light on the craniofacial evolution and biogeography of the weakly electric clade (Mormyridae). *Syst Biol*. 2022;71:1032–44.
20. Feng C, Tang Y, Liu S, Tian F, Zhang C, Zhao K. Multiple convergent events created a nominal widespread species: *Triplophysa stoliczkae* (Steindachner, 1866) (Cobitoidea: Nemacheilidae). *BMC Evol Biol*. 2019;19:177.
21. Winemiller KO, Fitzgerald DB, Bower LM, Pianka ER. Functional traits, convergent evolution, and periodic tables of niches. *Ecol Lett*. 2015;18:737–51.
22. Winemiller KO. Ecomorphological diversification in lowland fresh-water fish assemblages from 5 biotic regions. *Ecol Monogr*. 1991;61:343–65.
23. Zhao SQ. *Physical Geography of China*. Beijing: Science Press; 1986.
24. Ye Z, Zhu G, Damgaard J, Chen X, Chen P, Bu W. Phylogeography of a semi-aquatic bug, *Microvelia horvathi* (Hemiptera: Veliidae): an evaluation of historical, geographical and ecological factors. *Sci Rep*. 2016;6:21932.
25. Webb T, Bartlein PJ. Global changes during the last 3 million years: climatic controls and biotic responses. *Annu Rev Ecol Syst*. 1992;23:141–73.
26. Hewitt GM. Genetic consequences of climatic oscillations in the Quaternary. *Philos Trans R Soc Lond B Biol Sci*. 2004;359:183–95.
27. Li Y, Zhang XW, Fang YM. Landscape features and climatic forces shape the genetic structure and evolutionary history of an oak species (*Quercus chenii*) in East China. *Front Plant Sci*. 2019;10:1060.
28. Lu L, Mao L-F, Yang T, Ye J, Liu B, Li H-L, Sun M, Miller J, Mathews S, Hu H-H, Niu Y, Peng D-X, Chen Y, Smith S, Chen M, Xiang K, Le CT, Dang V-C, Lu A-M, Chen Z. Evolutionary history of the angiosperm flora of China. *Nature*. 2018;554:234–8.
29. Fu J, Wen L. Impacts of Quaternary glaciation, geological history and geography on animal species history in continental East Asia: a phylogeographic review. *Mol Ecol*. 2023;32:4497–514.
30. St John CA, Buser TJ, Kee VE, Kirilchik S, Bogdanov B, Neely D, Sandel M, Aguilar A. Diversification along a benthic to pelagic gradient contributes to fish diversity in the world's largest lake (Lake Baikal, Russia). *Mol Ecol*. 2022;31:238–51.
31. Seehausen O, Wagner CE. Speciation in freshwater fishes. *Annu Rev Ecol Syst*. 2014;45:621–51.
32. Kakioka R, Kokita T, Tabata R, Mori S, Watanabe K. The origins of limnetic forms and cryptic divergence in Gnathopogon fishes (Cyprinidae) in Japan. *Environ Biol Fishes*. 2012;96:631–44.
33. Tang QY, Yu D, Liu HZ. *Leptobotia zebra* should be revised as *Sinibotia zebra* (Cypriniformes: Botiidae). *Zool Res*. 2008;29:1–9.
34. Bohlen J, Šlechtová V. *Leptobotia bellacauda*, a new species of loach from the lower Yangtze basin in China (Teleostei: Cypriniformes: Botiidae). *Zootaxa*. 2016;4205:065–72.
35. Bohlen J, Šlechtová V. *Leptobotia micra*, a new species of loach (Teleostei: Botiidae) from Guilin, southern China. *Zootaxa*. 2017;4250:90–100.
36. Guo DM, Zhang E. Re-description of the loach species *Leptobotia citraurata* (Teleostei, Botiidae), with the description of *L. brachycephala* from southern Zhejiang Province, China. *Zookeys*. 2021;1017:89–109.
37. Guo DM, Cao L, Zhang E. Descriptions of two new species of the botiid genus *Leptobotia* Bleeker, 1870 (Teleostei: Cypriniformes) from South China. *J Fish Biol*. 2024;104:433–49.
38. Šlechtová V, Bohlen J, Tan HH. Families of Cobitoidea (Teleostei; Cypriniformes) as revealed from nuclear genetic data and the position of the mysterious genera *Barbucca*, *Psilorhynchus*, *Serpenticobitis* and *Vaillantella*. *Mol Phylogenet Evol*. 2007;44:1358–65.
39. Šlechtová V, Bohlen J, Freyhof J, Rab P. Molecular phylogeny of the Southeast Asian freshwater fish family Botiidae (Teleostei: Cobitoidea) and the origin of polyploidy in their evolution. *Mol Phylogenet Evol*. 2006;39:529–41.
40. Li M, Yang X, Ni X, Fu C. The role of landscape evolution in the genetic diversification of a stream fish *Sarcocheilichthys parvus* from Southern China. *Front Genet*. 2023;13:1075617.
41. Shao WH, Cheng JL, Zhang E. Eight in one: hidden diversity of the bagrid catfish *Tachysurus albomarginatus* s.l. (Rendhal, 1928) widespread in lowlands of South China. *Front Genet*. 2021;12:713793.
42. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80.
43. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. 2001;68:978–89.
44. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol*. 2017;34:3299–302.
45. Zhang D, Gao F, Jakovlic I, Zou H, Zhang J, Li WX, Wang GT. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour*. 2020;20:348–55.
46. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14:587–9.
47. Luo A, Qiao H, Zhang Y, Shi W, Ho SY, Xu W, Zhang A, Zhu C. Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. *BMC Evol Biol*. 2010;10:242.
48. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61:539–42.
49. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32:268–74.
50. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol*. 2013;30:1188–95.
51. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59:307–21.
52. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33:1870–4.
53. Clement M, Posada D, Crandall K. TCS: a computer program to estimate gene genealogies. *Mol Ecol*. 2000;9:1657–9.
54. Leigh JW, Bryant D, Nakagawa S. Popart: full-feature software for haplotype network construction. *Methods Ecol Evol*. 2015;6:1110–6.
55. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–59.
56. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour*. 2011;4:359–61.
57. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 2007;23:1801–6.
58. Rosenberg NA. Distruct: a program for the graphical display of population structure. *Mol Ecol Notes*. 2003;4:137–8.
59. Puillandre N, Brouillet S, Achaz G. ASAP: assemble species by automatic partitioning. *Mol Ecol Resour*. 2021;21:609–20.
60. Zhang J, Kapli P, Pavlidis P, Stamatakis A. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*. 2013;29:2869–76.
61. Kapli P, Lutteropp S, Zhang J, Kobert K, Pavlidis P, Stamatakis A, Flouri T. Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics*. 2017;33:1630–8.
62. Heled J, Drummond AJ. Bayesian inference of species trees from multilocus data. *Mol Biol Evol*. 2010;27:570–80.
63. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchene S, Fourment M, Gavryushkina A, Heled J, Jones G, Kuhnert D, De Maio N, Matschiner M, Mendes FK, Muller NF, Ogilvie HA, du Plessis L, Popinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu CH, Xie D, Zhang C, Stadler T, Drummond AJ. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*. 2019;15:e1006650.

64. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in bayesian phylogenetics using tracer 1.7. *Syst Biol*. 2018;67:901–4.
65. Bouckaert R, Heled J. DensiTree 2: Seeing trees through the forest. *bioRxiv*. 2014;012401:1–11.
66. Kottelat M. Indochinese nemacheilines: a review of nemacheiline loaches (Pisces: Cypriniformes) of Thailand, Burma, Laos, Cambodia and southern Vietnam. Mhncen: Pfeil; 1990.
67. Hammer O, Harper DAT, Ryan PD. PAST-Palaeontological statistics, ver. 1.89. *Paleontologia Electron*. 2009;4:1–92.
68. Chen G, Chang MM, Wang Q. Redescription of †*Cobitis longipectoralis* Zhou, 1992 (Cypriniformes: Cobitidae) from late early Miocene of East China. *Sci China Earth Sci*. 2010;53:945–55.
69. Yang L, Mayden RL, Sado T, He S, Saitoh K, Miya M. Molecular phylogeny of the fishes traditionally referred to *Cyprinini sensu stricto* (Teleostei: Cypriniformes). *Zool J Linn Soc*. 2010;39:527–50.
70. Chen G, Chang MM, Liu H. Revision of *Cyprinus maomingensis* Liu 1957 and the first discovery of *Procypris*-like cyprinid (Teleostei, Pisces) from the late Eocene of South China. *Sci China Earth Sci*. 2015;58:1123–32.
71. Maddison WP, Maddison DR. Mesquite: a modular system for evolutionary analysis. Version 3.8.1. 2023. Available online: <http://www.mesquiteproject.org>.
72. Chen Y, Chen Y. Three new species of cobitid fish (Teleostei, Cobitidae) from the River Xinjiang and the River Le'anjiang, tributaries of Lake Poyang of China, with remarks on their classification. *Folia Zool*. 2013;62:83–95.
73. Huang SP, Chen IS, Zhao YH, Shao KT. Description of a new species of the Gudgeon genus *Microphysogobio* Mori 1934 (Cypriniformes: Cyprinidae) from Guangdong Province, southern China. *Zool J Linn Soc*. 2018;57:e58.
74. Zink RM, Barrowclough GF. Mitochondrial DNA under siege in avian phylogeography. *Mol Ecol*. 2008;17:2107–21.
75. Chen WJ, Mayden RL. Phylogeny of suckers (Teleostei: Cypriniformes: Catostomidae): further evidence of relationships provided by the single-copy nuclear gene *IRBP2*. *Zootaxa*. 2012;3586:195–210.
76. Langerhans RB. Predictability of phenotypic differentiation across flow regimes in fishes. *Integr Comp Biol*. 2008;48:750–68.
77. Rundle H, Nosil P. Ecological speciation. *Ecol Lett*. 2005;8:336–52.
78. Bower LM, Saenz DE, Winemiller KO. Widespread convergence in stream fishes. *Biol J Linn Soc*. 2021;133:863–79.
79. McDowall RM. Jordan's and other ecogeographical rules, and the vertebral number in fishes. *J Biogeogr*. 2008;35:501–8.
80. Ward AB, Brainerd EL. Evolution of axial patterning in elongate fishes. *Biol J Linn Soc*. 2007;90:97–116.
81. McDowall RM. Variation in vertebral number in galaxiid fishes (Teleostei: Galaxiidae): a legacy of life history, latitude and length. *Environ Biol Fishes*. 2003;66:361–81.
82. Aguirre WE, Young A, Navarrete-Amaya R, Valdiviezo-Rivera J, Jiménez-Prado P, Cuccalón RV, Nugra-Salazar F, Calle-Delgado P, Borders T, Shervette VR. Vertebral number covaries with body form and elevation along the western slopes of the Ecuadorian Andes in the Neotropical fish genus *Rhoadsia* (Teleostei: Characidae). *Biol J Linn Soc*. 2019;126:706–20.
83. Wagner M, Bracun S, Skofitsch G, Kovacic M, Zogaris S, Iglesias SP, Sefc KM, Koblmüller S. Diversification in gravel beaches: a radiation of interstitial clingfish (*Gouania*, *Gobiesocidae*) in the Mediterranean Sea. *Mol Phylogenet Evol*. 2019;139:10652.
84. Ward AB, Azizi E. Convergent evolution of the head retraction escape response in elongate fishes and amphibians. *Zoology*. 2004;107:205–17.
85. Hendry AP, Kelly ML, Kinnison MT, Reznick DN. Parallel evolution of the sexes? Effects of predation and habitat features on the size and shape of wild guppies. *J Evol Biol*. 2006;19:741–54.
86. Shi XH, Yang Z, Dong YP, Qu HJ, Zhou B, Cheng B. Geomorphic indices and longitudinal profile of the Daba Shan, northeastern Sichuan Basin: evidence for the late Cenozoic eastward growth of the Tibetan Plateau. *Geomorphology*. 2020;353:107031.
87. An ZS, Kutzbach JE, Prell WL, Porter SC. Evolution of Asian monsoons and phased uplift of the Himalayan Tibetan plateau since Late Miocene times. *Nature*. 2001;411:62–6.
88. Sakka H, QuÉRÉ JP, Kartavtseva I, Pavlenko M, Chelomina G, Atopkin D, Bogdanov A, Michaux J. Comparative phylogeography of four *Apodemus* species (Mammalia: Rodentia) in the Asian Far East: evidence of Quaternary climatic changes in their genetic structure. *Biol J Linn Soc*. 2010;100:797–821.
89. Zhao J, Wang J, Yang X. Review, progress and prospect of the quaternary glaciations in eastern China (east to 105° E). *J Glaciology Geocryology*. 2019;41:75–92.
90. Chen W, Zhong Z, Dai W, Fan Q, He S. Phylogeographic structure, cryptic speciation and demographic history of the sharpbelly (*Hemiculter leucisculus*), a freshwater habitat generalist from southern China. *BMC Evol Biol*. 2017;17:216.
91. Zhang J, Wan S, Clift PD, Huang J, Yu Z, Zhang K, Mei X, Liu J, Han Z, Nan Q, Zhao D, Li A, Chen L, Zheng H, Yang S, Li T, Zhang X. History of Yellow River and Yangtze River delivering sediment to the Yellow Sea since 3.5 Ma: Tectonic or climate forcing? *Quaternary Sci Rev*. 2019;216:74–88.
92. Hu J, Li H, Sakai H, Mukai T, Young Suk H, Li C. Molecular phylogenetics of the freshwater sleepers *Odontobutis* (Gobiiformes: Odontobutidae) and its implications on biogeography of freshwater ichthyofauna of East Asia. *Mol Phylogenet Evol*. 2023;186:107871.
93. Jeon HB, Song H, Suk HY, Bang IC. Phylogeography of the Korean endemic (Cypriniformes: Gobionidae): the genetic evidence of colonization through Eurasian continent to the Korean Peninsula during Late Plio-Pleistocene. *Genes Genomics*. 2022;44:709–19.
94. Shi B, Pan D, Sun Y, Liu X, Lv X, Cumberlandidge N, Sun H. Regional climates shape the biogeographic history of a broadly distributed freshwater crab species complex. *J Biogeogr*. 2021;48:1432–47.
95. Wang YS, Wang HF, Sheng Y, Xiang BW. Early Cretaceous uplift history of the Dabie orogenic belt: evidence from pluton emplacement depths. *Sci China Earth Sci*. 2014;57:1129–40.
96. Zheng H, Clift PD, Wang P, Tada R, Jia J, He M, Jourdan F. Pre-Miocene birth of the Yangtze River. *Proc Natl Acad Sci U S A*. 2013;110:7556–61.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.