PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Zhong J, Yi S, Ma L, Wang W. 2019 Evolution and phylogeography analysis of diploid and polyploid *Misgurnus anguillicaudatus* populations across China. *Proc. R. Soc. B* **286**: 20190076.
http://dx.doi.org/10.1098/rspb.2019.0076

Received: 10 January 2019
Accepted: 3 April 2019

Subject Category:
Evolution

Subject Areas:
evolution, genetics, molecular biology

Keywords:
polyploidy, phylogeography, population genetics, geographical distribution, mtDNA, *M. anguillicaudatus*

Author for correspondence:
Weimin Wang
e-mail: wangwm@mail.hzau.edu.cn

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4466063.

THE ROYAL SOCIETY PUBLISHING

Evolution and phylogeography analysis of diploid and polyploid *Misgurnus anguillicaudatus* populations across China

Jia Zhong, Shaokui Yi, Laiyan Ma and Weimin Wang

College of Fisheries, Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education/Key Laboratory of Freshwater Animal Breeding, Ministry of Agriculture, Huazhong Agricultural University, Wuhan 430070, People's Republic of China

JZ, 0000-0002-5084-0388; SY, 0000-0003-2325-0179; LM, 0000-0002-2744-9647; WW, 0000-0001-7705-8321

The origin and evolution of polyploid organisms have been extensively studied in plants, but this topic remains only partially understood in vertebrates, where polyploidy is relatively rare. In this study, we used *Misgurnus anguillicaudatus*, a fish that comprises five ploidy levels in nature, as a model animal to improve our understanding of biogeographic history and evolution of polyploid vertebrates. After collecting samples from different geographical populations in China, their ploidy levels were determined using flow cytometry. Two mitochondrial markers (*cytochrome b* and control region) were then used for phylogeographic analyses to unravel the possible origins of diploids and tetraploids in China. The results showed that diploids have wider geographical distribution than tetraploids and triploids. There was no clear allopatric geographical range or boundary to divide diploid and polyploid populations. Rather, the analysis of mitochondrial DNA sequences indicated that tetraploids were autopolyploids, with lower genetic diversity than diploids. This suggests that tetraploids originated from sympatric diploids via multiple independent polyploidization events. Genetic structure patterns were similar between diploids and tetraploids, whereas complex genetic differentiation was found among different regions. The potential origin of *M. anguillicaudatus* was deduced to be in the Pearl River basin, which exhibited the highest nucleotide diversity and genetic differentiation. These findings provide insights into the evolution of polyploidy in vertebrates.

1. Introduction

Polyploidy, the state of having more than two homologous sets of chromosomes, is known to generate genetic variation, drive genomic repatterning, facilitate evolution of novel traits and enhance ecological transformation [1]. As a ubiquitous process, polyploidy is thought to play a key role in plant evolution [2], but compared to the high frequency of polyploidization in plants, polyploidization is rare in animals [3]. However, polyploidization has been documented in some vertebrate taxa, including teleost fishes, amphibians and reptiles [4,5]. Phylogenomic and comparative genomics studies indicate that several duplications have occurred in the evolutionary history of this group of animals, including a fish-specific genome duplication (FSGD) at the base of teleost radiation; so, FSGD may have been a crucial factor driving the species diversity, genomic complexity and astounding ecological diversity in the teleost evolution [6,7]. In different lineages of fishes, throughout Asia, Europe, America and Africa, there are reports of triploids, tetraploids, pentaploids, hexaploids and even octoploids [8,9]. Therefore, polyploid fishes are a good model to study the evolutionary history and trajectory of polyploidization in vertebrates.

Misgurnus anguillicaudatus (Actinopterygii: Cypriniformes: Cobitidae) is a small freshwater teleost with strong environmental adaptability, widely distributed in China, Japan, Korea and other Southeast Asian countries [10]. This species is interesting from the scientific perspective for its natural variability in ploidy

levels. Ojima & Takai [11] were the first to report the natural occurrence of triploid ($3n = 75$) and tetraploid ($4n = 100$) *M. anguillicaudatus*. Since then, researchers found that, along with diploids ($2n = 50$), which are the most common, some wild loach populations in Japan also contain relatively high proportions of triploid loaches [12,13]. In China, tetraploid *M. anguillicaudatus* were frequently recorded in addition to the most widespread diploids [14,15], and small proportions of triploids have been detected at several locations [16]. Some studies have also reported the existence of specimens with rare ploidy levels, including pentaploids ($5N = 125$) [17] and hexaploids ($6N = 150$) [16]. Although the polyploidy phenomenon has received ample scientific attention, previous studies of ploidy level variation and geographical distribution of polyploid loaches in China were limited to specific localities or regions. To better understand the geographical distribution of polyploid loaches across China, we sampled populations in seven major river basins in mainland China. This study aims to identify population boundaries and deepen our understanding of polyploid population structure and distribution.

Phylogeography provides the framework to display causal relationships between geography, ecological interactions and the evolution of taxa [18]. The invention and development of molecular tools has facilitated large-scale phylogeographic studies. Mitochondrial DNA (mtDNA) sequences, especially the *cytochrome b* (*cyt b*) gene and the control region (CR), are frequently used in fish population genetics, phylogenetics and evolutionary history studies [19]. The *cyt b* gene evolves relatively slowly because it encodes a protein, whereas the non-coding CR evolves rapidly in vertebrates owing to the lack of coding constraints. With a high ratio of substitutions, tandem repeat quantity changes and indels of variable length [20], CR is believed to have a two to five times higher evolutionary rate than mitochondrial protein-coding genes. Previous studies demonstrated that polyploidization not only spotted some redundancy, but also causes nuclear enlargement, and complicates the processes of managing and segregation of chromosomes during cell division [21]. Meanwhile, mitochondria play an active role in cell division and chromosomal replication. Therefore, mtDNA sequence variation could reflect the polyploidy-driven changes in genetic diversity and evolution.

In this study, we performed a field investigation to reveal the geographical distribution patterns of polyploid and diploid *M. anguillicaudatus* populations across the Chinese mainland. Moreover, phylogeny, genetic variability and intraspecific structure were studied using *cyt b* and CR sequences. Subsequently, we further explored the origin and evolutionary history of polyploids in loach. This study shall help unravel the phylogeography and evolutionary history of polyploidy in *M. anguillicaudatus*.

2. Material and methods

(a) Samples

A total of 168 populations were sampled throughout the species' distribution range across China, from seven major river basins: Songhua River basin, Liaohe River basin, Haihe River basin, Yellow River basin, Huaihe River basin, Yangtze River basin and Pearl River basin [22]. Among these, 19 populations were not included in the study owing to the small number of specimens sampled. Sample size per population ranged from 22 to 70 specimens, with an average of 45 specimens per locality (electronic supplementary material, table S1). Caudal fin and blood were obtained from all sampled specimens. Caudal fins were preserved in 100% ethanol for subsequent DNA extraction and polymerase chain reaction (PCR) amplification. On the basis of the ploidy analyses, we further selected 351 specimens for sequencing of mtDNA genes (electronic supplementary material, table S2).

(b) Determination of ploidy levels

Flow cytometry (FCM) is an accurate and convenient method to detect variation in the ploidy level in large samples [23]. We collected blood from the tail vein mainly following the protocol described by Yu *et al*. [16]. Blood cells were suspended in 1 ml staining buffer (Cystain DNA 1 step staining solution) for 3 min. The FCM measurements were taken using an Elite flow cytometer (CyFlow Space, Japan). Erythrocytes of karyologically identified *M. anguillicaudatus* with $2n = 50$ were used as the diploid standard (internal control) [24]. Chicken blood cells with known DNA content of 2.5 pg nucleus^{-1} were used as an internal quantitative standard for each specimen's FCM profile [25].

(c) DNA extraction, polymerase chain reaction amplification and sequencing

The total genomic DNA was extracted from ethanol-preserved fin tissues using the phenol–chloroform extraction. Two mtDNA regions, *cyt b* and CR, were amplified via PCR. For *cyt b*, we designed a primer pair L14322 (5′ GACTTGAAGAACCACC GTTGTTATTCAAC 3′) and H15576 (5′ GCGCTAGGGAGGAA TTTAACCTCC 3′). For CR, we used primers designed by Tang *et al*. [26]: DL1 (5′ ACCCCTGGCTCCCAAAGC 3′) and DH1 (5′ ATCTTA GCATCTTCAGTG 3′).

For both sequences, reaction amplifications were carried out in 20 µl reactions containing: 10 µl of each Premix Taq (TaKaRa Taq v. 2.0 Plus dye), 0.5 µl of each primer, 1 µl genomic DNA (100 ng ml^{-1}) and 8 µl distilled water. The amplification was performed using the following conditions: for *cyt b*, 5 min at 94°C, 35 cycles of 30 s at 94°C, 30 s at 54°C and 45 s at 72°C, and a final extension of 10 min at 72°C; for the CR, conditions were identical with the exception of annealing step temperature (53°C). The PCR products were sent to Quintarabio Biotech Co., Ltd. (Wuhan, China) for purification and sequencing. Most sequences (600 bp) were obtained with a single (forward) primer, but to ensure the accuracy of nucleotide identification, all specimens assigned to singleton haplotypes and all samples showing double peaks at any nucleotide position were sequenced in the reverse direction as well.

(d) Data analysis

In order to explore the relationship and genetic structure of diploid and tetraploid loaches, we selected 22 locations that harboured sympatric diploids and tetraploids (electronic supplementary material, table S2). All nucleotide sequences (a total of 216 diploid specimens and 135 tetraploid specimens) were initially aligned using ClustalW implemented in MEGA 6.0 [27]. Haplotype diversity (Hd), nucleotide diversity (π) and number of haplotypes (h) were calculated using Arlequin 3.5 [28] and DnaSP 5.0 [29]. To test the correlation between haplotype/nucleotide diversity and geographical longitude and latitude, we conducted the Pearson correlation analysis using SPSS 19.0.

To study the relationships between different ploidies and among populations, phylogenetic trees were constructed using two methods, maximum likelihood (ML) and Bayesian inference (BI), using PhyML3.0 [30] and MrBayes 3.1.2 [31], respectively. MEGA 6.0 was used to select the T92 + G mtDNA substitution models on the basis of the Bayesian information criterion [32] for *cyt b* and CR. In the ML analysis, the statistical robustness of the

resulting tree nodes was determined by 1000 bootstrap replicates. The BI analysis also used the best-fit model of T92 + G (nst = 2, rates = gamma). We ran 1 000 000 generations with four chains, and trees were sampled every 100 generations with the first 2500 samples discarded as burn-in. Trees and posterior probabilities were visualized with FigTree v. 1.4.2. In addition, to investigate the relationship among haplotypes, a statistical parsimony haplotype network with median-joining (MJ) was built using Network 4.6 [33], and each indel was treated as a single mutation event.

Analysis of molecular variance (AMOVA) and fixation indices (F_{ST}) were calculated using Arlequin 3.5 [34]. The significance of the test was assessed using 10 000 permutations of each pairwise comparison. Genetic distance calculations within/between populations were calculated using MEGA 6.0 under the Kimura 2 model. Isolation-by-distance (IBD) was examined by testing the correlation between genetic distance $F_{ST}/(1-F_{ST})$ and geographical distance for all populations using GenAlEx 6.5 [35]. We measured the geographical distance between sites using the shortest waterway with Google Earth v. 5 (electronic supplementary material, table S3).

Historical population dynamics was investigated using Tajima's *D* [36], Fu's *Fs* [37] and mismatch distributions of pairwise differences [38]. Tajima's *D* and Fu's *Fs* implemented in Arlequin 3.5, with the deviation from neutrality determined from 10 000 coalescent simulations. To test the population expansion hypothesis, mismatch distributions of pairwise sequences were calculated using Arlequin 3.5 with 1000 bootstrap replicates.

To assess the historic migration pattern of *M. anguillicaudatus*, we applied a coalescent-based approach using the Bayesian search strategy implemented in Migrate-N v. 3.2.1 [39]. The full migration model was conducted through a long chain of 1×10^5 recorded genealogies with a sampling increment of 100 iterations. The burn-in was set to 1×10^4 for each locus. To improve the Markov chain Monte Carlo searches, an adaptive heating scheme with four chains, with temperatures 1, 1.5, 3 and 1 000 000, and a swapping interval of 1, was applied.

3. Results

(a) Geographical distribution of different ploidy levels

We genotyped and analysed a total of 7439 *M. anguillicaudatus* specimens from 168 populations throughout its distribution range in China, except for the western plateau. Three ploidy levels were detected: diploid, triploid and tetraploid (figure 1). Diploids were prevalent, with 97.03% of all specimens, followed by 2.58% tetraploids and 0.38% triploids. Diploid specimens were found at all 168 studied localities. In contrast with the widely distributed diploids, polyploid populations were primarily found at locations in the central Yangtze River basin, namely Han River, Dongting Lake and Poyang Lake, but some were also found in the Haihe and Pearl River basins: triploids appeared at 14 localities (7.49%) and tetraploids at 27 localities (14.44%). Exclusively polyploid populations were not detected. In more than 50% of the populations where diploids and tetraploids were sympatric, triploids were not detected, and we also recorded a co-occurrence of diploids and triploids in a few populations where tetraploids were not present. In 11 populations, diploid, triploid and tetraploid specimens were sympatric. In conclusion, there was no apparent pattern to the ploidy composition in populations where polyploid specimens appeared; distribution of polyploid specimens was limited to some regions, but there were no clearly allopatric geographical ranges or boundaries dividing diploid and polyploid populations of *M. anguillicaudatus* in China.

(b) Mitochondrial DNA sequence data and phylogenetic structure

Partial CR and *cyt b* sequences were successfully amplified from 351 samples of *M. anguillicaudatus*, among which 216 were diploids and 135 were tetraploids. Triploids were not used owing to the small number of identified specimens and their low statistical significance. CR sequences were 600 bp in length and contained 162 variable sites, 140 of which were parsimony informative. *Cyt b* sequences (607 bp) had 189 variable sites, 173 of which were parsimony informative. Among the 351 specimens, we found 134 CR haplotypes and 152 *cyt b* haplotypes. There were 24 haplotypes shared by two or more populations in the CR dataset, and 22 in the *cyt b* dataset. Each population contained at least one distinct haplotype (except for HP and LC from Han River and Haihe River respectively in CR; electronic supplementary material, table S2). The MJ network analysis and phylogenetic trees produced by CR and *cyt b* had similar topologies, split into five lineages (figure 2; electronic supplementary material, figures S1 and S2). This phylogenetic division did not fully correspond to the ploidy structure across the distribution range. Instead, sympatric diploids and tetraploids generally clustered within the same lineage, congruent with the geographical distribution. Lineage I, spanning a comparatively extensive geographical range, included samples mainly from the Han River, Dongting Lake, Haihe River and some from the Pearl River. Lineage II included only specimens from Poyang Lake. Most Poyang Lake haplotypes grouped together, separated from other haplotypes, except for several haplotypes from the Poyang Lake estuary, which grouped in lineage I. Lineage III contained specimens from the Pearl River, and some specimens from the ZJJ (from Dongting Lake) and HP (from the Han River) populations. Lineage IV comprised specimens from the Pearl River and ZJJ population. Lineage V consisted of two haplotypes that were found only in the XZ population (from the Pearl River), which was first separated (figure 2*a*,*b*) and clearly differentiated from other lineages. Interestingly, the haplotype structure of the Pearl River was complex, where most of the GL samples grouped with lineage I, as did approximately half of the PL samples (both from the Pearl River). The rest of the GL and PL specimens clustered in clades III and IV distantly separated from other lineages, while the XZ population specimens formed a private haplogroup V. The results suggest multiple genetically distinct sources of Pearl River populations. Diploid haplotypes from HP and ZJJ grouped into clades III and IV, rather than grouping with sympatric populations in clade I. However, tetraploid samples from HP and ZJJ haplotypes belonged to clade I (electronic supplementary material, figure S2).

(c) Genetic diversity and population structure

We investigated whether geographical locality influences the genetic diversity in *M. anguillicaudatus*. Population genetic diversity variation was very similar in CR and *cyt b* analyses (table 1). The lowest haplotype and nucleotide diversities were detected in the populations of Haihe River, whereas the Pearl River populations were characterized by highest nucleotide diversity. Haplotype and nucleotide diversities in the three rivers of the Yangtze River basin exhibited high diversity values. In the lineage analyses, the highest haplotype and nucleotide diversities were discovered in the

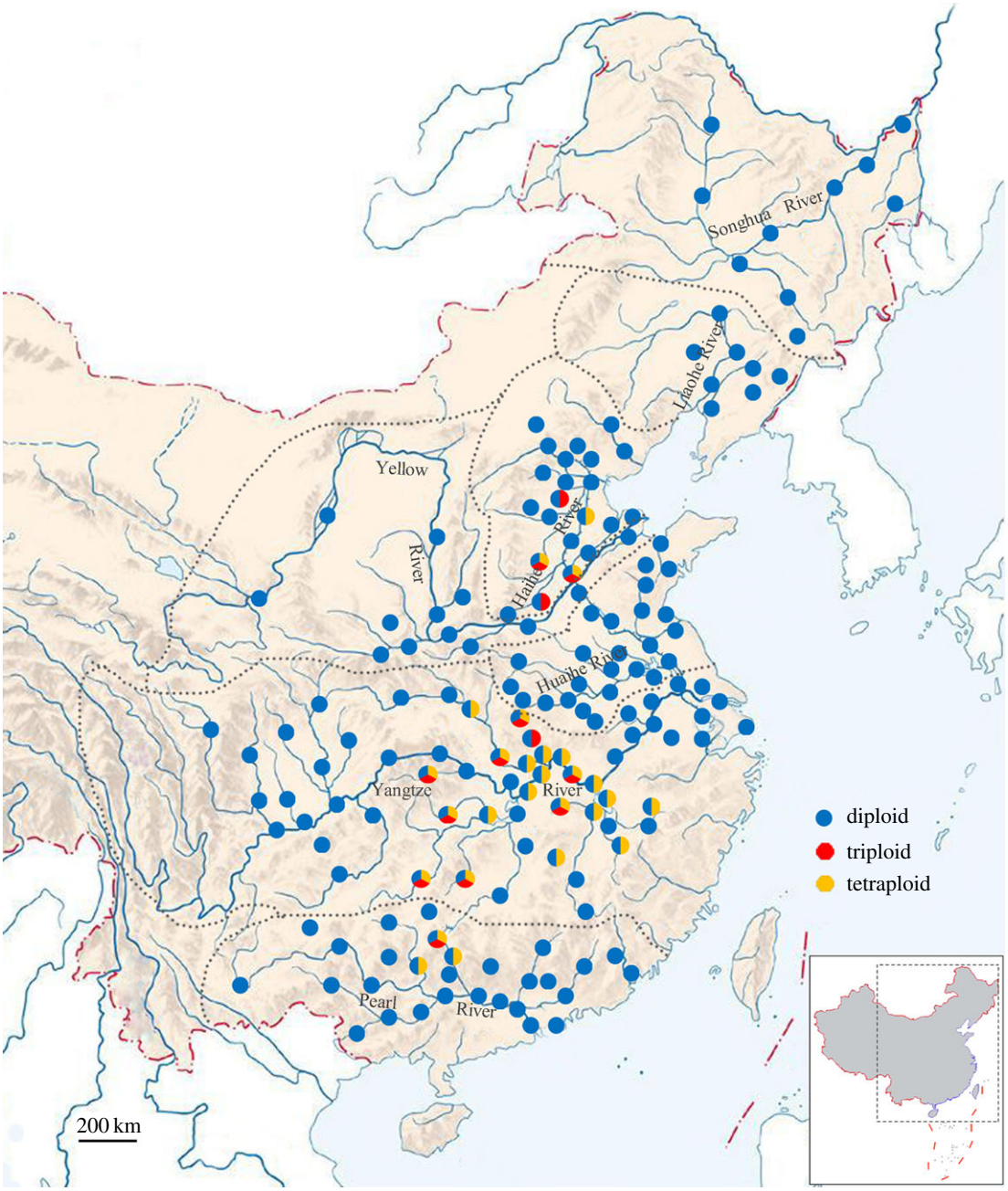

Figure 1. Geographical distribution of the 168 sampled populations. Dashed lines delineate the seven major drainage areas. Detailed information is available in the electronic supplementary material, table S1. (Online version in colour.)

populations of lineage II, whereas the lowest were identified in lineage V (electronic supplementary material, table S4). As regards the ploidy level, both diploid and tetraploid populations were characterized by high haplotype diversity values, but diploid populations had slightly higher nucleotide diversity values than tetraploids (electronic supplementary material, table S5). The diversity of both datasets (CR and *cyt b*) exhibited a negative correlation with latitude, i.e. haplotype and nucleotide diversity in both datasets decreases from south to north. Nucleotide diversity also decreased from west to east (visualized in the electronic supplementary material, figure S3).

Results of the AMOVA based on two mtDNA fragments revealed significant genetic differentiation among the populations of *M. anguillicaudatus* (electronic supplementary material, table S6). AMOVA revealed no genetic differentiation among different ploidy levels, as the interploidy variance component indices were negative and differences were not statistically significant ($p > 0.05$). By contrast, geographical regions and genetic lineages contributed substantially more than ploidy levels to the population structure (genetic differentiation); and we identified genetic differentiation among different regions ($p < 0.001$) and lineages ($p < 0.001$). Genetic distances among the five basins (1.749–5.357% for CR and 3.168–10.686% for *cyt b*) were notably higher than that within each basin (0.149–2.484% for CR and 0.117–4.697% for *cyt b*), except for the Pearl River, with extremely high intra-basin values of 6.474% and 9.869% for CR and *cyt b*, respectively. Among the intrapopulation genetic distances, XZ exhibited the highest value. In pairwise comparisons between location and population differentiation (F_{ST}), most population pairs (85.3% for CR and 76.6% for *cyt b*) were significantly different after Bonferroni's correction (electronic supplementary material, table S7).

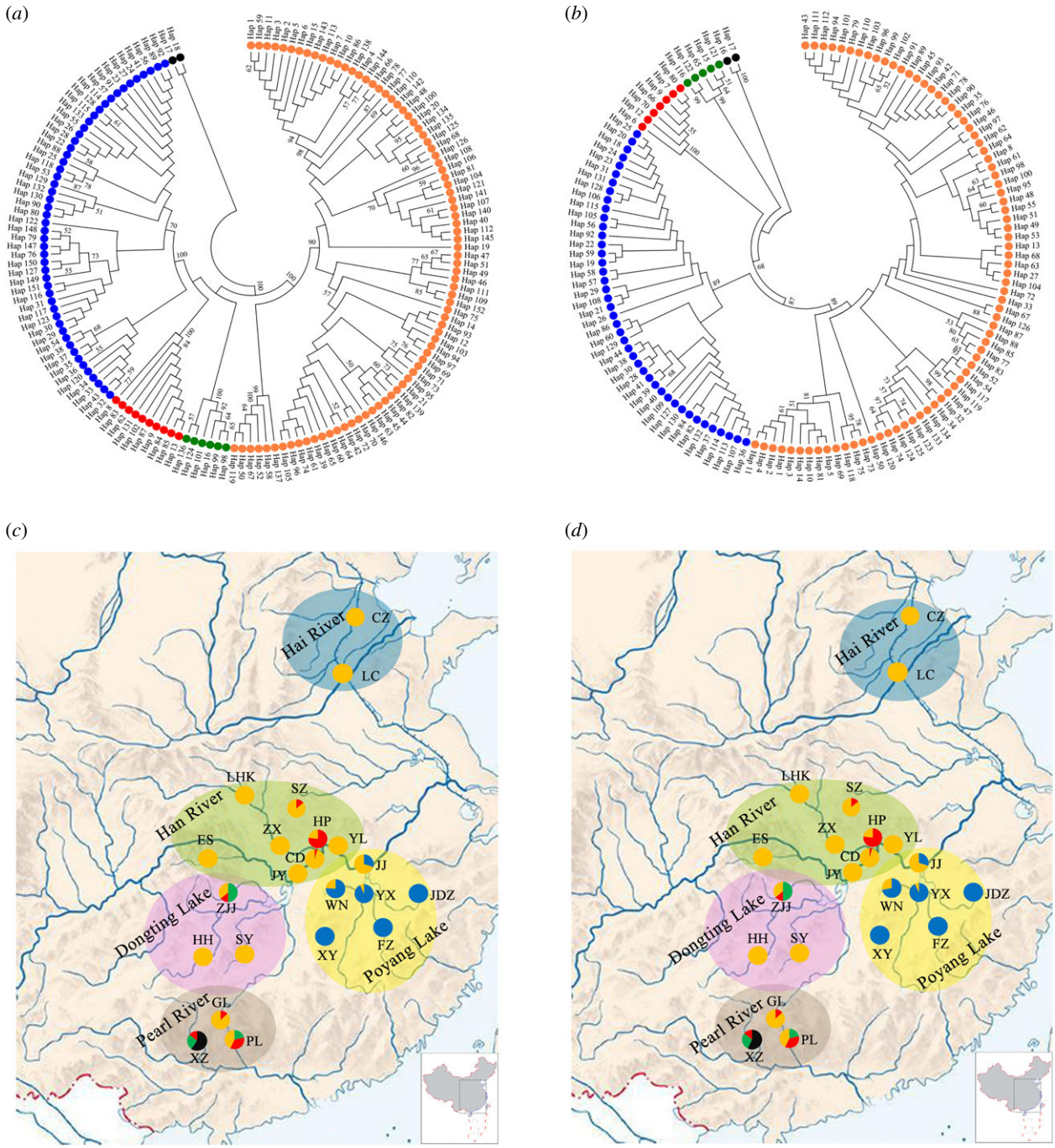

Figure 2. The structure of 22 mixed-ploidy *M. anguillicaudatus* populations inferred using *cyt b* and CR sequences (electronic supplementary material, tables S1 and S2). ML phylogenetic trees based on *cyt b* (*a*) and CR (*b*) datasets. Numbers above the branches correspond to the bootstrap support values (only values greater than 50% are shown). Clades are highlighted with different colours: yellow, lineage I; blue, lineage II; red, lineage III; green, lineage IV; black, lineage V. Maps showing the frequencies of the five clades in 22 populations in the *cyt b* (*c*) and CR (*d*) datasets. (*c*) and (*d*) panels differ only in the proportion of lineage V in the XZ population: 57% in *cyt b*, and 53% in CR; and lineage III in the GL population: 13% in *cyt b*, and 15% in CR. Basins are coloured with different backgrounds: blue, Haihe River; green, Han River; pink, Dongting Lake; yellow, Poyang Lake; grey, Pearl River. (Online version in colour.)

IBD was tested in diploid and tetraploid populations separately. In the *cyt b* dataset, comparison of F_{ST} versus distance (km) revealed non-significant IBD for tetraploid populations (one-tailed Mantel's tests: $r = 0.216$, $p = 0.067$), but significant IBD for diploid populations ($r = 0.345$, $p = 0.002$). For CR, tetraploid and diploid populations exhibited significant IBD (diploid: $r = 0.23$, $p = 0.018$; tetraploid: $r = 0.236$, $p = 0.036$). When we tested for IBD across mixed-ploidy populations, we also found that genetic distance increased with geographical distance in both datasets (CR: $r = 0.14$, $p = 0.045$; *cyt b*: $r = 0.39$, $p < 0.001$). In conclusion, genetic relationships among these populations are distance-dependent.

No evidence for demographic expansion was observed in multimodal mismatch distribution patterns (electronic supplementary material, figure S4a–j), or non-significant values for Fu's *Fs* and Tajima's *D* index ($p > 0.05$) (table 1) in the Han River, Dongting Lake, Poyang Lake and Pearl River, for which the expansion hypothesis was rejected. However, opposite neutrality test results were observed in the Haihe River, which may have been caused by the limited data available for this locality.

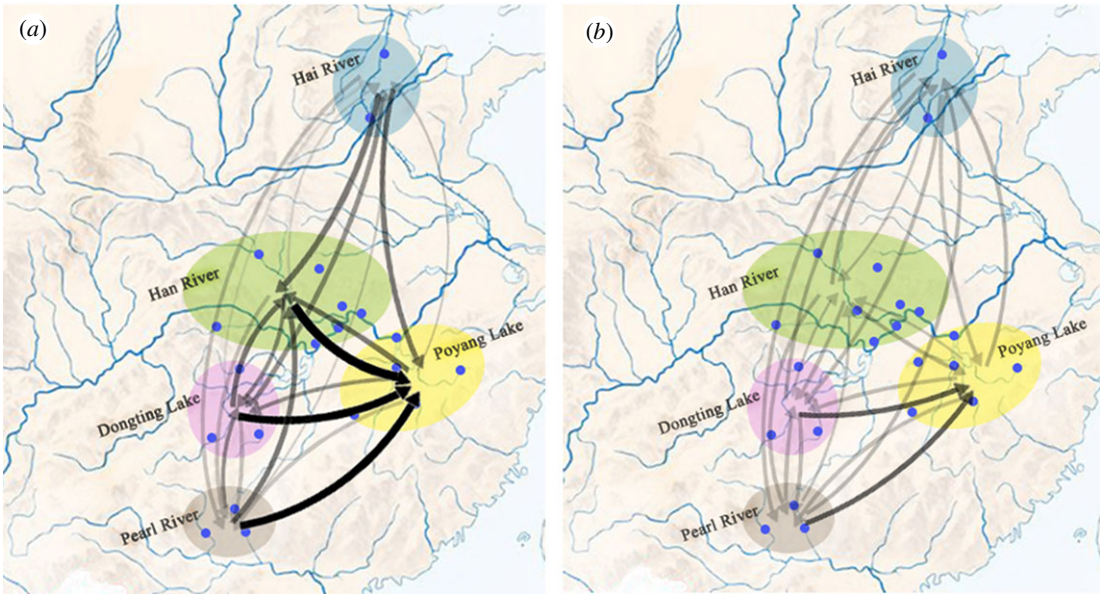

Figure 3. Migration of diploid populations (*a*) and tetraploid populations (*b*) estimated using MIGRATE-N. The arrows represent the direction of migration and their thickness is proportional to the number of migrants. (Online version in colour.)

Table 1. Genetic diversity indices and results of neutrality tests based on *cyt b* and CR datasets. (*n*, number of samples; *h*, number of haplotypes; Hd, haplotype diversity; π, nucleotide diversity; numbers in italics indicate statistically significant results ($p < 0.05$).)

sequence	region	*n*	*h*	Hd	π	Fu's *Fs*	Tajima's *D*
cyt b	Haihe River	25	6	0.367	0.00117	*−3.279*	*−2.064*
	Han River	130	55	0.954	0.03115	−7.579	−0.267
	Dongting Lake	47	24	0.966	0.04358	2.053	0.144
	Poyang Lake	98	57	0.985	0.03963	−9.441	0.450
	Pearl River	51	22	0.938	0.08729	12.519	2.177
CR	Haihe River	24	4	0.678	0.00149	−0.088	0.275
	Han River	128	41	0.936	0.01662	−5.976	−0.908
	Dongting Lake	47	27	0.966	0.02378	−2.370	0.412
	Poyang Lake	95	51	0.973	0.01490	−9.503	−0.730
	Pearl River	49	20	0.915	0.05917	8.978	1.378

(d) Gene flow between regions and populations

We used MIGRATE-N to infer the putative migration history, the effective population sizes and migration rates of diploids and tetraploids among regions and lineages (figure 3; electronic supplementary material, table S8). Estimates of directional gene flow in regions indicated that the migration among the tetraploid populations was lower than among the diploid populations. The migration among the tetraploid populations was also predominantly symmetrical (Nm = 0.49–0.72), with the exception of the migration rates from Dongting Lake and Pearl River to Poyang Lake (Nm = 1.29 and 1.28, respectively), which were almost two times larger than the gene flow in the opposite direction (figure 3*b*). By contrast, the migration between diploid populations was asymmetrical, exhibiting a general trend of migration towards the localities in the Yangtze basin, with north → centre, south → centre and west → east as dominant migratory directions (figure 3*a*). The diploid data showed that the populations at the ends of the sampled range (Pearl River and Haihe River) contributed to the genetic diversity across the central portion of the range (Nm = 0.89–1.64 and 0.80–1.11, respectively, electronic supplementary material, table S9), while they rarely received migrants (Nm = 0.63–0.69 and 0.04–0.09, respectively). When comparing diploids and tetraploids, all diploid gene flow between populations was higher than tetraploid gene flow, except for the migration rates towards the Haihe River (electronic supplementary material, table S9). A similar migration pattern was also inferred in the lineage analysis (electronic supplementary material, table S8). The migration between diploid populations was asymmetrical, with south → centre (lineage III/IV to lineage I/II) and west → east (lineage I to lineage II) directions, whereas the migration between tetraploid populations was symmetrical.

4. Discussion

In this study, diploid cytotypes were the dominant ploidy level, continuously distributed all over China. This is in

agreement with previous studies, both in China [16] and Japan [12]. The distribution of polyploids was fragmented, and they were mixed with diploids, which is also in agreement with previous studies [17]. The absence of a clearly defined geographical range or boundary for the distribution of diploids and polyploids of *M. anguillicaudatus* is in disagreement with studies in amphibians [40] and plants [41], where polyploids usually exhibit regular biogeographic patterns. The mtDNA analyses suggest a high level of genetic diversity and significant differentiation of populations in the Pearl River basin. Our data also indicate that the Pearl River basin could be the potential origin centre of *M. anguillicaudatus*, as the basal lineage V comprised only populations from this basin. However, there is no apparent geographical underpinning for samples from two Pearl River locations (GL and PL) and Han River clustering together, so we hypothesize that the most likely explanation for this anomalous result is human-mediated movement, which can generate complex genetic structuring at regional and local scales and affect population genetic patterns [42,43]. In other river basins, phylogenetic analysis results were congruent with geographical distribution, which is in agreement with previous studies [44].

Sympatric tetraploid and diploid specimens belonged to the same lineages and shared the same haplotypes. This was also observed before [45]. AMOVA revealed that populations did not exhibit a significant structuring by the ploidy level. From the theoretical standpoint, a single origin of polyploidy would result in a single ancestral mtDNA haplotype, and descendants of the original polyploid should form a single lineage. Therefore, the lack of monophyly of polyploids strongly suggests that they originated from diploids via multiple independent autopolyploidization events. This hypothesis is in agreement with a proposal that *M. anguillicaudatus* is an autotetraploid [46]. Multiple independent polyploidization events and recurrent formation of polyploid lineages are the norm in many species, not only in plants [47], but also in animals, including ostracods [48], reptiles [49], amphibians [50] and fishes [51–53].

Owing to polysomic inheritance, which enables them to maintain three or four alleles at a single locus, polyploids can have higher heterozygosity than diploids, so in some cases, polyploids can have higher environmental adaptability than their diploid counterparts [2]. As a result, polyploids are often able to occupy new habitats and survive a broader range of environmental conditions, i.e. they can be highly invasive [53]. Greater ecological amplitude, such as higher tolerance to altitude and extreme conditions, in comparison to their diploid progenitors has been demonstrated in some polyploid plants and animals [54,55]. Also, in some species, geographical distribution of diploids can be a subset of the distribution of polyploids, with *Plantago media* [56] and *Veronica chamaedrys* [57] as examples. In *M. anguillicaudatus*, we found an opposite pattern of cytotype distribution, in which polyploids exhibited a localized, fragmented distribution. Thus, our results indicate that in the current situation, polyploid *M. anguillicaudatus* specimens/populations probably exhibit slightly decreased fitness levels in comparison to diploids. Although the presence of palaeopolyploidy indicates that genome duplication is not always an evolutionary dead-end, the long-term survival of some ancient polyploids does not mean that genome multiplication necessarily leads to evolutionary innovations and increased diversity [58]. Both of our datasets consistently indicated a slightly higher nucleotide and haplotype diversity in diploids than in tetraploids. Also, while there was a definite directionality in the migration patterns of diploid populations, tetraploid populations exhibited low, symmetric migration rates, which indicates low dispersal ability. These results indicate that diploids possess better environmental adaptation abilities than tetraploids, which conforms to the pattern that recently originated polyploids usually fail to persist and diversify at lower rates. Our findings therefore suggest that multiple tetraploidization events occurred independently in *M. anguillicaudatus* populations on numerous occasions and in different regions, and that the observed frequencies of polyploids in some *M. anguillicaudatus* populations are probably a consequence of relatively high rates of their formation, rather than their high fitness. This is in agreement with observations in some other species [59,60].

Historically, geological events, such as the Qinghai-Tibet Plateau orogeny, would occasionally re-shape and re-arrange the drainage systems in the studied range [61]. This includes the formation of the Yangtze River, with its predominant west-to-east flow [62]. Evolutionary histories (including dispersion and gene flow) of many fish taxa were affected by these events [63]. We hypothesize that the northward migration pattern of diploid *M. anguillicaudatus* was also influenced by the geological history: during the Last Glacial Maximum (LGM), the sea level dropped significantly, the ancient coastline advanced significantly into the (nowadays) sea, and the continental shelf was mostly exposed (became land). This, along with relatively arid climate, often caused water shortages in the Yangtze River basin [64]. As opposed to this, the aquatic habitats in the lower Pearl River reaches were likely to be less affected by these events [65]. The Pearl River to Yangtze River migration pattern observed in our study (figure 3*a*; electronic supplementary material, tables S8 and S9) supports a hypothesis that when the water levels rose with the subsequent post-LGM warming, populations inhabiting the Pearl River basin colonized the available habitats in the Yangtze River basin. A similar scenario was proposed for *Hemiculter leucisculus* populations [66]. The southward migration pattern, with more migrants moving from the Haihe River to the Yangtze River than in the reverse direction (figure 3*a*; electronic supplementary material, table S9), may have been driven by more recent events, namely the construction of the Beijing-Hangzhou Grand Canal, which directly connected the Haihe River and the Yangtze River 2500 years ago.

Ethics. All experiments in this research were performed according to the guidelines established by Huazhong Agricultural University (HZAU), Chinese Academy of Sciences, and the experimental protocols were approved by the Institutional Animal Care and Use Committee of HZAU, Chinese Academy of Sciences.

Data accessibility. Sequence data of mtDNA are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.n3s0p79 [67].

Authors' contributions. W.W. conceived and supervised the whole study. J.Z. and S.Y performed fieldwork and blood sampling. J.Z and L.M. performed laboratory works and data analysis. W.W. and J.Z. wrote the manuscript.

Competing interests. The authors declare that they have no competing interests.

Funding. This study was supported by the National Natural Science Foundation of China (grant no. 31372180).

Acknowledgements. We would like to thank S. Wang, R.J. Geng, X.R. Song, J.T. Li, Y.X. Xie and Y.D. Shen who provided invaluable support in the sample collection. We thank S.M. Wan, L. Zhong and X.Z. Guo for their suggestions in data analyses.

References

1. Soltis PS, Liu X, Marchant DB, Visger CJ, Soltis DE. 2014 Polyploidy and novelty: Gottlieb's legacy. *Phil. Trans. R. Soc. B* **369**, 20130351. (doi:10.1098/rstb.2013.0351)
2. Otto SP, Whitton J. 2000 Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**, 401–437. (doi:10.1146/annurev.genet.34.1.401)
3. Mable BK. 2015 'Why polyploidy is rarer in animals than in plants': myths and mechanisms. *Biol. J. Linn. Soc.* **82**, 453–466. (doi:10.1111/j.1095-8312.2004.00332.x)
4. Cole CJ. 1979 Chromosome inheritance in parthenogenetic lizards and evolution of allopolyploidy in reptiles. *J. Hered.* **16**, 2539–2540. (doi:10.1039/dt9930002539)
5. Mable BK, Alexandrou MA, Taylor MI. 2011 Genome duplication in amphibians and fish: an extended synthesis. *J. Zool.* **284**, 151–182. (doi:10.1111/j.1469-7998.2011.00829.x)
6. Holland PWH, Sharman AC. 1995 Conservation, duplication, and divergence of developmental genes during chordate evolution. *Neth. J. Zool.* **46**, 47–67. (doi:10.1163/156854295X00050)
7. Meyer A, Van DPY. 2010 From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays* **27**, 937–945. (doi:10.1002/bies.20293)
8. Luo J *et al.* 2014 Tempo and mode of recurrent polyploidization in the *Carassius auratus* species complex (Cypriniformes, Cyprinidae). *Heredity* **112**, 415–427. (doi:10.1038/hdy.2013.121)
9. Mei J, Gui JF. 2015 Genetic basis and biotechnological manipulation of sexual dimorphism and sex determination in fish. *Sci. China Life Sci.* **58**, 124. (doi:10.1007/s11427-014-4797-9)
10. Zeng L, Wang J, Sheng J, Gu Q, Hong Y. 2012 Molecular characteristics of mitochondrial DNA and phylogenetic analysis of the loach (*Misgurnus anguillicaudatus*) from the Poyang Lake. *Mitochondrial DNA* **23**, 187–200. (doi:10.3109/19401736.2012.668893)
11. Ojima Y, Takai A. 1979 The occurrence of spontaneous polyploid in the Japanese common loach, *Misgurnus anguicaudatus*. *P. Jpn Acad. B Phys.* **55**, 487–491. (doi:10.2183/pjab.55.487)
12. Arai K. 2003 Genetics of the loach, *Misgurnus anguillicaudatus*: recent progress and perspective. *Folia Biol.* **51**, 107–117. (doi:10.1117/12.499216)
13. Fujimoto T *et al.* 2017 Development of nuclear DNA markers to characterize genetically diverse groups of *Misgurnus anguillicaudatus* and its closely related species. *Fish Sci.* **5**, 743–756. (doi:10.1007/s12562-017-1108-y)
14. Yin J, Zhao ZS, Chen XQ, Li YQ, Zhu LY. 2005 Karyotype comparison of diploid and tetraploid loach, *Misgurnus anguillicanudatus*. *Acta Hydrobiol. Sin.* **29**, 469–472. (doi:10.1360/biodiv.050121)
15. Feng B, Yi SV, Zhang MM, Zhou XY. 2018 Development of novel EST-SSR markers for ploidy identification based on de novo transcriptome assembly for *Misgurnus anguillicaudatus*. *PLoS ONE* **13**, e0195829. (doi:10.1371/journal.pone.0195829)
16. Yu YY, Abbas K, Wang WM, Zhou XY. 2014 Geographical distribution of ploidy level variation of loach *Misgurnus anguillicaudatus* in China. *Pak. J. Agr. Sci.* **51**, 273–281. (doi:10.1016/j.compag.2014.04.005)
17. Cui L, Abbas K, Yu YY, Wang WM, Zhou L, Zhou XY. 2013 First record of the natural occurrence of pentaploid loach, *Misgurnus anguillicaudatus* in Hubei Province, China. *Folia Zool.* **62**, 14–18. (doi:10.1016/j.beproc.2013.01.002)
18. Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF, Yoder AD. 2010 Phylogeography's past, present, and future: 10 years after. *Mol. Phylogenet. Evol.* **54**, 291–301. (doi:10.1016/j.ympev.2009.09.016)
19. Peng Z, He S, Zhang Y. 2004 Phylogenetic relationships of glyptosternoid fishes (Siluriformes: Sisoridae) inferred from mitochondrial cytochrome *b* gene sequences. *Mol. Phylogenet. Evol.* **31**, 979–987. (doi:10.1016/j.ympev.2003.10.023)
20. Sbisà E, Tanzariello F, Reyes A, Pesole G, Sacone C. 1997 Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene* **205**, 125–140. (doi:10.1016/S0378-1119(97)00404-6)
21. Comai L. 2005 The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **6**, 836–846. (doi:10.1038/nrg1711)
22. Lu XP, Fu ZC. 2010 General theory. In *A survey of major river systems in China* (eds XP Lu, ZC Fu), pp. 1–6. Beijing, China: China Water Power Press
23. Kron P, Suda J, Husband BC. 2007 Applications of flow cytometry to evolutionary and population biology. *Annu. Rev. Ecol. Evol. Syst.* **38**, 847–876. (doi:10.1146/annurev.ecolsys.38.091206.095504)
24. Hardie DC, Gregory TR, Hebert PDN. 2002 From pixel to picograms: a beginner's guide to genome quantification by Feulgen image analysis densitometry. *J. Histochem. Cytochem.* **50**, 735–749. (doi:10.1177/002215540205000601)
25. Tiersch TR, Chandler RW. 1989 Chicken erythrocytes as an internal reference for analysis of DNA content by flow cytometry in grass carp. *Trans. Am. Fish. Soc.* **118**, 713–717. (doi:10.1577/1548-8659(1989)118<0713:CEAAIR>2.3.CO;2)
26. Tang QY, Liu HZ, Mayden R., Xiong BX. 2006 Comparison of evolutionary rates in the mitochondrial DNA cytochrome *b* gene and control region and their implications for phylogeny of the Cobitoidea (Teleostei: Cypriniformes). *Mol. Phylogenet. Evol.* **39**, 347–357. (doi:10.1016/j.ympev.2005.08.007)
27. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013 MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729. (doi:10.1093/molbev/mst197)
28. Excoffier L, Laval G, Schneider S. 2005 Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform.* **1**, 47–50. (doi:10.1143/JJAP.34.L418)
29. Librado P, Rozas J. 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452. (doi:10.1093/bioinformatics/btp187)
30. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010 New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321. (doi:10.2307/25677586)
31. Ronquist F, Huelsenbeck JP. 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574. (doi:10.1093/bioinformatics/btg180)
32. Schwarz G, Schwartz GJ. 1978 Estimating dimension of a model. *Ann. Stat.* **6**, 461–464. (doi:10.1214/aos/1176344136)
33. Bandelt HJ, Forster P, Rohl A. 1999 Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48. (doi:10.1093/oxfordjournals.molbev.a026036)
34. Excoffier L, Lischer HEL. 2010 Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567. (doi:10.1111/j.1755-0998.2010.02847.x)
35. Peakall R, Smouse PE. 2012 GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**, 2537–2539. (doi:10.1093/bioinformatics/bts460)
36. Tajima F. 1989 Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.
37. Fu YX. 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925.
38. Rogers AR, Harpending H. 1992 Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**, 552–569. (doi:10.1093/oxfordjournals.molbev.a040727)
39. Beerli P. 2006 Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* **22**, 341–345. (doi:10.1093/bioinformatics/bti803)
40. Smith MJ, Roberts JD, Hammond TJ, Davis RA. 2003 Intraspecific variation in the advertisement call of the sunset frog *Spicospina flammocaerulea* (Anura: Myobatrachidae): a frog with a limited geographic distribution. *J. Herpetol.* **37**, 285–291. (doi:10.1670/0022-1511(2003)037[0285:IVITAC]2.0.CO;2)
41. Trávníček P, Kubátová B, Čurn V, Rauchová J, Krajníková E, Jersákováet J, Suda J. 2011 Remarkable coexistence of multiple cytotypes of the *Gymnadenia conopsea* aggregate (the fragrant orchid): evidence from flow cytometry. *Ann. Bot.-London* **107**, 77–87. (doi:10.1093/aob/mcq217)
42. Bock DG, Zhan A, Lejeusne C, Maclsaac HJ, Cristescu ME. 2011 Looking at both sides of the invasion:

patterns of colonization in the violet tunicate *Botrylloides violaceus*. *Mol. Ecol.* **20**, 503–516. (doi:10.1111/j.1365-294X.2010.04971.x)

43. Yi SK, Wang WM, Zhou XY. In press. Genomic evidence for the population genetic differentiation of *Misgurnus anguillicaudatus* in the Yangtze River basin of China. *Genomics*, S0888754318301113. (doi:10.1016/j.ygeno.2018.02.011)
44. Abbas K, Zhou XY, Wang WM. 2017 Microsatellite markers reveal genetic differentiation of Chinese dojo loach *Misgurnus anguillicaudatus* in the Yangtze river basin. *Turk. J. Fish Aquat. Sci.* **17**, 1167–1177. (doi:10.4194/1303-2712-v17_6_10)
45. Yi SK, Zhou XY, Li J, Zhang MM, Luo SS. 2018 Full-length transcriptome of *Misgurnus anguillicaudatus* provides insights into evolution of genus *Misgurnus*. *Sci. Rep.-UK* **8**, 11699. (doi:10.1038/s41598-018-29991-6)
46. Li YJ, Yu Z, Zhang MZ, Qian C, Abe S, Arai K. 2011 The origin of natural tetraploid loach *Misgurnus anguillicaudatus* (Teleostei: Cobitidae) inferred from meiotic chromosome configurations. *Genetica* **139**, 805–811. (doi:10.1007/s10709-011-9585-x)
47. Rebernig CA, Weiss-Schneeweiss H, Schneeweiss GM, Schönswetter P, Obermayer R, Villaseñor JL, Stuessy TF. 2010 Quaternary range dynamics and polyploid evolution in an arid brushland plant species (*Melampodium cinereum*, Asteraceae). *Mol. Phylogenet. Evol.* **54**, 594–606. (doi:10.1016/j.ympev.2009.10.010)
48. Little TJ, Hebert PDN. 1997 Clonal diversity in high arctic ostracodes. *J. Evol. Biol.* **10**, 233–252. (doi:10.1046/j.1420-9101.1997.10020233.x)
49. Bogart JP. 1980 Evolutionary implications of polyploidy in amphibians and reptiles. *Polyploidy* **13**, 341–378. (doi:10.1007/978-1-4613-3069-1_18)
50. Holloway AK, Cannatella DC, Gerhardt HC, Hillis DM. 2006 Polyploids with different origins and ancestors form a single sexual polyploid species. *Am. Nat.* **167**, 88–101. (doi:10.1086/501079)
51. Lukáš C, Karel J, Koen DG, Jörg B, Věra S, Marie R, Ráb P. 2012 Synthesis of clonality and polyploidy in vertebrate animals by hybridization between two sexual species. *Evolution* **66**, 2191–2201. (doi:10.2307/23262063)
52. Liu XL, Jiang FF, Wang ZW, Li XY, Zhi L. 2017 Wider geographic distribution and higher diversity of hexaploids than tetraploids in *Carassius* species complex reveal recurrent polyploidy effects on adaptive evolution. *Sci. Rep.-UK* **7**, 5395. (doi:10.1038/s41598-017-05731-0)
53. Jakovlić I, Gui JF. 2011 Recent invasion and low level of divergence between diploid and triploid forms of *Carassius Auratus* complex in Croatia. *Genetica* **139**, 789–804. (doi:10.1007/s10709-011-9584-y)
54. Thompson JD, Lumaret R. 1992 The evolutionary dynamics of polyploid plants: origins, establishment and persistence. *Trends Ecol. Evol.* **7**, 302–307. (doi:10.1016/0169-5347(92)90228-4)
55. Hoffmann A *et al.* 2015 Genetic diversity and distribution patterns of diploid and polyploid hybrid water frog populations (*Pelophylax esculentus* complex) across Europe. *Mol. Ecol.* **24**, 4371–4391. (doi:10.1111/mec.13325)
56. Van DP, Hartog M, Van DW. 1992 Single cytotype areas in autopolyploid *Plantago media* L. *Biol. J. Linn. Soc.* **116**, 315–331. (doi:10.1111/j.1095-8312.1992.tb00867.x)
57. Bardy KE, Albach DC, Schneeweiss GM, Fischer MA, Schönswetter P. 2010 Disentangling phylogeography, polyploid evolution and taxonomy of a woodland herb (*Veronica chamaedrys* group, Plantaginaceae s.l.) in southeastern Europe. *Mol. Phylogenet. Evol.* **57**, 771–786. (doi:10.1016/j.ympev.2010.06.025)
58. Symonds VV, Soltis PS, Soltis DE. 2010 Dynamics of polyploid formation in Tragopogon (Asteraceae): recurrent formation, gene flow, and population structure. *Evolution* **64**, 1984–2003. (doi:10.1111/j.1558-5646.2010.00978.x)
59. Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP. 2011 Recently formed polyploid plants diversify at lower rates. *Science* **333**, 1257. (doi:10.1126/science.1207205)
60. Escudero M *et al.* 2014 Karyotypic changes through dysploidy persist longer over evolutionary time than polyploid changes. *PLoS ONE* **9**, e85266. (doi:10.1371/journal.pone.0085266)
61. Clark MK, Schoenbohm LM, Royden LH, Whipple KX, Burchfiel BC, Zhang X, Tang W, Wang E, Chen L. 2004 Surface uplift, tectonics, and erosion of eastern Tibet from large-scale drainage patterns. *Tectonics* **23**, 1–18. (doi:10.1029/2002tc001402)
62. Li J, Xie S, Kuang M. 2001 Geomorphic evolution of the Yangtze gorges and the time of their formation. *Geomorphology* **41**, 125–135. (doi:10.1016/S0169-555X(01)00110-6)
63. Yu D, Chen M, Tang QY, Li XJ, Liu HZ. 2014 Geological events and Pliocene climate fluctuations explain the phylogeographical pattern of the cold water fish *Rhynchocypris oxycephalus* (Cypriniformes: Cyprinidae) in China. *BMC Evol. Biol.* **14**, 1–12. (doi:10.1186/s12862-014-0225-9)
64. Yang D. 1986 The paleoenvironment of the mid-lower regions of Yangtze River in the full-glacial period of late Pleistocene. *Acta Geogr. Sin.* **41**, 302–310. (doi:10.11821/xb198604002)
65. Tian Z, Jiang D. 2016 Revisiting last glacial maximum climate over China and East Asian monsoon using PMIP3 simulations. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **453**, 115–126. (doi:10.1016/j.palaeo.2016.04.020)
66. Chen W, Zhong Z, Dai W, Fan Q, He S. 2017 Phylogeographic structure, cryptic speciation and demographic history of the sharpbelly (*Hemiculter leucisculus*), a freshwater habitat generalist from southern China. *BMC Evol. Biol.* **17**, 216. (doi:10.1186/s12862-017-1058-0)
67. Zhong J, Yi S, Ma L, Wang W. 2019 Data from: Evolution and phylogeography analysis of diploid and polyploid *Misgurnus anguillicaudatus* populations across China. Dyrad Digital Repository. (https://doi.org/10.5061/dryad.n3s0p79)