

Article

Origin of Polyploidy, Phylogenetic Relationships, and Biogeography of Botiid Fishes (Teleostei: Cypriniformes)

Lei Yang ^{1,*}, Richard L. Mayden ² and Gavin J. P. Naylor ¹¹ Florida Museum of Natural History, University of Florida, 1659 Museum Rd., Gainesville, FL 32611, USA² Biology Department, Saint Louis University, 3507 Laclede Avenue, St. Louis, MO 63103, USA

* Correspondence: lyang@floridamuseum.ufl.edu

Simple Summary: In the present study, we found that fishes in the subfamily Botiinae are likely of allotetraploid origin, meaning they possess four sets of chromosomes derived from distinct parental species. This tetraploidization event appears to have occurred only once, in the common ancestor of the subfamily. Additionally, our results provide new insights into the phylogenetic relationships and biogeographical history of the family Botiidae.

Abstract: Botiidae is a small family of freshwater fishes distributed across Southeast Asia, South Asia, and East Asia. It comprises two subfamilies: the diploid Leptobotiinae and the tetraploid Botiinae. Whether species in the Botiinae are autotetraploids or allotetraploids and how many polyploidization events occurred during the evolution of this subfamily remain open questions. The phylogenetic relationships and biogeography of the Botiidae also require further investigation. In the current study, we compared phylogenetic trees constructed using DNA sequences from the mitochondrial genome and five phased nuclear genes. We also performed whole genome sequencing for two tetraploid species: *Chromobotia macracanthus* and *Yasuhikotakia modesta*. Genome profiling of five botiine species suggests that they are likely of allotetraploid origin. Nuclear gene tree topologies indicate that the tetraploidization of the Botiinae occurred only once in the common ancestor of this subfamily. Although the possible maternal progenitor and paternal progenitor of the Botiinae cannot be determined, the subfamily Leptobotiinae can be excluded as a progenitor. The gene trees built in this study generally agree on the following sister group relationships: Leptobotiinae/Botiinae, *Leptobotia*/*Parabotia*, *Chromobotia*/*Botia*, *Yasuhikotakia*/*Syncrossus*, and *Sinibotia*/*Ambastakia*. Clades formed by the last two generic pairs are also sisters to each other. Timetree analyses and ancestral range reconstruction suggest that the family Botiidae might have originated in East Asia and Mainland Southeast Asia approximately 51 million years ago and later dispersed to South Asia and the islands of Southeast Asia.

Keywords: allopolyploidy; genome; loach; mitogenome; nuclear; timetree



Academic Editor: Qinbo Qin

Received: 15 April 2025

Revised: 5 May 2025

Accepted: 9 May 2025

Published: 11 May 2025

Citation: Yang, L.; Mayden, R.L.; Naylor, G.J.P. Origin of Polyploidy, Phylogenetic Relationships, and Biogeography of Botiid Fishes (Teleostei: Cypriniformes). *Biology* **2025**, *14*, 531. <https://doi.org/10.3390/biology14050531>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As an important evolutionary and ecological force, polyploidy (whole genome duplication) has received more and more attention in recent years (e.g., [1–4]). Compared to plants, much fewer polyploids have been found in fishes [5,6], and research on polyploid fishes has lagged behind that of polyploid plants in many aspects. Polyploids can be broadly categorized into two main types: autopolyploids and allopolyploids. Autopolyploids possess more than two sets of chromosomes derived from a single species, whereas allopolyploids carry more than two sets of chromosomes originating from different species,

often through hybridization followed by genome doubling. For some known polyploid fish groups (e.g., *Tor*, *Probarbus*, *Spinibarbus*, and their allies), it has been unclear whether they are autopolyploids or allopolyploids until relatively recently [7,8]. For some other polyploid fish groups, such as the polyploid members of the family Botiidae, the situation remains unresolved.

Botiidae is a primary freshwater fish family distributed across Southeast Asia, South Asia, and East Asia. Some species, such as the Clown loach (*Chromobotia macracanthus*), are very popular in the aquarium trade. The family currently comprises around 65 valid species in two subfamilies and eight genera [9]. The subfamily Leptobotiinae includes two genera, *Leptobotia* and *Parabotia*, while the subfamily Botiinae consists of six genera, *Ambastaia*, *Botia*, *Chromobotia*, *Sinibotia*, *Syncrossus*, and *Yasuhikotakia*. Early studies on the karyotypes of botiid species (e.g., [10–13]) found that species in Leptobotiinae are diploids ($2n = 50$), whereas species in Botiinae are tetraploids ($2n = c.100$). However, none of these studies addressed whether fishes in Botiinae are autotetraploids or allotetraploids. Kaewmad et al. (2014) [14] explicitly stated that they are autotetraploids but provided no convincing evidence to support this claim, merely citing the aforementioned early studies (i.e., [10–12]). In contrast, the molecular phylogenetic study by Šlechtová et al. (2006) and the molecular cytogenetic work by Sember et al. (2018) both concluded that the underlying mechanism of polyploidization in Botiinae remains unclear [15,16]. More recently, Bitsikas et al. (2024) and Lv et al. (2024) published the genome assemblies for several botiine species (i.e., *Chromobotia macracanthus*, *Yasuhikotakia modesta*, *Ambastaia sidhimunki*, *Botia almorhae*, *B. kubotai*, and *Sinibotia reevesae*) [17,18]. However, they did not comment on whether these species are autopolyploids or allopolyploids.

Over the years, numerous methods have been developed to distinguish the two types of polyploids. Earlier studies relied on taxonomic investigation, chromosome counts, and cytological observation (e.g., [19]). In recent years, however, most studies have employed molecular cytogenetics, molecular genetics, or genomics approaches. A detailed review of many of these methods and their limitations is provided in [20]. In the current study, we applied a phylogeny-based genetic method and a k-mer-based genomic method to investigate whether botiines are autotetraploids or allotetraploids. Previous studies have shown that comparing phylogenetic trees built from mitochondrial genes and nuclear genes can help distinguish different types of polyploidization [8,21,22]. That is because mitochondrial genes are maternally inherited, whereas nuclear genes are biparentally inherited. Such analyses require separating the different nuclear gene copies in polyploid species and using them to construct gene trees. The second method utilized GenomeScope 2.0 [23], which distinguishes autotetraploids from allotetraploids based on the relative abundance of different nucleotide heterozygosity forms in k-mers generated from unassembled sequencing reads.

Resolving the phylogenetic relationships of a relatively small group like Botiidae may superficially seem straightforward. Indeed, several phylogenetic studies have already been conducted on the Botiidae using either mitochondrial genes alone (e.g., [15,24]) or both mitochondrial and nuclear genes (e.g., [16,25]). However, the use of nuclear genes in phylogenetic studies of polyploid species presents significant challenges. Every single-copy nuclear gene has two copies (four alleles) in tetraploids, like the subfamily Botiinae. These gene copies need to be sorted out properly, and only orthologous sequences can be used to resolve sister relationships among taxa. Unfortunately, previous studies using nuclear genes (e.g., [25]) did not separate the different gene copies and, thus, may be compromised by paralogy issues. In the present study, we attempted to separate the different gene copies of five single-copy nuclear genes: EGR2B (early growth response protein 2B gene), EGR3 (early growth response protein 3 gene), RAG1 (recombination activating gene 1, exon 3), RAG2 (recombination activating gene 2), and IRBP2 (interphotoreceptor retinoid-binding

protein gene 2). We also reconstructed the phylogenetic relationships of botiids using whole mitochondrial genome sequences. By comparing the topologies of nuclear gene trees built using all gene copies and the mitochondrial tree, we were able to better resolve the phylogenetic relationships within the family Botiidae.

To our knowledge, no previous studies have estimated a timetree specifically focusing on the family Botiidae. A timetree is needed to know the divergence times of Botiidae, its subfamilies, genera, and species. It is also important if we want to investigate when the whole genome duplication (polyploidization) event might have happened. Moreover, botiid species have a very interesting distributional pattern. The diploid subfamily Leptobotiinae is endemic to East Asia, while the tetraploid subfamily Botiinae is mainly distributed in Southeast Asia and South Asia [15]. Both subfamilies contain about the same number of valid species (33 vs. 32; see [9]). A timetree will allow us to explore the biogeographical history of botiid fishes and may shed light on the potential roles that polyploidization might have played during the diversification of these fishes.

In summary, the objectives of this study are threefold: (1) to investigate whether members of the subfamily Botiinae are autotetraploids or allotetraploids and to assess how many polyploidization events have occurred during the evolution of this subfamily; (2) to reconstruct the phylogenetic relationships among botiid fishes using DNA sequences from both mitochondrial genomes and phased nuclear genes; and (3) to build a timetree and use it to reconstruct the biogeographical history of botiid fishes.

2. Materials and Methods

2.1. Taxon Sampling, DNA Extraction, PCR, Library Preparation, and Sequencing

A total of 21 new samples representing both subfamilies and all eight genera of the family Botiidae were used for sequencing in this study. The detailed sample information can be found in Tables S1 and S2. The total genomic DNA of these samples was extracted using the E.Z.N.A. Tissue DNA Kit (Omega Bio-tek, Inc., Norcross, GA, USA) following the manufacturer's instructions. The whole mitochondrial genome (mitogenome) sequences were obtained using a gene capture method followed by next-generation sequencing [26,27]. Briefly, an aliquot of the extracted total genomic DNA (0.5–3 µg) of each sample was sheared to ~500 bp using a Covaris M220 Focused-ultrasonicator (Covaris, Inc., Woburn, MA, USA). After that, the dual-indexed genomic libraries were prepared, and they were subjected to one round of gene capture using RNA baits designed for vertebrate mitogenome enrichment [27].

For 14 out of the 21 samples, we tried to separate the different gene copies of five single-copy nuclear genes (i.e., EGR2B, EGR3, RAG1, RAG2, and IRBP2) using the “Next-Generation Sequencing followed by data phasing” method detailed in [28]. Primers and protocols used for the PCR amplification of these genes can be found in [29,30]. For each sample, PCR amplicons of the five nuclear genes were pooled together based on the brightness of bands on the agarose gel and then sheared to c.500 bp using acoustic ultrasonication on a Covaris M220 Focused-ultrasonicator (Covaris, Inc., Woburn, MA, USA). Dual-indexed Illumina sequencing libraries were then prepared for all samples following [26]. For the indexing step, the number of PCR cycles was set to five to minimize PCR-induced errors. The individual mitogenome libraries and nuclear gene libraries were pooled in equimolar ratios, respectively, and sent to the Interdisciplinary Center for Biotechnology Research of the University of Florida (UF-ICBR) for quality control and paired-end 300 bp sequencing on an Illumina MiSeq benchtop sequencer (Illumina, Inc., San Diego, CA, USA).

2.2. Mitogenome and Nuclear Gene Assembling

Adaptors and low-quality reads were trimmed from the demultiplexed raw reads using Trim Galore v0.6.4 [31]. For each sample, the duplicated reads were discarded, and unique reads were mapped to a reference in Geneious v.11.1.5 (<https://www.geneious.com>). For the mitogenome assembling of each species, the mitogenome sequence of a closely related species from the GenBank (*Ambastatia sidthimunki*, KP319024; *Botia lohachata*, KP729183; *Chromobotia macracanthus*, AB242163; *Sinibotia robusta*, KP979711; *Syncrossus beauforti*, AP011231; *Yasuhikotakia modesta*, KY131962; *Leptobotia elongata*, JQ230103; or *Parabotia fasciata*, KM393223) was used as the reference. For nuclear gene assembling, the following sequences from the GenBank were combined into a single FASTA file and used as the reference: EU409737 (*Botia dario*) for EGR2B, KP694684 (*Syncrossus beauforti*) for EGR3, EU711110 (*Sinibotia superciliaris*) for RAG1, KP696287 (*Yasuhikotakia lecontei*) for RAG2, and JN177266 (*Chromobotia macracanthus*) for IRBP2. For each sample, the mapping results of each gene were sorted out automatically by Geneious. We then deleted the references and reads that were not directly mapped to the reference for each read alignment. Next, we used the R package *copyseparator* 1.2.0 [28] to separate and assemble the two gene copies (if they existed) from the read alignment of each gene of each sample. The function “sep_assem” was used in most cases. The “copy_number” was set as 2, “read_length” was set as 300, and “overlap” was set as 225. In some cases, the function “copy_separate” was used, followed by manual assembly.

2.3. Datasets and Phylogenetic Analyses

2.3.1. Mitogenome Dataset

The “Mitogenome dataset” was built by combining the 21 mitogenome sequences newly obtained in this study and the 99 mitogenome sequences (14 for Botiidae, 85 for outgroups) downloaded from the GenBank (Tables 1 and S1). The 120 whole mitogenome sequences were aligned following the same procedures detailed in [32]. The final alignment is 14,888 bp in length, with 11,391 sites from the 13 protein-coding genes, 2096 sites from the two rRNA genes, and 1401 sites from the 22 tRNA genes.

Table 1. Taxon sampling and characteristics of different datasets used in this study.

	Mitogenome	EGR2B	EGR3	RAG1	RAG2	IRBP2	7-Nuclear	All-Gene
Botiidae	35	14	14	14	13	17	14	14
Outgroup	85	18	17	18	10	17	18	18
Total species	120	32	31	32	23	34	32	32
Total sequences	120	40	41	32	23	34	32	32
Nucleotides (bp)	14,888	828	953	1497	1314	864	7240	22,128
Variable characters (bp)	8372	348	393	650	613	446	3070	10,266
Parsimony-informative characters (bp)	7188	246	288	513	396	320	2155	7909

2.3.2. Individual Nuclear Gene Datasets

A separate dataset was built for each of the five nuclear genes. Taxon sampling information and characteristics of each dataset can be found in Tables 1 and S2. For the genes EGR2B and EGR3, most tetraploid species have sequences from two gene copies (I and II). For the genes RAG1, RAG2, and IRBP2, only one gene copy was identified for each species.

2.3.3. Concatenated Nuclear Gene Dataset

For each of the EGR2B dataset and the EGR3 dataset, a Copy I dataset was created by removing all Copy II sequences, and a Copy II dataset was created by removing all Copy I sequences. The “7-nuclear dataset” was then constructed by concatenating the EGR2B Copy I and II datasets, the EGR3 Copy I and II datasets, the RAG1 dataset, the RAG2 dataset, and the IRBP2 dataset. The 7-nuclear dataset contains 14 botiids and 18 outgroups and is 7240 bp in length (Table 1).

2.3.4. All-Gene Dataset

The “All-gene dataset” was built by adding sequences from corresponding species in the Mitogenome dataset to the 7-nuclear dataset. The All-gene dataset contains 14 botiids and 18 outgroups and is 22,128 bp in length (Table 1).

2.3.5. Phylogenetic Analyses

Maximum Likelihood (ML) analyses and bootstrap (MLBP) analyses were performed for each of the above datasets using RAxML v.8.0.26 [33–35]. For individual nuclear gene datasets, the sequence alignments were partitioned by codon positions, and the GTR + I + G substitution model was used. A total of 200 independent runs were performed. Based on preliminary results, we also built ML trees for RAG1, RAG2, and IRBP2 to show the placements of gene alleles of some species. The best partition schemes and partition models suggested by PartitionFinder v2.1.1 [36] were employed for the Mitogenome dataset and the two concatenated datasets. A total of 1000 independent runs were performed. The number of nonparametric bootstrap replications was set as 1000 for all MLBP analyses. The 50% majority rule consensus tree and bootstrap values (BP) were then calculated using PAUP *4.0.b10 [37].

2.4. Whole Genome Sequencing and Genome Profiling

Whole genome sequencing (WGS) and genome profiling were performed for two tetraploid species: *Chromobotia macracanthus* and *Yasuhikotakia modesta*. The WGS libraries were prepared using the NEBNext[®] Ultra[™] II FS DNA Library Prep Kit (#E7805S; New England Biolabs, Ipswich, MA, USA). The dual-indexed libraries were then sent to UF-ICBR for quality control and paired-end 150 bp sequencing on an Illumina NovaSeq X Plus sequencer (Illumina, Inc., San Diego, CA, USA). Adaptors and low-quality reads were trimmed using Trim Galore v0.6.4 [31]. For each species, k-mers (length = 21) were counted from the short read data, and a histogram was exported using *Jellyfish* v2.3.0 [38]. The histogram was then loaded into the online server of GenomeScope 2.0 (<http://genomescope.org/genomescope2.0/>; accessed on 1 March 2025; see [23]) to generate the GenomeScope profile plot. The ploidy level was set as 4 for both species. The above analyses were also performed for four other samples (i.e., *Yasuhikotakia modesta*, *Ambastaia sidthimunki*, *Botia almorhae*, and *B. kubotai*) based on the short read data (NCBI BioProject accession #: PRJNA1067307) obtained in [17].

2.5. Timetree Analyses

There is no reliable fossil from Botiidae to be used for the timetree calibration. To build a timetree for the family Botiidae, we must add taxa from other major families and build a timetree for the order Cypriniformes. Fortunately, some previous studies (e.g., [39–41]) have already performed such timetree analyses. Most of these studies (e.g., [39,40]) only used one or two botiid species to represent the family. Hirt et al. (2017), however, used ten botiid species in their timetree analyses for Cypriniformes [41]. Moreover, they collected sequence data from six nuclear genes and 154 species and used eight carefully

selected fossils. In the current study, we decided to use the divergence time estimate of [41] for the most recent common ancestor (MRCA) of Botiidae (~51 Mya—million years ago) to infer the divergence times of all the botiid taxa we analyzed. We first generated 1000 random numbers following a normal distribution (mean = 51.0, sd = 6.08, so that 5th percentile = 41.0 and 95th percentile = 61.0) in R 4.4.2 [42], using the function `rnorm`, and then randomly picked 500 numbers that were larger than 41 but smaller than 61. Next, a total of 500 separate timetree analyses were performed in TreePL [43] using 41.0 Mya as the minimum age and each of the 500 numbers generated above as the maximum age for the MRCA of Botiidae. We pruned all non-Botiidae taxa from the best ML tree inferred from the Mitogenome dataset and used it as the input tree. The trees built based on the nuclear datasets and concatenated datasets were not used because they contain fewer botiid species. TreePL uses a penalized likelihood approach to generate timetrees in a Maximum Likelihood framework. It uses tree topologies as input files. Whether the trees were inferred from mitochondrial gene sequences or nuclear gene sequences does not impact the results. For the analyses, we performed cross-validation analyses using the “cv” and “randomcv” commands and tested the performance of the available optimization routines using the “prime” command. We then used the “thorough” command to ensure the preferred optimization routine converged. The 500 timetrees obtained were merged and imported into TreeAnnotator 2.7.7 [44] to generate the maximum clade credibility (MCC) tree (no burn-in; posterior probability limit, 0.0; median node heights).

2.6. Ancestral Range Reconstruction

Ancestral range reconstruction of the Botiidae was conducted in R 4.4.2 [42] using the package BioGeoBEARS [45]. The timetree built above was used in the analysis. Based on the current distribution range of botiid species, four areas were defined, i.e., A: Mainland Southeast Asia minus Myanmar (Thailand, Laos, Vietnam, etc.); B: Maritime Southeast Asia (Borneo, Sumatra, Java, etc.); C: South Asia (India, Bangladesh, Pakistan, etc.) plus Myanmar; and D: East Asia (mainly China). Because each area is large, the maximum number of ancestral areas allowed at each node was set as two. Distances among neighboring areas were measured in Google Earth Pro version 9.3.6. using the shortest distance between the assumed centers of distribution. The distance matrix was re-scaled by dividing all the distances by the shortest distance in the matrix. Values in the matrix were then rounded to the first decimal digit. This re-scaling and rounding can greatly reduce the impact of measurement uncertainty. A total of six models (DEC, DEC + J, DIVALIKE, DIVALIKE + J, BAYAREALIKE, and BAYAREALIKE + J) implemented in BioGeoBEARS were tested.

3. Results

3.1. Nuclear Gene Copies

We have identified two gene copies (Copies I and II) for the genes EGR2B and EGR3 for most species, while we only found one gene copy for the other three genes (i.e., RAG1, RAG2, and IRBP2). No indels were found in the EGR3 and IRBP2 gene alignments. In the RAG1 alignment, *Sinibotia pulchra* has an insertion at position 633. This insertion interrupts the reading frame and was ignored during phylogenetic analyses. In the RAG2 alignment, *Yasuhikotakia eos* has deletions at positions 956–958. In the EGR2B alignment, *Yasuhikotakia modesta* (Copy I) and *Y. eos* (Copy I) share the insertion from position 190 to position 192. *Yasuhikotakia modesta* (Copy I) has deletions at the following alignment positions: 223–240 and 472–489. *Botia lohachata* (Copy I) has deletions at positions 373–381. *Sinibotia robusta* (Copy I) and *S. pulchra* (Copy I) share deletions at positions 511–513. *Yasuhikotakia modesta* (Copy II) has deletions at positions 502–513.

3.2. Phylogenetic Relationships

In the tree built from the Mitogenome dataset, the subfamilies Leptobotiinae and Botiinae formed two reciprocally monophyletic groups (Figure 1). Within Leptobotiinae, the two genera *Leptobotia* and *Parabotia* are also reciprocally monophyletic. Within Botiinae, besides the monotypic *Chromobotia*, the other five genera are all monophyletic. The sister group formed by *Ambastaia* and *Sinibotia* is sister to the sister group formed by *Syncrossus* and *Yasuhikotakia*. The group formed by these four genera is then sister to the sister group formed by *Chromobotia* and *Botia*. All but four nodes in the Botiidae are robustly supported (BP = 99% or 100%). The generic-level relationships shown by the mitogenome tree also held true for the trees inferred from the 7-nuclear dataset, the All-gene dataset, the EGR2B dataset (Copy I), and the RAG1 dataset (Figures 2–4). *Ambastaia* is not placed as the sister to *Sinibotia* in the RAG2 tree, *Yasuhikotakia* is not sister to *Syncrossus* in the IRBP2 tree, and *Chromobotia* is not sister to *Botia* in the EGR2B tree (Copy II). However, the relevant nodes have low bootstrap values (all <60%; see Figures 2 and 3). In the EGR3 tree (Copy I), the two *Sinibotia* species did not form a monophyletic group. However, the relevant node is not strongly supported (67%; see Figure 2b). In the EGR3 tree (Copy II), *Ambastaia* and *Sinibotia* are not reciprocal sister groups, and this relationship is strongly supported (BP = 96%; see Figure 2b). The ML trees built for RAG1, RAG2, and IRBP2 with gene alleles shown can be found in Figures S1–S3.

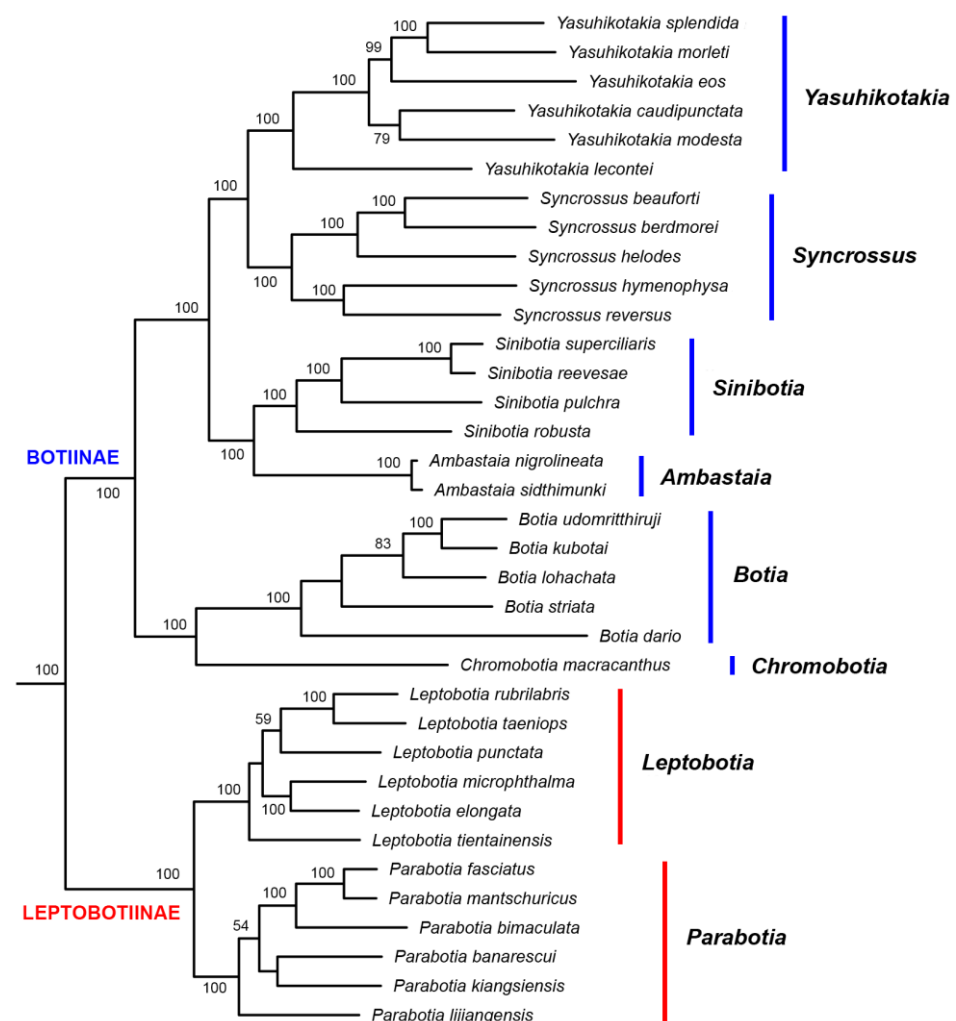


Figure 1. The best Maximum Likelihood tree ($-LnL = 458,271.170016$) built based on the Mitogenome dataset. Only botiid species are shown. Numbers beside nodes are bootstrap support values (BP). Only those values $\geq 50\%$ are shown.

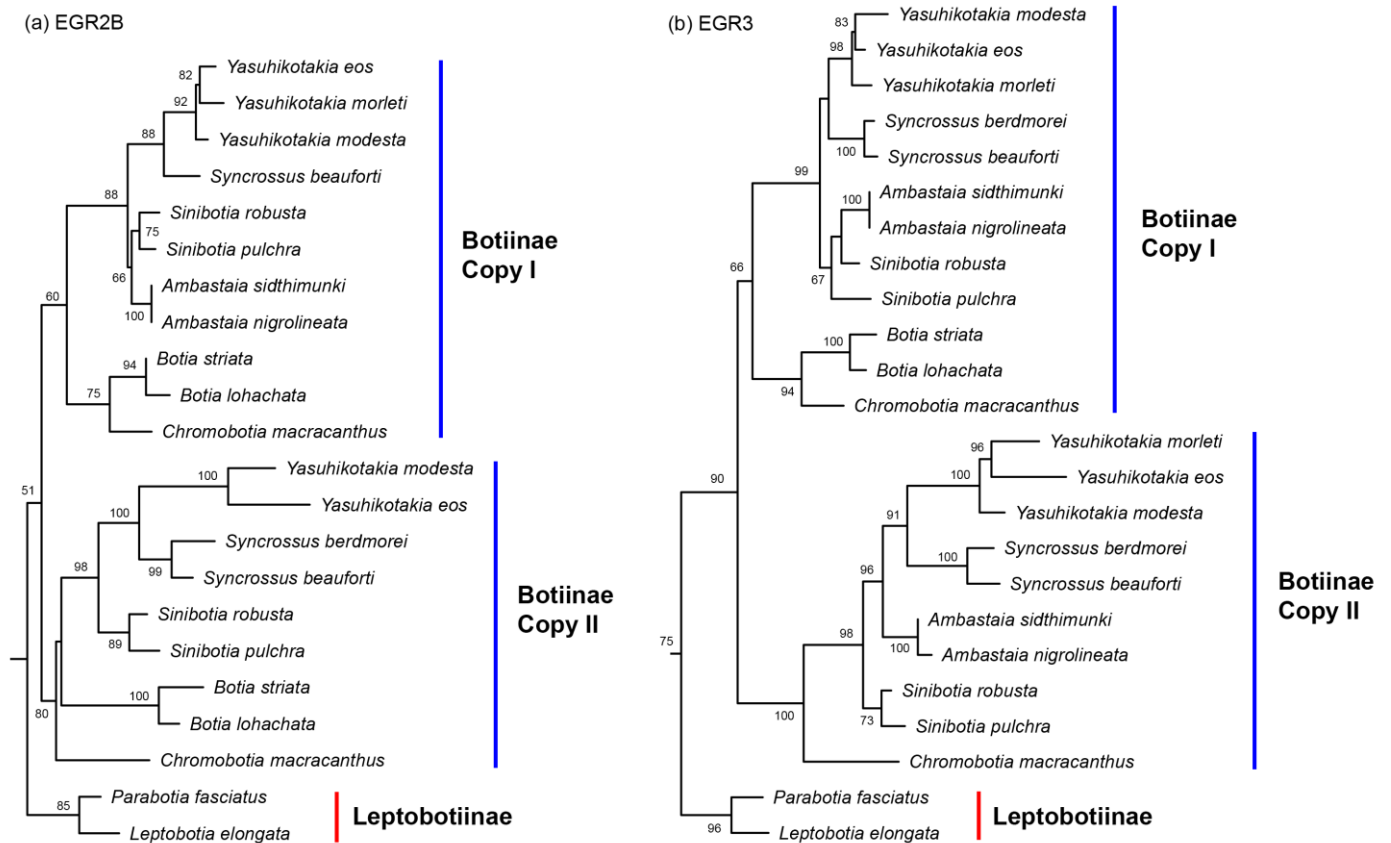


Figure 2. The best Maximum Likelihood trees built based on: (a) EGR2B dataset ($-LnL = 5871.561986$), and (b) EGR3 dataset ($-LnL = 6741.165360$). Only botiid species are shown. Numbers beside nodes are bootstrap support values (BP). Only those values $\geq 50\%$ are shown.

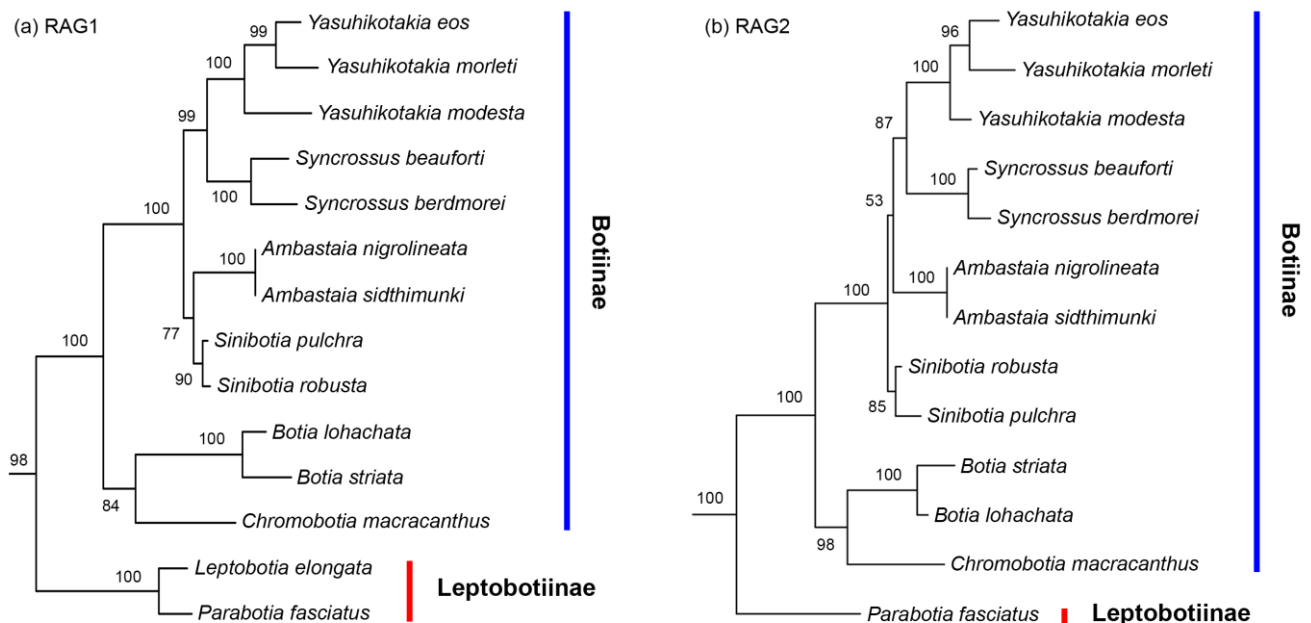


Figure 3. Cont.

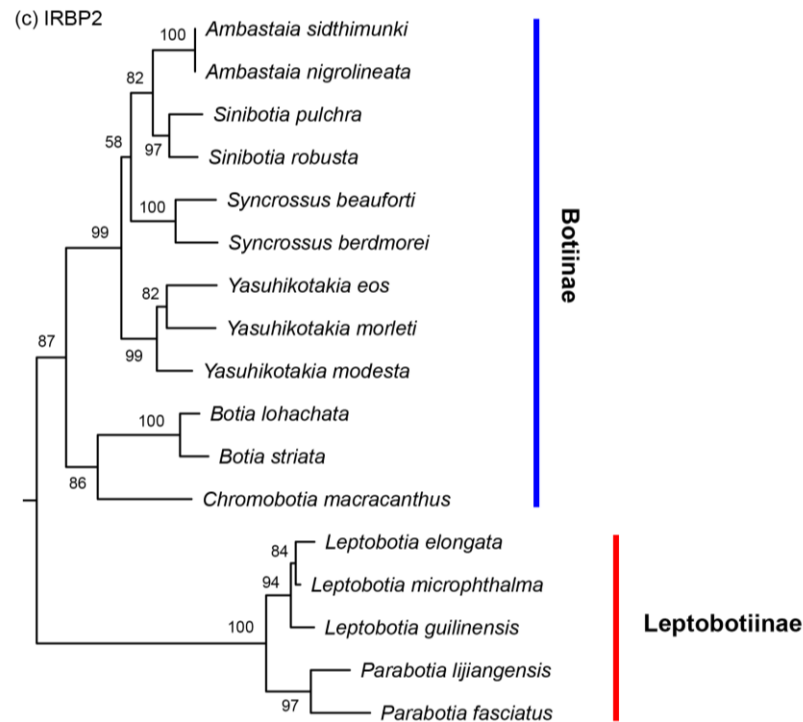


Figure 3. The best Maximum Likelihood trees built based on the (a) RAG1 dataset ($-LnL = 11,118.905157$), (b) RAG2 dataset ($-LnL = 7967.061854$), and (c) IRBP2 dataset ($-LnL = 6904.434141$). Only botiid species are shown. Numbers beside nodes are bootstrap support values (BP). Only values $\geq 50\%$ are shown. See Figures S1–S3 to see the trees with alleles of some species shown.

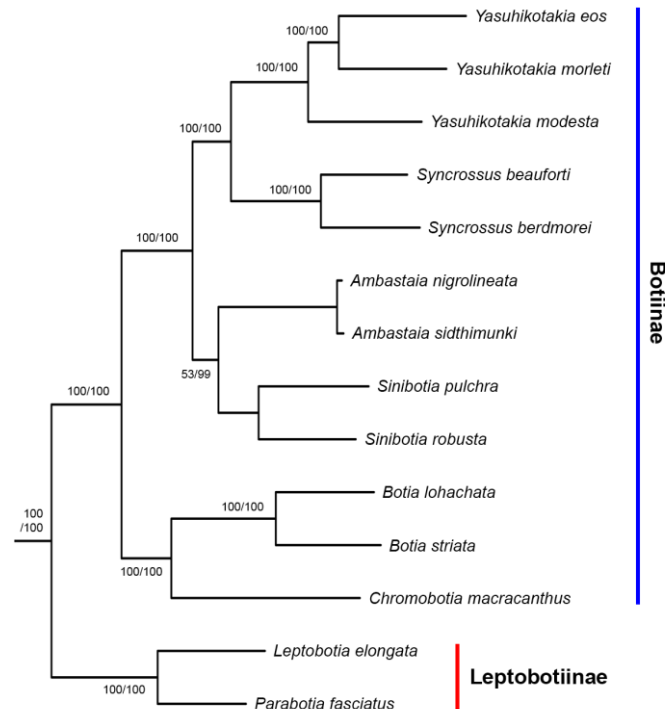


Figure 4. The best Maximum Likelihood trees built based on the 7-nuclear dataset ($-LnL = 47,826.851238$) and the All-gene dataset ($-LnL = 196,328.028457$). Only the latter tree is shown here because the two trees share the same topology. Only botiid species are shown. Numbers beside nodes are bootstrap support values (BP) for the 7-nuclear dataset (before slash) and the All-gene dataset (after slash). Only those values $\geq 50\%$ are shown.

3.3. Whole Genome Sequencing and GenomeScope Profile

A total of 301,552,512 and 295,865,186 sequence reads have been produced by NovaSeq for *Chromobotia macracanthus* and *Yasuhikotakia modesta*, respectively. The GenomeScope profile plots for both species can be found in Figure 5a,b. For *C. macracanthus*, the estimated percentage for the heterozygosity form *aabb* is 12.5%, much higher than that of *aaab* (0.71%). For *Y. modesta*, the percentage of *aabb* (12.2%) is also much higher than that of *aaab* (0.658%). For the four samples from [17], the same pattern has been observed (Figure 5c–f).

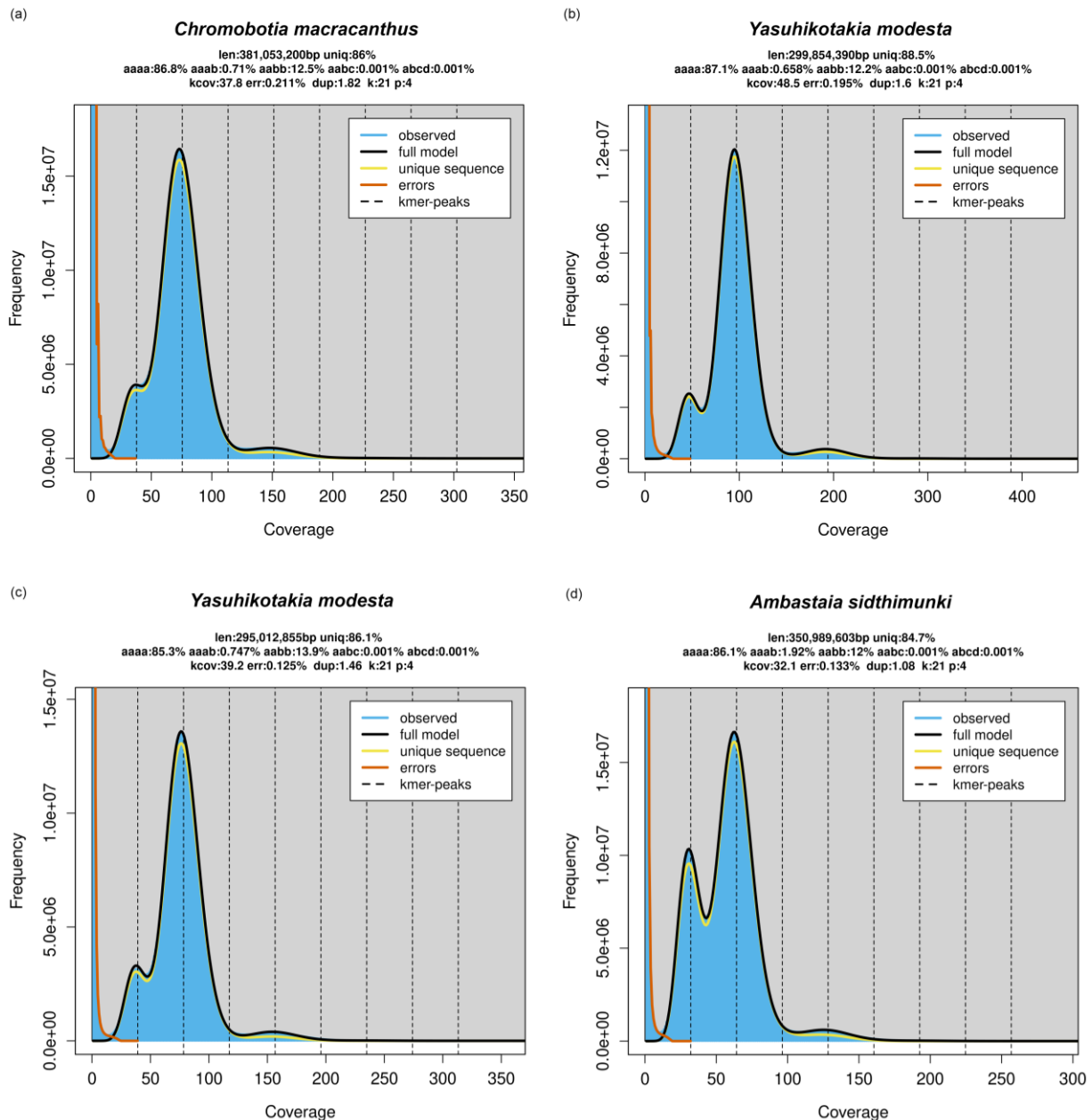


Figure 5. Cont.

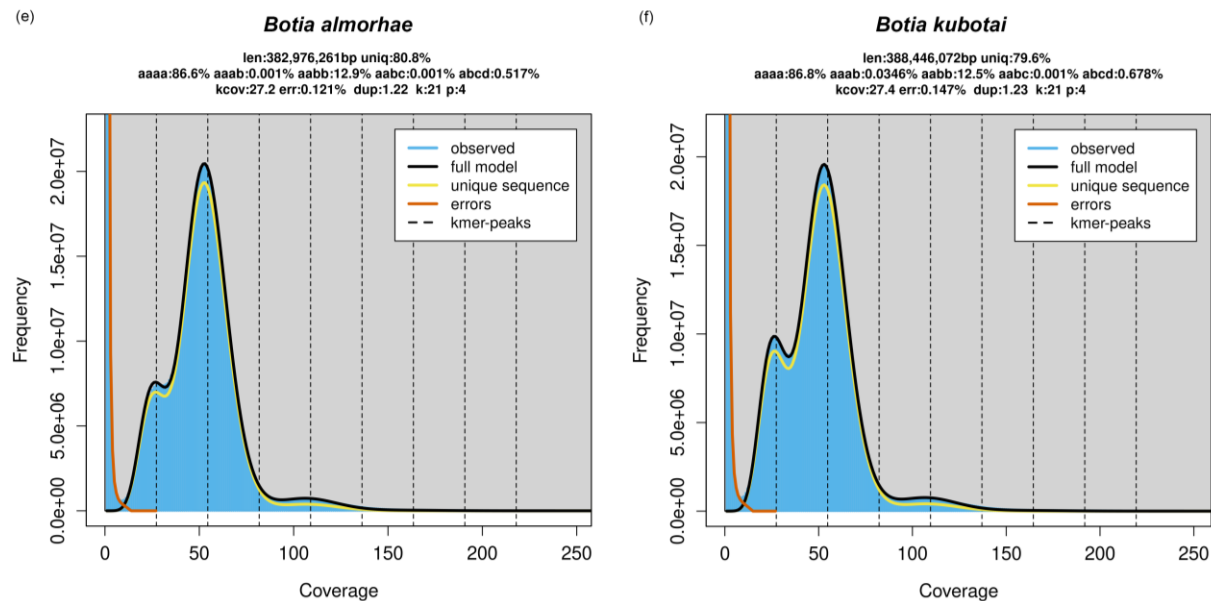


Figure 5. GenomeScope 2.0 profile plots for botiid fishes. (a,b) were based on data (NCBI BioProject accession #: PRJNA1257825) generated in the current study. (c–f) were based on data (NCBI BioProject accession #: PRJNA1067307) obtained by Bitsikas et al. (2024) [17]. For (b), the “average k-mer coverage” was set as 50 based on preliminary results. For other samples, this parameter was set as default (−1).

3.4. Divergence Time Estimation

According to our timetree (Figure S4), the subfamily Botiinae began to diverge at 41.6 Mya, while the first split in the subfamily Leptobotiinae happened at 25.1 Mya. The MRCAs of the tetraploid genera *Botia*, *Sinibotia*, *Syncrossus*, and *Yasuhikotakia* all have similar ages; they are 22.5 Mya, 21.1 Mya, 21.5 Mya, and 21.8 Mya, respectively. The monotypic genus *Chromobotia* originated at 35.7 Mya. The two species of *Ambastatia* did not form until very recently (0.8 Mya). The two diploid genera *Parabotia* and *Leptobotia* began to diverge at 18.2 Mya and 15.6 Mya, respectively. The 95% Highest Posterior Density (HPD) intervals of those ages can be found in Figure S4.

3.5. Ancestral Range Reconstruction Results

Our ancestral range reconstruction results show that the analysis based on the model DIVALIKE + J returned the highest likelihood value ($LnL = -25.97$). Botiidae might have an ancestral distribution in East Asia and Mainland Southeast Asia (Figure 6). After the split of Botiinae and Leptobotiinae, the former stayed in East Asia and later diverged into the current generic and specific diversity, while the latter dispersed to Maritime Southeast Asia, South Asia, and East Asia and diverged into the current generic and specific diversity during the process.

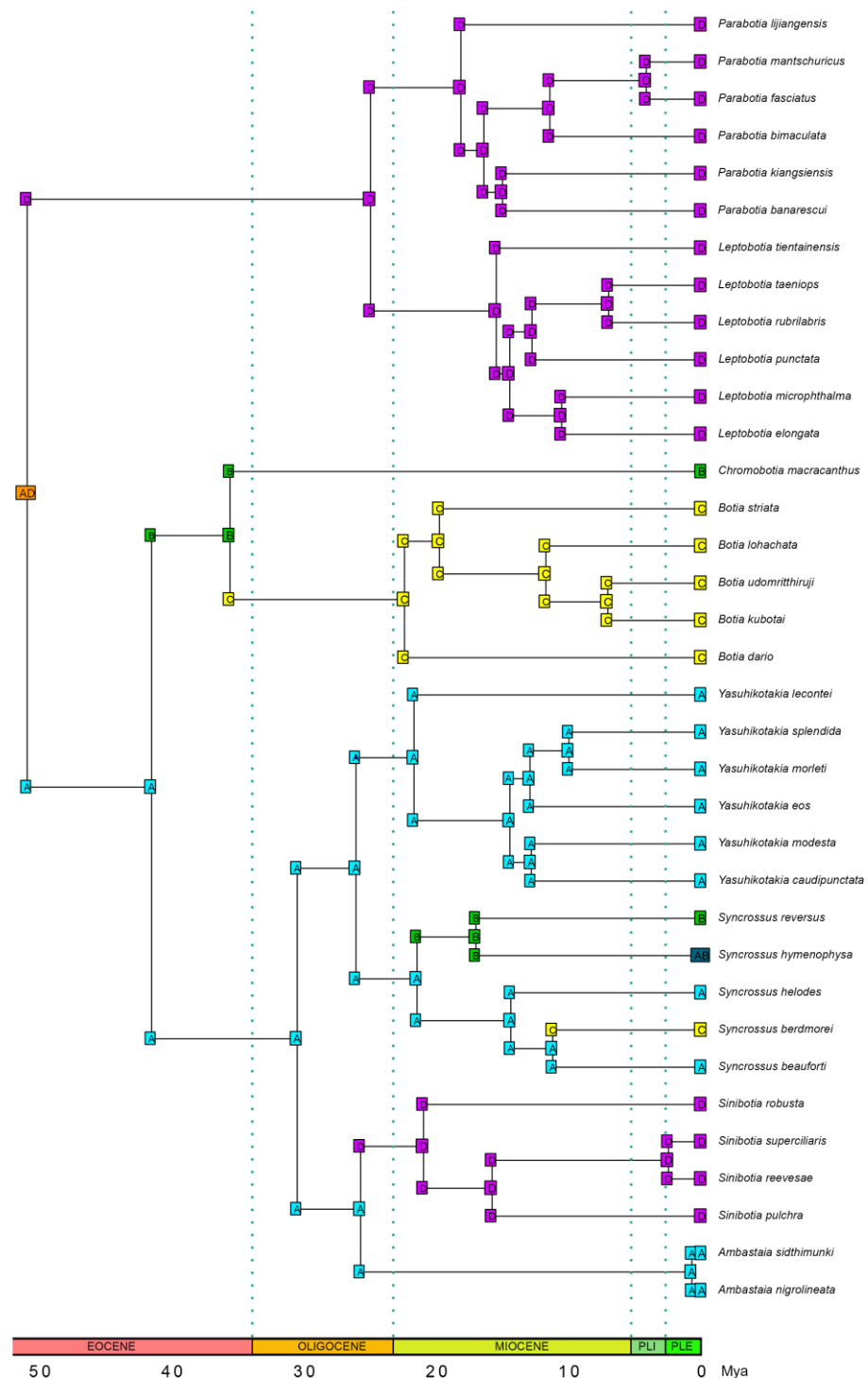


Figure 6. Biogeographical history of the family Botiidae inferred using BioGeoBEARS based on the DIVALIKE + J model. A: Southeast Asia (mainland minus Myanmar); B: Southeast Asia (Maritime); C: South Asia (plus Myanmar); D: East Asia.

4. Discussion

4.1. The Origin of Polyploidy in Botiinae

For the three nuclear genes RAG1, RAG2, and IRBP2, we obtained sequences corresponding to a single gene copy in all botiine species analyzed (Figure 3). This could be due to the loss of one gene copy or the possibility that the PCR primers we used were copy-specific. Another explanation could be recombination between gene copies—i.e.,

homoeologous exchanges [46]—which may homogenize the sequences and obscure the presence of multiple copies. As a result, these three genes are uninformative for determining whether the subfamily Botiinae is of allopolyploid or autopolyploid origin. In contrast, for the nuclear genes EGR2B and EGR3, we successfully recovered two distinct copies for each gene in Botiinae, confirming that species within this subfamily are indeed tetraploids (Figure 2). However, because each gene copy formed a distinct clade and the two clades were sisters to each other in the phylogenetic trees, we are still unable to determine whether botiine species are allopolyploids or autopolyploids. In some known allotetraploid lineages, such as the family Catostomidae, similar tree topologies have been observed for five nuclear genes [28]. In autotetraploid lineages, such as salmonid fishes, some genes that evolved under the so-called AORE (Ancestral Ohnologue Resolution) model can also exhibit similar tree topologies [47]. Under the AORE model, rediploidization occurs in the common polyploid ancestor, and the ohnologues (duplicated gene copies) begin to diverge prior to the subsequent speciation events that give rise to descendant lineages.

Robertson et al. (2017) also proposed another evolutionary model specific to autopolyploids: the LORe (Lineage-specific Ohnologue Resolution) model [47]. Under this model, speciation predates the sequence divergence of ohnologues, leading to lineage-specific divergence patterns. That is because, in some genes, the rediploidization process took too long. When it was completed, speciation events had already taken place [48–50]. Under the LORe model, both gene copy sequences of one or more (but not all) descendant species of a tetraploid group should be clustered together in the gene tree. Multiple such small clusters may appear in the tree. Within each cluster, if there is more than one sample, all Copy I sequences are grouped together, and all Copy II sequences are grouped together; if there is only one sample, then its Copy I and Copy II sequences are sisters to each other. In our current study, the two alleles of the RAG2 sequences of *Sinibotia robusta* and *Ambastaia sidthimunki* differ by 25 bp and 17 bp, respectively. For each species, its two alleles were sisters to each other in the phylogenetic tree (Figure S2). Should the two alleles of these species be treated as two distinct gene copies? We do not know, as there is no clear boundary between alleles and gene copies. If treated as separate gene copies, the RAG2 gene in these species may have evolved under the LORe model, which would be a sign of autopolyploidy. However, we did not identify any larger clades in our trees that would indicate widespread delayed rediploidization. In summary, based on the DNA sequences of the five single-copy nuclear genes analyzed and the resulting gene trees, we are unable to draw a definitive conclusion as to whether the subfamily Botiinae is of allopolyploid or autopolyploid origin. Only five nuclear loci were used in this study. Future research employing hundreds of nuclear loci may provide a clearer resolution of this issue from a phylogenomic perspective.

The GenomeScope profile plots for five botiine species—*Chromobotia macracanthus*, *Yasuhikotakia modesta*, *Ambastaia sidthimunki*, *Botia almorhae*, and *B. kubotai*—representing four of six genera of the subfamily Botiinae, revealed that the estimated percentage of the heterozygosity form *aabb* is consistently much higher than that of *aaab* (Figure 5). According to [23], such a pattern is indicative of an allopolyploid origin. That is because, in allotetraploids, the two subgenomes are from different parental species. During meiosis, homologous chromosomes from the same subgenome tend to pair preferentially, often resulting in the observation of two homologs from each subgenome after recombination. In contrast, autotetraploids possess subgenomes originating from the same parental species. During meiosis, chromosomes from both subgenomes may engage in polysomic inheritance, frequently forming quadrivalents. This pattern reduces the likelihood of recovering exactly two homologs from each subgenome following recombination. A more detailed explanation can be found in [23]. Based on this genomic evidence, it appears that fishes in the subfamily

Botiinae likely had an allotetraploid origin. Nevertheless, we acknowledge that drawing definitive conclusions about the nature of polyploidy can be challenging. Features once thought to be exclusive to autopolyploids—such as polysomic inheritance and multivalent formation—have also been reported in some allopolyploids [20,51].

Because the two gene copies are sisters to each other in both the EGR2B tree and the EGR3 tree (Figure 2), it is not possible to distinguish which copy is maternally inherited and which is paternally inherited. Consequently, we are unable to identify the potential maternal or paternal progenitor of the subfamily Botiinae. It is likely that one of the progenitor lineages is now extinct. Furthermore, the subfamily Leptobotiinae is unlikely to represent either the maternal or the paternal progenitor of Botiinae, as it did not group with either of the Botiinae gene copies in the EGR2B or EGR3 trees.

Allopolyploidy appears to be the predominant mechanism underlying the origin of polyploidy in fishes. To date, only a few fish groups have been confirmed to be autopolyploids, including Salmonidae, Schizothoracinae, Schizopygopsinae, and some species of Acipenseriformes [8,47,52–54]. Besides Botiinae, we already know that all the following polyploid fish groups are allopolyploids: Probarbinae, Cyprininae, Spinibarbininae, tetraploid Torinae, hexaploid Torinae, tetraploid Barbininae, hexaploid Barbininae, and Catostomidae [7,8,28,55]. Some sturgeons (Acipenseriformes) may also be of allopolyploid origin [56]. Our preliminary data also suggest that the tetraploid Smiliogasterinae are allopolyploids as well. In plants, there is still debate on whether allopolyploids are also more common than autopolyploids [20,51,57,58].

The topologies of the EGR2B and EGR3 gene trees clearly indicate that the tetraploidization of the subfamily Botiinae occurred only once, during the formation of the common ancestor of this subfamily (Figure 2). Šlechtová et al. (2006) expressed the same opinion after mapping ploidy levels on their mitochondrial phylogeny, without considering that the same pattern could also be produced if multiple independent tetraploidization events occurred within Botiinae [15]. Sember et al. (2018) reached the same conclusion as ours by analyzing the dynamics of some tandemly repeated DNA sequences and detected a high degree of rediploidization [16]. Based on our timetree, this tetraploidization event might have happened after the Botiinae originated 51.0 Mya and before it began to diversify 41.6 Mya, corresponding to the early to middle Eocene (Figure S4). Therefore, the polyploidization of Botiinae appears to predate most, if not all, polyploidization events in Cyprinidae (<44 Mya; see [8]) but is later than the polyploidization of Catostomidae (>63 Mya; see [59]). Our biogeographical reconstruction suggests that this polyploidization event likely took place in Mainland Southeast Asia, which is home to several large river systems (e.g., Mekong River and Chao Phraya River) (Figure 6). Interestingly, the polyploidization of the small allotetraploid cyprinid subfamily Probarbinae might have also occurred in Mainland Southeast Asia, as all its extant species are restricted to this region [9].

4.2. Phylogenetic Relationships

Previous phylogenetic studies on Botiidae that used mitochondrial genes (e.g., [15,16,24,25]) usually only used one or a few genes (e.g., Cytochrome *b*, 12S rRNA gene). In the current study, however, the whole mitochondrial genome sequences were employed. All generic-level relationships were resolved with 100% bootstrap support, and most other ingroup nodes were also robustly supported (Figure 1). This is not the case for any previous phylogenetic studies on Botiidae. Moreover, we also tried to separate the gene copies for five nuclear genes and built phylogenetic trees based on their sequences. This has not been achieved by any previous studies that used nuclear genes (e.g., [16,25]), which makes the results from these studies potentially plagued with paralog issues.

All phylogenetic trees built in this study (Figures 1–3), except for the EGR3 tree (Copy I), strongly supported or at least did not reject the following relationships among Botiidae. Leptobotiinae/Botiinae, *Leptobotia*/*Parabotia*, *Chromobotia*/*Botia*, *Yasuhikotakia*/*Syncrossus*, and *Sinibotia*/*Ambastaia* are all sister taxa. Clades formed by the last two generic pairs are sisters to each other. The clade formed by these four genera is sister to the clade formed by *Chromobotia*/*Botia*. The above relationships have also been shown by the RAG1 tree of Bohlen et al. (2016) [25]. In the EGR3 tree (Copy I), *Ambastaia* and *Sinibotia* did not form a sister group (Figure 3). That might be the result of incomplete lineage sorting.

All our phylogenetic trees indicated that the two species of *Ambastaia*, *A. sidthimunki* and *A. sidthimunki*, are very closely related to each other. Especially in the trees built based on the RAG1 dataset, the RAG2 dataset, and the IRBP2 dataset, the two alleles of both species were mixed with each other (Figures S1–S3), which is a sign that they have diverged from their common ancestor relatively recently. This is supported by our timetree, which shows that the two species diverged only 0.8 million years ago (Figure S4).

Due to resource limitations, this study only included 35 species in the mitogenome dataset and 14 species in the nuclear datasets, representing 53.8% and 21.5% of all currently recognized botiid species, respectively. Furthermore, only five nuclear genes were analyzed. Future studies incorporating broader taxon sampling and more extensive genomic data will be essential for a more comprehensive understanding of phylogenetic relationships within the family Botiidae.

4.3. Biogeographical History

According to our ancestral range reconstruction results (Figure 6), the family Botiidae likely originated in East Asia and Mainland Southeast Asia around 51 Mya. This time frame coincides with the initial collision of the Indian plate with the Eurasian plate [60]. The ancestors of Indian botiid fishes likely dispersed from Southeast Asia in two distinct waves. The first wave may have occurred during the Eocene, introducing the common ancestor of *Botia* to the Indian subcontinent. The second wave likely took place in the middle Miocene, bringing *Syncrossus* into the region. Most Chinese botiid species are diploids belonging to the subfamily Leptobotiinae, which originated during the initial divergence of Botiidae. In contrast, the common ancestor of the tetraploid genus *Sinibotia* likely dispersed from Mainland Southeast Asia during the Early Miocene. The presence of *Chromobotia* and *Syncrossus* on the large islands of Southeast Asia likely resulted from two separate dispersal events from the mainland. The first wave might have occurred in the late Eocene to early Oligocene, while the second wave likely took place during the Miocene.

The DIVALIKE + J model inferred a long-distance dispersal event from Maritime Southeast Asia (B) to South Asia and Myanmar (C) along the branch leading to *Chromobotia* and *Botia* (Figure 6). This inference is based on the sister relationship between *Chromobotia*—which is currently distributed in Sumatra and Borneo—and *Botia*, which occurs in South Asia and Myanmar. This dispersal likely took place during the late Eocene (Figure 6). Although the model did not explicitly reconstruct the route, we propose that the dispersal may have occurred via the nearby Malay Peninsula, which could have served as a stepping stone. We agree with Šlechtová et al. (2006) that *Chromobotia* and *Botia* might once have had a broader and continuous distribution [17].

The family Botiidae began to diverge around 51.0 Mya. Sometime after this, the ancestor of the subfamily Botiinae became tetraploid through hybridization followed by whole genome duplication. Subsequently, around 41.6 Mya, the subfamily Botiinae began to diverge. These major events occurred during the early to middle Eocene. This period was part of a global greenhouse phase, characterized by elevated atmospheric CO₂ levels and high global temperatures [61]. At the same time, the region was undergoing significant

tectonic shifts, driven by the collision of the Indo-Australian plate with the Eurasian plate [60,62]. The tectonic processes also led to the development of complex river systems and the establishment of new ecological niches, promoting the diversification of freshwater fish species. The reorganization of river systems might have created opportunities for the maternal progenitor and the paternal progenitor of the Botiinae to meet and hybridize with each other and eventually form the tetraploid ancestor of the subfamily.

The subfamily Leptobotiinae began to diverge around 25.1 Mya, during the late Oligocene. East Asia underwent significant geological and climatic transformations during this period, which might have contributed to the diversification of the subfamily. Globally, this period was marked by a cooler and drier climate [63]. In East Asia, the ongoing collision between the Indian and Eurasian plates resulted in the progressive uplift of the Tibetan Plateau, reshaping regional topography and altering drainage patterns [64,65]. These tectonic processes played an important role in the early development of the East Asian monsoon system and the formation of new river systems and freshwater habitats [66], which later might have contributed to the further diversification of the subfamily Leptobotiinae.

5. Conclusions

Previous studies have remained inconclusive about whether the tetraploid subfamily Botiinae originated through autopolyploidy or allopolyploidy. In the present study, we employed both a phylogeny-based genetic approach and a k-mer-based genomic method to address this question. Our results indicate that Botiinae likely have an allopolyploid origin, with the tetraploidization event occurring only once in the common ancestor of the subfamily. Phylogenetic trees inferred from mitochondrial genome sequences and five phased nuclear gene sequences were largely congruent, supporting the following sister group relationships: Leptobotiinae/Botiinae, *Leptobotia*/*Parabotia*, *Chromobotia*/*Botia*, *Yasuhikotakia*/*Syncrossus*, and *Sinibotia*/*Ambastaia*. We also estimated divergence times and reconstructed the ancestral range distribution of botiid fishes. Our analyses suggest that the family Botiidae likely originated in East Asia and Mainland Southeast Asia approximately 51 million years ago, followed by its dispersal into South Asia and the islands of Southeast Asia, ultimately giving rise to the group's current generic and species diversity. Future studies will need to expand taxon sampling and incorporate data from more nuclear loci. This is critical for improving our understanding of the phylogenetic relationships, polyploid evolution, and biogeography of botiid fishes.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/biology14050531/s1>: Figure S1: Maximum Likelihood tree ($-LnL = 11,796.194434$) built based on the nuclear RAG1 dataset; Figure S2: Maximum Likelihood tree ($-LnL = 8541.995023$) built based on the nuclear RAG2 dataset; Figure S3: Maximum Likelihood tree ($-LnL = 6922.093239$) built based on the nuclear IRBP2 dataset; Figure S4: Divergence time estimations for Botiidae; Table S1: Sample information for the mitogenome dataset; Table S2: Sample information for the nuclear dataset.

Author Contributions: Conceptualization, L.Y.; Methodology, L.Y.; Software, L.Y.; Validation, L.Y.; Formal Analysis, L.Y.; Investigation, L.Y.; Resources, R.L.M. and G.J.P.N.; Data Curation, L.Y.; Writing—Original Draft Preparation, L.Y.; Writing—Review and Editing, L.Y., R.L.M. and G.J.P.N.; Visualization, L.Y.; Supervision, R.L.M. and G.J.P.N.; Project Administration, R.L.M. and G.J.P.N.; Funding Acquisition, R.L.M. and G.J.P.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Foundation (EF 0431326, DEB 0956370, and DEB 1021840 to R.L.M. and DEB 1541556 to G.J.P.N.) and the Florida Museum of Natural History.

Institutional Review Board Statement: Ethical review and approval were waived for this study because the animal samples involved were preserved tissue samples sent by collaborators who followed local legislation and guidelines. No live animals were involved in any experiments conducted by the authors.

Informed Consent Statement: Not applicable.

Data Availability Statement: All mitochondrial and nuclear gene sequence data used in this study can be found in GenBank (accession numbers are provided in Tables S1 and S2). The short Illumina read data used in this study can be found in NCBI's Sequence Read Archive (SRA) with the BioProject accession #: PRJNA1257825 and PRJNA1067307.

Acknowledgments: We thank Terry Lott and Larry Page for their help with tissue samples. We are very grateful to Shengchen Shan for the helpful suggestions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rice, A.; Šmarda, P.; Novosolov, M.; Drori, M.; Glick, L.; Sabath, N.; Meiri, S.; Belmaker, J.; Mayrose, I. The global biogeography of polyploid plants. *Nat. Ecol. Evol.* **2019**, *3*, 265–273. [CrossRef] [PubMed]
2. Van de Peer, Y.; Ashman, T.; Soltis, P.S.; Soltis, D.E. Polyploidy: An evolutionary and ecological force in stressful times. *Plant Cell* **2021**, *33*, 11–26. [CrossRef] [PubMed]
3. Rothfels, C.J. Polyploid phylogenetics. *New Phytol.* **2021**, *230*, 66–72. [CrossRef]
4. Morris, J.P.; Baslan, T.; Soltis, D.E.; Soltis, P.S.; Fox, D.T. Integrating the study of polyploidy across organisms, tissues, and disease. *Annu. Rev. Genet.* **2024**, *58*, 297–318. [CrossRef] [PubMed]
5. Otto, S.P.; Whitton, J. Polyploid incidence and evolution. *Annu. Rev. Genet.* **2000**, *34*, 401–437. [CrossRef]
6. Mable, B.K.; Alexandrou, M.A.; Taylor, M.I. Genome duplication in amphibians and fish: An extended synthesis. *J. Zool.* **2011**, *284*, 151–282. [CrossRef]
7. Yang, L.; Sado, T.; Hirt, M.; Pasco-Viel, E.; Arunachalam, M.; Li, J.; Wang, X.; Freyhof, J.; Saitoh, K.; Simons, A.M.; et al. Phylogeny and polyploidy: Resolving the classification of cyprinine fishes (Teleostei: Cypriniformes). *Mol. Phylogenet. Evol.* **2015**, *85*, 97–116. [CrossRef]
8. Yang, L.; Naylor, G.J.P.; Mayden, R.L. Deciphering reticulate evolution of the largest group of polyploid vertebrates, the subfamily Cyprininae (Teleostei: Cypriniformes). *Mol. Phylogenet. Evol.* **2022**, *166*, 107323. [CrossRef]
9. Fricke, R.; Eschmeyer, W.N.; van der Laan, R. (Eds.) *Eschmeyer's Catalog of Fishes: Genera, Species, References, Electronic Version*; California Academy of Sciences: San Francisco, CA, USA, 2025; Available online: <https://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp> (accessed on 15 February 2025).
10. Ojima, Y.; Yamamoto, K. Cellular DNA contents of fishes determined by flow cytometry. *La Kromosomo II* **1990**, *57*, 1871–1888.
11. Suzuki, A.; Taki, Y. Tetraploidization in the cobitid subfamily Botinae (Pisces, Cypriniformes). *Cytobios* **1996**, *85*, 229–245.
12. Donsakul, T.; Magtoon, W. Karyotypes of Five Cobitid Fishes (Family Cobitidae): *Botia lecontei*, *B. berdmorei*, *B. kubotai*, *Yasuhikotakia sidhimunki* and *Y. eos* in Thailand. In Proceedings of the 48th Kasetsart University Annual Conference: Fisheries, Bangkok, Thailand, 3–5 March 2010; pp. 235–242.
13. Arai, R. (Ed.) *Fish Karyotypes: A Check List*; Springer: Tokyo, Japan, 2011.
14. Kaewmad, P.; Monthatong, M.; Supiwong, W.; Saowakoon, S.; Tanomtong, A. Natural autotetraploid and chromosomal characteristics in the subfamily Botiinae (Cypriniformes, Cobitinae) from Northeast Thailand. *Cytologia* **2014**, *79*, 299–313. [CrossRef]
15. Šlechtová, V.; Bohlen, J.; Freyhof, J.; Ráb, 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*, 529–541. [CrossRef] [PubMed]
16. Sember, A.; Bohlen, J.; Šlechtová, V.; Altmanová, M.; Pelikánová, Š.; Ráb, P. Dynamics of tandemly repeated DNA sequences during evolution of diploid and tetraploid botiid loaches (Teleostei: Cobitoidea: Botiidae). *PLoS ONE* **2018**, *13*, e0195054. [CrossRef] [PubMed]
17. Bitsikas, V.; Cubizolles, F.; Schier, A.F. A vertebrate family without a functional Hypocretin/Orexin arousal system. *Curr. Biol.* **2024**, *34*, 1532–1540. [CrossRef]
18. Lv, Y.; Li, Y.; Huang, Y.; Wang, J.; Tian, Z.; He, Y.; Shi, J.; Huang, Z.; Wen, Z.; Shi, Q.; et al. Deciphering genome-wide molecular pathways for exogenous *Aeromonas hydrophila* infection in wide-bodied sand loach (*Sinibotia reevesae*). *Aquac. Rep.* **2024**, *35*, 102033. [CrossRef]
19. Stebbins, G.L. *Variation and Evolution in Plants*; Oxford University Press: London, UK, 1950.

20. Lv, Z.; Nyarko, C.A.; Ramtekey, V.; Behn, H.; Mason, A.S. Defining autopolyploidy: Cytology, genetics, and taxonomy. *Am. J. Bot.* **2024**, *111*, e16292. [CrossRef]
21. Evans, B.J.; Kelley, D.B.; Melnick, D.J.; Cannatella, D.C. Evolution of RAG-1 in polyploid clawed frogs. *Mol. Biol. Evol.* **2005**, *22*, 1193–1207. [CrossRef]
22. Gu, H.; Wang, S.; Yang, C.; Tao, M.; Wang, Z.; Liu, S. Global cooling and hot springs may have induced autotetraploidy and autohexaploidy in *Schizothorax* ancestors, and its implications for polyploid breeding. *Aquaculture* **2024**, *584*, 740659. [CrossRef]
23. Ranallo-Benavidez, T.R.; Jaron, K.S.; Schatz, M.C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **2020**, *11*, 1432. [CrossRef]
24. Tang, Q.; Yu, D.; Liu, H. *Leptobotia zebra* should be revised as *Sinibotia zebra* (Cypriniformes: Botiidae). *Zool. Res.* **2008**, *29*, 1–9. [CrossRef]
25. Bohlen, J.; Šlechtová, V.; Šlechtá, V.; Šlechtová, V.; Sember, A.; Ráb, P. A ploidy difference represents an impassable barrier for hybridisation in animals. Is there an exception among botiid loaches (Teleostei: Botiidae)? *PLoS ONE* **2016**, *11*, e0159311. [CrossRef]
26. Li, C.; Hofreiter, M.; Straube, N.; Corrigan, S.; Naylor, G.J. Capturing protein-coding genes across highly divergent species. *BioTechniques* **2013**, *54*, 321–326. [CrossRef] [PubMed]
27. White, W.T.; Corrigan, S.; Yang, L.; Henderson, A.C.; Bazinet, A.L.; Swofford, D.L.; Naylor, G.J.P. Phylogeny of the manta and devilrays (Chondrichthyes: Mobulidae), with an updated taxonomic arrangement for the family. *Zool. J. Linn. Soc.* **2018**, *182*, 50–75. [CrossRef]
28. Yang, L.; Mayden, R.L.; Naylor, G.J.P. Phylogeny and polyploidy evolution of the suckers (Teleostei: Catostomidae). *Biology* **2024**, *13*, 1072. [CrossRef] [PubMed]
29. Chen, W.-J.; Miya, M.; Saitoh, K.; Mayden, R.L. Phylogenetic utility of two existing and four novel nuclear gene loci in reconstructing tree of life of ray-finned fishes: The order Cypriniformes (Ostariophysi) as a case study. *Gene* **2008**, *423*, 125–134. [CrossRef]
30. Lovejoy, N.R.; Collette, B.B. Phylogenetic relationships of New World needlefishes (Teleostei: Belontiidae) and the biogeography of transitions between marine and freshwater habitats. *Copeia* **2001**, *2001*, 324–338. [CrossRef]
31. Krueger, F. Trim Galore! 2012. Available online: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (accessed on 1 March 2024).
32. Saitoh, K.; Sado, T.; Mayden, R.L.; Hanzawa, N.; Nakamura, K.; Nishida, M.; Miya, M. Mitogenomic evolution and interrelationships of the Cypriniformes (Actinopterygii: Ostariophysi): The first evidence toward resolution of higher-level relationships of the World's largest freshwater fish clade based on 59 whole mitogenome sequences. *J. Mol. Evol.* **2006**, *63*, 826–841. [CrossRef]
33. Stamatakis, A. RAxML-VI-HP: Maximum Likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006**, *22*, 2688–2690. [CrossRef]
34. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and postanalysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef]
35. Stamatakis, A.; Hoover, P.; Rougemont, J. A rapid bootstrap algorithm for the RAxML web-servers. *Syst. Biol.* **2008**, *57*, 758–771. [CrossRef]
36. Lanfear, R.; Frandsen, P.B.; Wright, A.M.; Senfeld, T.; Calcott, B. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **2017**, *34*, 772–773. [CrossRef] [PubMed]
37. Swofford, D.L. PAUP*: *Phylogenetic Analysis Using Parsimony (* and Other Methods)*, version 4.0b10; Sinauer Associates: Sunderland, MA, USA, 2022.
38. Marçais, G.; Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **2011**, *27*, 764–770. [CrossRef]
39. Saitoh, K.; Sado, T.; Doosey, M.H.; Bart, H.L., Jr.; Inoue, G.; Nishida, M.; Mayden, R.L.; Miya, M. Evidence from mitochondrial genomics supports the lower Mesozoic of South Asia as the time and place of basal divergence of cypriniform fishes (Actinopterygii: Ostariophysi). *Zool. J. Linn. Soc.* **2011**, *161*, 633–662. [CrossRef]
40. Chen, W.-J.; Lavoué, S.; Mayden, R.L. Evolutionary origin and early biogeography of otophysan fishes (Ostariophysi: Teleostei). *Evolution* **2013**, *67*, 2218–2239. [CrossRef]
41. Hirt, M.V.; Arratia, G.; Chen, W.J.; Mayden, R.L.; Tang, K.L.; Wood, R.M.; Simons, A.M. Effects of gene choice, base composition and rate heterogeneity on inference and estimates of divergence times in cypriniform fishes. *Biol. J. Linn. Soc.* **2017**, *121*, 319–339. [CrossRef]
42. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2024; Available online: <https://www.R-project.org/> (accessed on 8 May 2025).
43. Smith, S.A.; O'Meara, B.C. treePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **2012**, *28*, 2689–2690. [CrossRef]

44. Bouckaert, R.; Heled, J.; Kühnert, D.; Vaughan, T.; Wu, C.-H.; Xie, D.; Suchard, M.A.; Rambaut, A.; Drummond, A.J. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* **2014**, *10*, e1003537. [\[CrossRef\]](#)
45. Matzke, N.J. Probabilistic historical biogeography: New models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Front. Biogeogr.* **2013**, *5*, 242–248. [\[CrossRef\]](#)
46. Mason, A.S.; Wendel, J.F. Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution. *Front. Genet.* **2020**, *11*, 1014. [\[CrossRef\]](#)
47. Robertson, F.M.; Gundappa, M.K.; Grammes, F.; Hvidsten, T.R.; Redmond, A.K.; Lien, S.; Martin, S.A.M.; Holland, P.W.H.; Sandve, S.R.; Macqueen, D.J. Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biol.* **2017**, *18*, 111. [\[CrossRef\]](#)
48. Martin, K.J.; Holland, P.W. Enigmatic orthology relationships between Hox clusters of the African butterfly fish and other teleosts following ancient whole-genome duplication. *Mol. Biol. Evol.* **2014**, *31*, 2592–2611. [\[CrossRef\]](#)
49. Macqueen, D.J.; Johnston, I.A. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc. Biol. Sci.* **2014**, *281*, 20132881. [\[CrossRef\]](#)
50. Lien, S.; Koop, B.F.; Sandve, S.R.; Miller, J.R.; Kent, M.P.; Nome, T.; Hvidsten, T.R.; Leong, J.S.; Minkley, D.R.; Zimin, A.; et al. The Atlantic salmon genome provides insights into rediploidization. *Nature* **2016**, *533*, 200–205. [\[CrossRef\]](#)
51. Doyle, J.J.; Sherman-Broyles, S. Double trouble: Taxonomy and definitions of polyploidy. *New Phytol.* **2017**, *213*, 487–493. [\[CrossRef\]](#)
52. Du, K.; Stöck, M.; Kneitz, S.; Klopp, C.; Woltering, J.M.; Adolphi, M.C.; Feron, R.; Prokopov, D.; Makunin, A.; Kichigin, I.; et al. The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. *Nat. Ecol. Evol.* **2020**, *4*, 841–852. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Li, X.; Wang, M.; Zoo, M.; Guan, X.; Xu, S.; Chen, W.; Wang, C.; Chen, Y.; He, S.; Guo, B. Recent and recurrent autopolyploidization fueled diversification of snow carp on the Tibetan Plateau. *Mol. Biol. Evol.* **2024**, *41*, msae221. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Wang, B.; Wu, B.; Liu, X.; Hu, Y.; Ming, Y.; Bai, M.; Liu, J.; Xiao, K.; Zeng, Q.; Yang, J.; et al. Whole-genome sequencing reveals autooctoploidy in Chinese sturgeon and its evolutionary trajectories. *Genom. Proteom. Bioinform.* **2024**, *22*, qzad002. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Krabbenhoft, T.J.; MacGuigan, D.J.; Backenstose, N.J.; Waterman, H.; Lan, T.; Pelosi, J.A.; Tan, M.; Sandve, S.R. Chromosome-level genome assembly of Chinese sucker (*Myxocyprinus asiaticus*) reveals strongly conserved synteny following a catostomid-specific whole-genome duplication. *Genome Biol. Evol.* **2021**, *13*, evab190. [\[CrossRef\]](#)
56. Fontana, F.; Congiu, L.; Mudrak, V.A.; Quattro, J.M.; Smith, T.I.J.; Ware, K.; Doroshov, S.I. Evidence of hexaploid karyotype in shortnose sturgeon. *Genome* **2008**, *51*, 113–119. [\[CrossRef\]](#)
57. Barker, M.S.; Arrigo, N.; Baniaga, A.E.; Li, Z.; Levin, D.A. On the relative abundance of autopolyploids and allopolyploids. *New Phytol.* **2016**, *210*, 391–398. [\[CrossRef\]](#)
58. Spoelhof, J.P.; Soltis, P.S.; Soltis, D.E. Pure polyploidy: Closing the gaps in autopolyploid research. *J. Syst. Evol.* **2017**, *55*, 340–352. [\[CrossRef\]](#)
59. Bagley, J.C.; Mayden, R.L.; Harris, P.M. Phylogeny and divergence times of suckers (Cypriniformes: Catostomidae) inferred from Bayesian total-evidence analyses of molecules, morphology, and fossils. *PeerJ* **2018**, *6*, e5168. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Najman, Y.; Appel, E.; Boudagher-Fadel, M.; Bown, P.; Carter, A.; Garzanti, E.; Godin, L.; Han, J.; Liebke, U.; Oliver, G.; et al. Timing of India–Asia collision: Geological, biostratigraphic, and palaeomagnetic constraints. *J. Geophys. Res.* **2010**, *115*, B12416. [\[CrossRef\]](#)
61. Zachos, J.C.; Dickens, G.R.; Zeebe, R.E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **2008**, *451*, 279–283. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Hall, R. Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: Computer-based reconstructions, model and animations. *J. Asian Earth Sci.* **2002**, *20*, 353–434. [\[CrossRef\]](#)
63. Zachos, J.; Pagani, M.; Sloan, L.; Thomas, E.; Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **2001**, *292*, 686–693. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Tapponnier, P.; Xu, Z.; Roger, F.; Meyer, B.; Arnaud, N.; Wittlinger, G.; Yang, J. Oblique stepwise rise and growth of the Tibet Plateau. *Science* **2001**, *294*, 1671–1677. [\[CrossRef\]](#)
65. Harrison, T.M.; Copeland, P.; Kidd, W.S.F.; Yin, A. Raising Tibet. *Science* **1992**, *255*, 1663–1670. [\[CrossRef\]](#)
66. Clift, P.D.; Hodges, K.V.; Heslop, D.; Hannigan, R.; Long, H.; Calves, G. Correlated evolution of Himalayan and Tibetan uplift and Asian monsoon intensity. *Nat. Geosci.* **2008**, *1*, 875–880. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.