

The Phylogeographic Relationships and the Evolution History of *Carassius Auratus* Complex With a Newborn Homodiploid Crucian Carp-like Fish (2nNCRC)

Qianhong Gu

Hunan Normal University

Shi Wang

Hunan Normal University

Hui Yuan

Hunan Normal University

Hui Zhong

Hunan Normal University

Junliu Yang

Hunan Normal University

Conghui Yang

Hunan Normal University

Xuexue Huang

Hunan Normal University

Xiaowei Xu

Hunan Normal University

Yude Wang

Hunan Normal University

Zehong Wei

Hunan Normal University

Jing Wang

Hunan Normal University

Shaojun Liu (✉ lsj@hunnu.edu.cn)

Hunan Normal University

Research Article

Keywords: Eurasia, *Carassius auratus* complex, hybrid speciation, divergence time, phylogeography

Posted Date: June 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-604445/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: One of the important aspects of studying evolution is to understand how new species are formed and their uniqueness maintained. Hybridization can lead to the formation of new species with the reorganization of adaptive system and significant changes in phenotype. It is wondrous that eight stable strains of 2nNCRC derived from the interspecies hybridization have been established in our laboratory. To examine the phylogeographical pattern of the widely distributed genus *Carassius* in the Eurasia, and investigate the possible hybrid origin of *Carassius auratus* lineage, in light of past climatic events, the mitochondrial genome (mtDNA) were used to reconstruct the phylogenetic relationship between the *C. auratus* complex and the 2nNCRC, and to assess how demographic history, dispersal and barriers to gene flow have led to the current distribution of mtDNA lineages for *C. auratus* complex.

Results: As expected, the 2nNCRC had a very close relationship with the *C. auratus* complex, which was distinctly separated with other three species of *Carassius*. The *C. auratus* lineage possibly originated from China during the Late Pliocene, far postdated the diversification of *C. carassius* in Europe and *C. cuvieri* in Japan. The admixture of mtDNA haplotype lineages of *C. auratus* detected across the whole Eurasia has experienced a rapid diversification since Early Pleistocene

Conclusion: Combined the molecular dating analyses, species distribution modeling and ancestral area reconstruction, the speciation of *C. auratus* seemed not to be the processing of lineage diversification from the most recent common ancestor of *C. carassius* or *C. cuvieri*. The formation of 2nNCRC in our laboratory could be a good candidate explaining for the hybrid origin species for *C. auratus* lineage, as well as the paleoclimate oscillation and geological event during Pliocene and Pleistocene in China supplying an opportunity for the distant hybridization. The most widely distributed *C. auratus* lineage could be attributed to the dispersal during the glacial period and the recent human-facilitated dispersal.

Background

A fundamental question in evolutionary biology is whether speciation is gradual or punctuated [1]. The hybridization is regarded as an 'evolutionary catalyst' which may be an important source for the origin of a new species or increase the quantum of genetic variability. Hybridization can result in the reorganization of adaptive system and lead to the formation of new species [2]. Traditional concept holds that the speciation was the processing of lineage diversification [3]. However, recent studies have confirmed that the hybridization speciation by means of Lineage fusion was also one of the important species diversification [4, 5]. Hybridization between species can trigger vast genetic and genomic imbalances, including a high rate of DNA mutations and combinations [6], often results in significant changes in phenotypes and genotypes of hybrid offspring, which may facilitate speciation and adaptive radiation.

Drastic climatic and geological oscillations, such as the glacial expansion and mega-droughts forced freshwater species to contract their distribution ranges and reside in small refuge in many cases. The

potential gene exchange and hybrid speciation was likely to happen owing to the close contact between distant species. Increasingly, studies suggest that it is not uncommon on the inter-species hybridization and gene introgression in fish, such as the families Poeciliidae, Atherinidae, Cyprinidae, Cobitidae, and even the Carcharhinidae [7–9], especially the East African cichlids, which have been attracted much attention in recent years [10–12]. As the most number of species in vertebrate, more than 34000 species have been recorded (<https://www.fishbase.de/>). It was deduced that the formations of many fishes were associated with hybridization [13, 14]. The ancestral hybridization has been suggested to play a key role in facilitating species diversification of cichlids fish in the Lake Malawi, Lake Victoria, as well as Lake Tanganyika [10–12]. Such hybridization would be common when allopatric species come into second contact due to the drainage evolution or huge environmental changes, such as the mega-droughts caused strong water level changes and probably led to severe reduction of their natural habitat, which could increase the odds of interaction between fish species and eventually cause the hybridization between distant relatives. The ancient hybridization of Lake Victoria cichlids fish was considered to be related to the capture of Malagarasi (Congo) tributaries and the East African mega-droughts, providing an opportunity for the Congolese lineage colonized in the Lake Victoria region and Congo-Nilotic admixture event [10].

However, for a long time, it lacks of enough and direct evidences to prove that fish hybridization will produce new species. The distant hybridization can generate allotetraploid and autotetraploid fish, which supply an important and direct evidence for the hybridization speciation in fish [15, 16]. It was reported that the 2nNCRC (2n = 100) could be derived from the interspecies hybridization between *Cyprinus carpio* (female, 2n = 100) and *Megalobrama amblycephala* (male, 2n = 48) in our lab. In addition, eight stable generations for 2nNCRC with a very clear genetic background have been founded in our lab since 2014, supplying a great value model system for studying the hybrid speciation in fish.

It is widely believed that there exist five species in genus *Carassius*, including *Carassius carassius*, *Carassius cuvieri*, *Carassius langsdorfii*, *Carassius gibelio* and *Carassius auratus*, which have been considered as valid species in Fishbase (<https://www.fishbase.de/>). Well, whether the origin of the *C. auratus* is identical with that of 2nNCRC or not, and whether the *C. auratus* is evolved from 2nNCRC are unknown. It is very necessary to illuminate the historical evolutionary processes of *C. auratus*. The phylogeographic study might settle this issue because it is usually used to reveal the historical processes affecting vicariance, dispersal, extinctions and radiations that lead to the geographic distributions of genetic lineages [17, 18].

The recent climatic oscillation and geological events during Pliocene-Pleistocene have played a fundamental role in shaping contemporary patterns of biodiversity and species diversification and distribution, which are of main interest in phylo- and biogeography [19, 20]. Pleistocene glaciations are known to have exerted a far-reaching influence on the evolution of organisms in the Northern Hemisphere [21, 22]. For example, palaeontological and genogeographical studies indicate that European and North American species experienced repeated episodes of contraction and expansion of their ranges due to major climatic oscillations [23]. The *Carassius* species complex has a wide distribution across the

Eurasian continent and neighboring islands [24, 25]. According to the fossil record of *Carassius* species in Pliocene epoch (5.3–2.6 Mya, million years ago) discovered in north of China [26], the Quaternary palaeoenvironmental changes in East Asia and Europe would have a great influence on the speciation and diversification of *Carassius* species. In freshwater fishes, the dynamics of re-colonization are tightly linked to the history of river drainage systems [27]. During glacial melt periods, ephemeral rivers and periglacial lakes could arise, as well as the reconfiguration of the landscape caused by drastic climatic change and geological events may allow species to disperse into new habitats, providing opportunities for colonization and new species interactions.

These processes have resulted in complicated recolonization scenarios for the *Carassius* species complex in East Asia, where the haplotypes of mitochondrial control region and *tf* alleles were clustered into four and three major lineages respectively, and they both speculated that the Yangtze River basin was the potential origin center for *Carassius* species complex, and then radiated across East Asia [25, 28], the highest genetic diversity in the Yangtze River basin suggesting that it should be the refuge for *Carassius* species during glaciations. Gao et al. (2012) deduced that the close relationship between the *C. auratus* complex from eastern mainland China and the south-central Ryukyus was the result of a natural Pleistocene dispersal [24]. However, the existence of two distinct lineages of *C. carassius* in Europe [27], was mainly due to the Danubian catchment separated with other river system by the Alps, the Sudety Mountains and the Carpathian Mountains.

According to the formation of the newborn of 2nNCRC, combined with the close phylogenetic relationship between *C. carpio* and *Carassius* species complex [29, 30], we speculated that the existence of possible route through distant hybridization under natural conditions can generate *C. auratus*, which has happened during Pliocene–Pleistocene based on these hypotheses as follows: (1) the origin times of *Cyprinus* and *Megalobrama* are much earlier than that of *Carassius* [24, 30, 31]; (2) according to the distribution of current fish species and fossile data [26, 32], as well as the divergence time estimation for species in *Carassius*, the speciation of *C. auratus* in China might be punctuated and not be derived from the processing of lineage differentiation with other species; (3) the land bridge between China and Japan during glacial cycling providing an opportunity for the expansion of *C. auratus* from China to Japan or other areas. Thus, establishing a robust time-calibrated phylogeny is a first requirement for tracing the possible origin and the diversification patterns for *Carassius* species complex.

We use mitochondria DNA sequences and Bayesian inference to estimate phylogenetic relationships and lineage divergence times among the newborn of 2nNCRC and the genus *Carassius*, and investigate the ancestral geographic ranges and the biogeographical history of *C. auratus*.

Results

Morphological traits

Twelve to Twenty samples of each generation of 2nNCRC were random selected to summarize the morphological characteristics on qualitative and quantitative traits. 2nNCRC have no barbells, 29–32 lateral scales, the first gill rakers number ranged from 49–53, 3–4 Spines total and 15–18 Soft-rays total in Dorsal fin, and 3 Spines total and 5–6 Soft-rays total in Anal fin (Table S1 in Supplementary Information). Especially, both of 2nNCRC and *C. auratus* have the same morphotype and formula of pharyngeal teeth, four tabular teeth on each side of the pharyngeal bone. All of that are very different from *C. carpio* and *M. amblycephala*.

Phylogenetic tree and haplotype network

The aligned and concatenated sequences (14 mitochondria genes) were 13634 bp with 5841 variable sites. Of these, 816 were parsimony-informative. The Maximum Likelihood (ML) and Bayesian tree showed a clear identification on the species of *Carassius* and most species of Cyprinidae (Fig. 1). All species from *Carassius*, containing the 2nNCRC formed a monophyletic clade with high support values, and three valid species of *Carassius* were observed (Fig. 1, in bold). However, the 2nNCRC, together with the *C. gibelio*, were unable to disentangle from the *C. auratus* but form a single genetic cluster, and the genetic distances among 2nNCRC and *C. auratus* and *C. gibelio* were much lower than that among 2nNCRC and other species of *Carassius* (Table S2 in Supplementary Information). Consequently, the *C. auratus* and *C. gibelio* were named as *C. auratus* complex in the present study.

The Median Joining Network of *cytb* showed that there was no shared haplotype among the *C. carassius*, *C. cuvieri* and *C. auratus* complex, and the latter had an admixture distribution across the Eurasia and no phylogeographical pattern was found. However, the haplotype of *C. carassius* just distributed in Europe with two distinct sub-lineages, and the haplotype of *C. cuvieri* had a narrow distribution in Japan (Fig. 2). In addition, there existed shared haplotypes between 2nNCRC and *C. auratus*.

Divergence time estimation

The fossil calibrated molecular clock using the incorporated 14 genes from mitochondria and only *cytb* yielded similar results for the divergence between *C. carassius* and other four species of *Carassius*, indicating the *C. carassius* lineage diverged from other species of *Carassius* during the Late Miocene (about 8 Mya) (Fig. 3a, 3b). And they both dated the most recent common ancestor (TMRCA) of *C. auratus* during the Late Pliocene (at approximately 3.56 and 3.12 Mya, respectively). However, the split between *C. cuvieri* and *C. langsdorfii* using the incorporated 14 genes was estimated at approximately 6.43 Mya, which was much earlier than that using only *cytb* gene (4.03 Mya).

Within *C. auratus* complex, *C. gibelio* from Russia (Far East) and Mongolia firstly diverged from other sublineages of *C. auratus* complex. *C. auratus* from Fujian in China and Vietnam were estimated to have separated at approximately 1.6 Mya, followed shortly thereafter by the divergences of *C. gibelio* widely distributed in Eurasia at approximately 1.49 Mya. While other separated sublineages of *C. auratus* complex distributed in Mainland China, Taiwan Island, Japan and Europe diverged during 1.26 Mya and 0.47 Mya, but without a distinct phylogeographic pattern (Fig. 3b).

In addition, based on the incorporated 14 genes, the *Cyprinus* most likely diverged from the *Carassius* at approximately 21.98 Mya, and the most recent common ancestor of *Cyprinus* was dated to 10.14 Mya (Fig. 3a), while the *C. carpio* and the species of *Carassius* separated at approximately 11.86 Mya using only the *cytb* gene (Fig. 3b).

Ancestral range reconstruction

The results of the ancestral range reconstruction are depicted in Fig. 4. Among the six models of geographic-range evolution compared in a likelihood framework in Bio-GeoBEARS, the DIVA + *j* model was chosen according to the best likelihood and AICc associated scores (Table S3 in Supplementary Information). According to the ancestral area reconstruction, 34 dispersal and 18 vicariance events occurred within the evolution of the studied *Carassius*. The TMRCA of *Carassius* probably diverged into two distinct clades in the Late Miocene (Fig. 4). One clade included the *C. carassius* mainly distributed in Europe, while the other clade further split into three separate lineages through three vicariance events, with the first event corresponding to *C. cuvieri* and *C. langsdorfii*, whose ancestor was distributed in Japan (node 43, G: 66.94%), and the second event corresponding to the *C. gibelio*, whose ancestor was likely to live in Siberia and Mongolia (node 39, D: 41.32%), and the third event corresponding to the *C. gibelio* and *C. auratus*, whose ancestor mainly distributed in China and Southeast Asia (node 38, F: 54.39%). Hence, the widely distribution of *C. auratus* in Eurasia and North America can be explained by dispersal events.

Species distribution modeling

Maxent modeling predicted a current range similar to that known for *C. auratus* in East Asia with little variance (Fig. S1a in Supplementary Information), the current range of other species of *Carassius* was shown in the Fig. S2a–5a in Supplementary Information. Mean area under the curve (AUC) was 0.962 (SD = 0.018; training AUC range: 0.962–0.971, test AUC range: 0.917–0.983). The similar results were also found in each species of *Carassius* with high AUC values. For *C. auratus*, precipitation of warmest quarter (bio18; 74.9%), precipitation of coldest quarter (bio19; 8.8%), annual precipitation (bio8; 7.3%), and isothermality (bio3; 6.5%) were the largest contributors to the model contributing 97.5% as supported by jackknifing. The similar bio18, bio19, bio3 and bio8 were the largest contributors for *C. cuvieri* and *C. langsdorfii* to the model contributing 98.5% and 99.2%, respectively. The bio19, bio8, bio3 and mean diurnal range (bio2) were the largest contributors for *C. carassius*. While the bio 19, precipitation seasonality (bio15) and bio3 were the largest contributions for *C. gibelio*.

When the model used current conditions to predict suitable habitat for the species of *Carassius* during the last glacial maximum (LGM), the main area of suitable habitat with high probability for *C. auratus* was as follows: the Yunnan-Guizhou Plateau, the middle and lower reaches of the Yangtze River, Taiwan Island and Japan in East Asia for *C. auratus* (Fig. S1b in Supplementary Information), and a considerable dispersal event between mainland China and Japan was observed. The main areas of suitable habitat for other four species of *Carassius* were shown in Fig. S2b-5b.

During the last interglacial period (LIG; ~120–140), suitable habitat reflected that of the present distribution, with greater levels of suitable habitat in east Asia, the southeast Asia and the central Europe for *C. auratus* (Fig. S1c in Supplementary Information), suggesting that *C. auratus* expanded into new areas after the ice sheets of the LGM receded; with great levels of suitable habitat in the whole Europe for *C. carassius* and *C. gibelio* (Fig. S2c and S4c in Supplementary Information); the great levels of suitable habits in East Asia including China, Japan and the lower elevations of the Himalayas for *C. cuvieri* seem that this species had a wild distribution in East Asia during the LIG (Fig. S3c in Supplementary Information), which was not consistent with current distribution; a very restricted distribution in the islands of the East Asia, including the Taiwan islands, Ryukyu Islands and South Japan (Fig. S5c in Supplementary Information), suggesting that the *C. langsdorfii* distributed in Europe current was likely the introduced species, rather than a native species.

Discussion

This study represents the first attempt to better understand the relationship of the artificial hybrid fish species and the natural evolution fish. The 2nNCRC derived from the interspecies hybridization between *C. carpio* and *M. amblycephala* is very similar to the *C. auratus* in morphological and genetic characteristics with a very close phylogenetic relationship. Our phylogenetic analysis showed that the genus *Carassius* was a monophyletic clade in Cyprinidae and three species including *C. carassius*, *C. cuvieri* and *C. langsdorfii* were further confirmed as valid species, however, *C. auratus* and *C. gibelio* were unable to disentangle from each other based on mtDNA. The admixture of mtDNA haplotype lineages and wildy distribution of *C. auratus* across the whole Eurasia seem that it has a very different evolution history. In the following, we discuss our findings in the context of the paleoclimate oscillation occurred during Pliocene and Pleistocene.

Divergence and distribution pattern of *Carassius* reflect that the *C. auratus* complex was not likely to be the lineage differentiated from the common ancestor of *C. cuvieri* or *C. carassius*

The *Carassius* species complex was widely distributed across Eurasia and exhibits remarkable morphological and genetic diversity [24, 33], and even some lineages were endemic to particular geographical regions, such as *C. gibelio* was restricted to the northern Amur River systems and eastern Europe [34, 35]. Both *C. langsdorfii* and *C. cuvieri* occurred on the main islands of Japan [36, 37], *C. carassius* mainly distributed in Northwest and Central Europe [38]. According the molecular dating, the common ancestor of *C. langsdorfii* and *C. cuvieri* distributed in Japan split with that of *C. auratus* complex during Late Miocene and Early Pliocene (Fig. 3a and 3b), which was much earlier than that estimated by Gao et al. (2012) [24], but coincide with that estimated by Podlesnykh et al. (2012) [39] and Takada et al. (2010) [40], corresponding to the divergence of island and continental limnetic fauna and the settlement of Japanese islands by ancestral species of *Carassius*. The regressions of the ocean level were probably the diversifying factors that resulted in the formation of land bridges between Japanese archipelago islands and the continent, suggesting that dispersal rather than vicariance has created the current distribution pattern of *Carassius* in Eurasia.

It was believed that the freshwater ichthyofaunas from Japanese Islands and China resembled each other during Miocene and Pliocene as the main islands of Japan and mainland China are known to have formed a contiguous land mass in the late Pliocene [41]. Fossils indicate that in the late Pliocene, *C. auratus* was distributed in North of China and Japan [26, 42], coinciding with the dating of TMRCA of *C. auratus* complex in this study. In that case, we speculate that the *C. auratus* complex lineage should be separated from *C. cuvieri* lineage or in turn, and they should have overlap distribution in East Asia at current and history. However, the *C. cuvieri* has a narrow distribution on the main islands of Japan at current, and the speciation of *C. cuvieri* was much earlier than that of *C. auratus* complex in the present and previous studies [24, 25]. Furthermore, the *C. auratus* distributed in Japan at present should be dispersed from China during the glacial period and the recent human-facilitated dispersal according to the ancestral areal distribution and SDM in the present study (Fig. 4 and Fig. S1 in Supplementary Information). According to the divergence and current distribution pattern of *Carassius* was mainly caused by the dispersal with the help of the formation of land bridges between Japanese archipelago islands and the continent [43, 44]. The common ancestor *Carassius* was likely to dispersal from Eurasia to Japan Island, rather than in reverse direction. The most recent common ancestor of *C. cuvieri* and *C. langsdorfii* was distributed in Japan, while *C. auratus* in China (Fig. 4). We speculate that if there exist a recent common ancestor for the three species, it should be distributed in Chinese mainland, rather than in Japan Islands (Fig. 4, node 44, G: 25.97%). Hence, there are reasons to believe that the *C. auratus* should not be lineage diverged from the most recent common ancestor of *C. cuvieri* and *C. langsdorfii*.

Similarly, the most recent common ancestor of *C. carassius* was distributed in Europe. However, the *C. auratus* distributed in Europe at present was verified dispersing from China or Southeast Asia. For instance, recent genetic evidence supports anthropogenic introduction of the crucian carp to the UK during the 15th century [45]. The *C. carassius* is native to parts of central, eastern and northern Europe and almost exclusively restricted to lentic ecosystems [38, 46]. The strong geographic structure was found because of the susceptibility of *C. carassius* to genetic isolation and bottlenecks caused by their small population sizes and especially the low dispersal [27]. Two genetically distinct sub-lineages of *C. carassius* distributed in Europe was found in the present study, coinciding with the previous study that one found throughout northern and central-eastern European drainages and a second almost exclusively confined to the Danubian catchment. Consequently, the Carpathian Mountains and the Central European Highlands was inferred to have acted as a barrier to the colonization of *C. carassius* into northern European drainages during the Pleistocene. According to the reconstructed ancestral areas and SDM, the *C. carassius* originated from Europe and was restricted to their native areas (Fig. 4 and Fig. S2 in Supplementary Information). Consequently, very few records of *C. carassius* distributed in Asia were found, except West Siberia, Kazakhstan, Uzbekistan and Turkey, all of which were very close to Europe and there existed no significant geographical barrier with Europe. However, the Central Siberian Plateau, Mongolian Plateau, Altay Mountains, Tianshan Mountains, Pamir Mountains, Himalayas, Iran Plateau and Great Caucasus, with an elevation more than 1000 m, formed natural dispersal barriers like watersheds impeding freshwater species exchange. According to the geological event of Eurasia, the diversification of *C. auratus* from *C. carassius* far postdated the formation of high-elevation orogens such

as the Tibetan Plateau, Mongolian Plateau and Tian-Shan, which happened during Miocene and even earlier [47], as well as the well-dated *Carassius* fossils have been found in North of China [26], all far postdating the formation of those mountains. Furthermore, in view of the current and paleogeographic distribution of *C. carassius* (Fig. S2 in Supplementary Information), the recent common ancestor of *C. carassius* was almost absent in East and Southeast Asia. This all suggests that the *C. auratus* should not be the lineage diverged from the most recent common ancestor of *C. carassius*.

The paleoclimate and geological event facilitated the distant hybridization

Based on the above analysis, and the fact of the 2nNCRC derived from distant hybridization in our laboratory [16], our results also reinforce a possible hybrid origin route for *C. auratus* in nature. Rapid hybrid speciation was recently shown to occur by Lamichhane et al. (2018) [5], and many species of cichlid fishes in East Africa have been verified that the ancestral distant hybridization has been suggested to play a key role in facilitating species diversification [10–12]. Our finding further supported that this was very likely to have happened in Cyprinidae.

The morphological characteristics on qualitative and quantitative traits between 2nNCRC and *C. auratus* are very similar. Furthermore, it has been certificated that the paternal (*M. amblycephala*) mitochondria DNA fragments were stably embedded in the mitochondria genomes of F₁-F₃ generations of 2nNCRC to form a chimaera, and most of these mutation sites were similar and even consistent with the existing diploid *C. auratus* [48].

In the phylogenetic tree, the closest relationship between *Carassius* and *Cyprinus* was further revealed, suggesting the origin of *Carassius* species could be correlated with *Cyprinus*. The robust, time-calibrated molecular phylogeny suggested that the recent cyprinid fish fauna in East Asia probably originated at 21.98 Ma (Fig. 3a), coinciding with that was estimated from early to middle Miocene [49]. The separation of *C. carpio* and *Carassius* occurred about 11.86 Mya (Fig. 3b), close with the divergence between goldfish and common carp (10.0 Mya) [50], and the origin of the common ancestor of common carp and crucian carp (11.4–8.1 Mya) [31], all of that close to the time of hybrid speciation of allotetraploid *C. carpio* (12.4 Mya) estimated by Xu et al. (2019) [30]. The most recent common ancestor of *C. auratus* and *C. gibelio* was estimated approximately at 3.12 Ma, whereas there is no very significant differentiation between the two species in the present study, maybe correlation with the interspecies hybridization, as they are often sympatry in the same river system, and the hybridization between interspecies is not uncommon in genus *Carassius* [51, 52]. Our results showed that the *C. auratus* most originated from Asia, especially in China (Fig. 4), and the speciation time seem to be related with the violent climate oscillation and geological event occurred during Pliocene and Pleistocene.

Landscape features and geological history were hypothesized to strongly affect the biogeographical processes [53, 54], and this can be especially true in freshwater systems [55, 56]. One major consequence of drastic climatic changes and geological activities are ecosystem and landscape reconfigurations. A

specifically large event is the drastic uplift of the Tibet Plateau ultimately leading to the formation of flowing eastward drainage system in China, such as Yangtze River [57, 58], which was built between Oligocene and Miocene, but the unification of the upper and middle Yangtze River in the Three Gorges mountain region was happened between Late Pliocene and Early Pleistocene [59, 60]. Followed by the Yangtze River and Han River flowed into the middle of Jiangnan Basin. The connection of the upper and middle Yangtze River would facilitate species dispersal from west to east, similar to what has been observed for the frog *Quasipaa boulengeri* [61] and the freshwater snail *Bellamya aeruginosa* [62, 63]. Therefore, the establishment of the east-flowing Yangtze River and the changes of river systems could offer opportunities for the dispersal, gene flow and even for secondary contact and hybridization between closely related freshwater species [63], especially for fish. Furthermore, the intensified tectonic movement and glacial cycling in Pleistocene, as well as the prevailing monsoon had a major impact on the East Asian biota since the late Pliocene [64, 65], as a result the weather were cold and dry [66, 67], rivers and lakes experienced frequent range expansions and regressions in Jiangnan Basin and Dongting Lake basin [68, 69]. The regressions of lakes and rivers could dramatically reduce the habitat region of fish, and this precarious situation was further exacerbated in glacial period. Consequently, the restricted habitat region would cause hybridization between closely and even not closely related species. An example of a specifically large event is in the LGM, the sea level would be 130–150 m lower than today, resulting in strengthened down-cutting processes of the rivers, most lakes in Plains region of Eastern China have opened and dried, which has been verified in Taihu Lake, Boyang Lake and Dongting Lake [70]. The cooler and drier climates intensified during this period, and the fish habitats in these lakes would be fragmented or even lost. The secondary contact of different species would be common when the repeated loss of habitat as the vast majority of freshwater fishes were in vitro fertilization. Specially, the hybridization could be common when allopatric species come into secondary contact [71], hence creating opportunities for new speciation and diversification. The distant hybridization between *C. carpio* and *M. amblycephala*, with the similar ecological niches, most likely occurred during the late Pliocene. A typical example is the repeated range expansions and regressions of lakes likely contributed to the high diversity of African cichlids [72, 73], since East Africa underwent dramatic climatic and geological changes in Pleistocene over the past few million years, the constant expansion and regression of the great East African lakes have led to the repeated loss of habitats or formation of new habitats [63, 73], such as the hybridization might have facilitated these speciation bursts for the cichlids in Lake Tanganyika, and the Time-calibrated trees supported that the radiation of Tanganyika cichlids coincided with lake formation [11].

However, during the interglacial period with the warmer and wetter conditions, the *C. auratus* adapted to their environment thrive and leave more offspring and ultimately would have dispersed northward to Heilongjiang, eastwards to Zhejiang and Fujian, and south-westward to Yunnan and Vietnam. It was believed that *C. auratus* complex had a Pleistocene fast radiation [24], coinciding with the diversification of *C. auratus* lineage in present study (Fig. 3a, 3b). The divergence time among several sublineages of *C. auratus* complex was ahead of 0.34 Mya and posterior to 0.47 Mya (Fig. 3b) coinciding with the end of China's 'Penultimate Interglacial Period' (0.46–0.33 Mya) as defined by Yi et al. (2005) [74], while the

divergence among most of sublineages were ahead of 0.86 Mya, which was earlier than the start of Wangkun Glaciation (0.70–0.50 Mya) in the Middle Pleistocene dated by ESR and paleo-magnetism, widely accepted as the earliest glaciation in China [75]. Consequently, the dispersal rather than vicariance was most responsible for the phylogenetic pattern of *C. auratus* in China.

Conclusions

The newborn homodiploid crucian carp-like fish derived from the hybridization of *C. carpio* and *M. amblycephala* was exactly the same as the diploid *C. auratus* in nature. The molecular phylogenetic analyses revealed an intraspecies relationship between the 2nNCRC and the diploid *C. auratus*. According to the reconstruction of ancestral areas and estimated divergence time, the *C. auratus* lineage most likely originated from the East Asia during the later Pliocene when the hydrological changes and connectivity happened in China, as well as the repeated expansion and regression of lakes facilitating the distant hybridization between the species of Cyprinidae, and especially the several glacial-interglacial cycles contributed to the dispersal and lineage differentiation. Considering the paleoclimate oscillation and geological event during Pliocene and Pleistocene, we speculated that the existence of possible route through distant hybridization under natural conditions generated the *C. auratus* due to the unlikeliness that the *C. auratus* lineage diverged from the most recent common ancestor of *C. carassius* or *C. cuvieri*. However, it's only one speculate that the *C. auratus* would have the similar hybrid origin with the newborn crucian carp-like fish, further studies should be executed to study the population genetics and biogeography on the *C. auratus*, *C. carpio* and *M. amblycephala*, in order to verify our hypotheses or negate it.

Material And Methods

Samples and sequence data preparation

All of the 2nNCRC samples, including F₁ to F₅ generations used in this study were cultured in ponds at the Protection Station of Polyploidy Fish, Hunan Normal University, and fed artificial feed. The F₁ generation of 2nNCRC is denoted as 2nNCRC-F₁, the F₂ generation of 2nNCRC was the F₁ self-cross offspring and denoted as 2nNCRC-F₂, and so on the 2nNCRC-F₃, 2nNCRC-F₄ and 2nNCRC-F₅. One specimens of each generation (F₁ to F₃) were collected and a total of three mitochondrial genomes of 2nNCRC were sequenced to construct a phylogeny tree based on two both rRNAs and twelve protein-coding genes (ND6 was excluded) (the methods for DNA extraction, amplification, and sequencing are given in additional file 1 in Supplementary Information). To perform a comprehensive phylogenetic analysis, another 85 specimens of representative Cyprinidae fish (including 31 specimens of *Carassius*, 19 of *Cyprinus*, 14 of *Megalobrama* and 24 of other cyprinid fish) and 10 Catostomidae fish complete mitochondrial genomes (both rRNAs and twelve protein-coding genes) were retrieved from the Genbank (see additional file 2 in Supplementary Information). In addition, 837 cyt b genes of *Carassius* were also acquired from the

Genbank (see additional file 3 in Supplementary Information) to obtain the phylogeographic structure of *Carassius* across the Eurasia.

Previous study showed that the morphometric of crucian carp-like fish (2nNCRC) were very similar with that of *C. auratus* on the lateral scales and the conventional measures [16]. In this study, the morphological characteristics on qualitative and quantitative traits were further compared among the species of *Carassius* and 2nNCRC, as well as its parents (see Table S1 in Supplementary Information). Furthermore, Pairwise genetic distance of inter- and intraspecific variation in *Carassius* genus (see Table S2 in Supplementary Information) were calculated under the K2P model for the cytb dataset using MEGA 7 [76].

Phylogenetic analysis

All sequences were aligned with MAFFT [77] as implemented in Geneious 4.8.3. For mitochondrial genomes, all alignments (twelve proteins and two rRNAs) were combined. We tested the 14 genes data for saturation using DAMBE v. 6.4.41 [78]. The test revealed an ISS-value that was significantly lower than the ISS.c in all cases ($P < 0.0000$), indicating the suitability of the fourteen genes for phylogenetic analysis. The homogeneity of the fourteen genes was tested in *PAUP [79], and the P value was 0.69 (> 0.05). To test our hypothesis that the 2nNCRC is precisely the *C. auratus*, the ML and Bayesian trees incorporating the mitochondrial genomes (14 genes data) were generated using RAxML v. 8.0 [80] with 1000 bootstrap replicates and MrBayes v. 3.2 [81], respectively. In the phylogenetic tree analysis, 98 sequences (see additional file 2 in Supplementary Information) were used, and the species of Catostomidae were used as outgroup. We determined the best-fitting substitution models for each gene fragment using Modeltest 3.7 [82], which were shown in Table S3 in Supplementary Information. Mixing of the MCMC chains of the two independent runs was monitored with TRACER v. 1.7.1 [83] and the analysis was terminated after the average standard deviation of the split frequencies fell under 0.01. The first 25% of the sampled approximately 20 million generations were discarded as burn-in. The final trees were visualized in FIGTREE v. 1.4.4 [84].

To further visualize haplotype diversity and distribution of *Carassius*, we generated two haplotype network using 837 cytb genes, and colored each haplotype by the geographic region from where it was collected. The haplotype networks were constructed using network v. 5.0 (www.fluxus-engineering.com/sharenet.htm) and applying the median-joining and Maximum Parsimony options.

Divergence time estimation

To test our hypothesis that the divergence times of *Cyprinus* and *Megalobrama* are much earlier than that of *Carassius*. The Divergence time was estimated using a molecular clock approach as implemented in BEAST. We used the combined data in the phylogenetic tree and employed a (uncorrelated lognormal) relaxed clock, as a likelihood-ratio test (LRT) [85] rejected the strict molecular clock hypothesis for the data ($P < 0.01$), with three calibration points. We used the following calibration points with normal distribution priors: the oldest fossil of *Plesiomyxocyprinus arratiae* similar to *Myxocyprinus asiaticus* was constrained to from the middle Eocene or earlier, approximately 40–38 Mya [86], the minimum age of

Catostomidae is 60 Ma based on a catostomid fossil from the Paleocene [87], and the timing of the drastic uplift of the Tibet Plateau between 25–17 Mya [88], which was utilized as one calibration point for the Schizothoracinae as the species endemic to Qinghai-Tibet Plateau [89–91]. The clade corresponding to each calibration point was constrained to be monophyletic. The GTR + I + G model was the best fit for the combined dataset (Table S3 in Supplementary Information). The ‘speciation: Yule process’ tree prior was used to construct the tree. We ran four independent runs for 20 million generations logging trees every 2,000 generations. Convergence was checked with TRACER v. 1.7.1 [83]. Posterior trees from the four runs were combined after removing the first 10% as burn-in in LogCombiner v. 2.5.2 (<http://beast.bio.ed.ac.uk/logcombiner>). The maximum credibility tree was created in TreeAnnotator v. 2.5.2 available in the BEAST package.

Furthermore, another divergence time was estimated using the *cytb* data. We used a reduced dataset of *cytb*: specimens for which two or more sequences from the same region were included. This resulted in 250 terminals – 237 representatives of the five species of *Carassius* and 13 outgroup sequences (*C. carpio*) from GenBank (see additional file 4 in Supplementary Information). We used a conservative approach by employing calibration points from previous studies [27]. The most recent common ancestor (TMRCA) of *C. carassius* in the northern and central-eastern European drainages and the Danubian catchment was constrained to 2.18–2.15 Mya. Another calibration point was the fossil of *Carassius* in Pliocene epoch (5.3–2.6 Mya) in north of China (Yushe, Shanxi province), The GTR + I + G model was also the best fit for the *cytb* dataset (Table S3 in Supplementary Information). The ‘speciation: Yule process’ tree prior was used to construct the tree. We ran four independent runs for 50 million generations logging trees every 5,000 generations. The other settings are same with the above.

Reconstruction of ancestral areas

To test our hypothesis that whether the *C. auratus* was derived from the processing of lineage differentiation with other species in *Carassius* or not, we further performed a biogeographic reconstruction of ancestral areas for species of *Carassius* using BioGeoBEARS [92] in RASP 4.02 [93]. The analyses were conducted on a fully resolved topology from the BEAST analysis containing five species of *Carassius*, eight major geographical areas were defined based on the worldwide distribution of *Carassius* according to their current distribution and Jeffries et al. (2016) [27]: (A) the Europe, except for the south of Alpes mountains and Danube River basin, (B) the south of Alpes mountains and Danube River basin in Europe, (C) Western Asia, (D) Siberia in Russia and Mongolia, (E) the Amur River (including the Amur River in China, Mongolia and Russia), (F) the Yangtze River Basin and its south area in China and Southeast Asia, (G) the East Asia–Japan, (H) the North America. All six models of geographic range evolution were compared in a likelihood framework (DEC, DEC + *j*, DIVALIKE, DIVALIKE + *j*, BAYAREALIKE and BAYAREALIKE + *j*). The best-fit model was assessed on comparing the Akaike's information criterion and likelihood-ratio tests and the DIVALIKE + *j* was choosed (Table S4 in Supplementary Information). Ancestral distributions were reconstructed using 24 sequences in *Carassius* (see additional file 5 in Supplementary Information). To account for phylogenetic uncertainty, 4000 post-burn-in trees resulting

from the BEAST analysis were integrated for inference. The maximum number of ancestral areas was set to four, as *C. auratus* and *C. gibelio* can be widespread.

Species distribution and paleodistribution modeling

We used species distribution modeling (SDM) to construct a model of current, Last Glacial Maximum (LGM, about 22,000 years ago), and Last Interglacial (LIG; ~120,000–140,000 years before present) *Carassius* distributions. The occurrences records for nearly all Eurasia species in the genus *Carassius* (including *C. carassius*, *C. cuvieri*, *C. langsdorfii*, *C. gibelio* and *C. auratus*) were collected from the Global Biodiversity Information Facility (GBIF, <http://data.gbif.org/>), Fishbase (<https://www.fishbase.de/>), the Fish database of Taiwan (<http://fishdb.sinica.edu.tw/>) and complementary literature (sources summarized in additional file 6 in Supplementary Information). Data records were inspected and occurrences outside of Eurasia, without geo-referencing, were excluded from the analyses. Furthermore, we performed a careful quality check and meticulously scrutinized all records for any geographic or taxonomic issues. Geographic duplicate and very adjacent records, as well as records from introduced species or those pre-dating 1970 were removed, as the current bioclimatic dataset ranges over a 30-year period (1970–2000), and a total of 1,499 unique, geo-referenced and quality-checked occurrence records were finally retained across Eurasia for species in the genus *Carassius*.

We extracted current bioclimatic data from the WORLDCLIM dataset (v 2.1, <http://www.worldclim.org/>) at 30 arc-seconds resolution, LGM bioclimatic data from the Model for Interdisciplinary Research on Climate (MIROC) dataset at 2.5-min resolution [94], and LIG bioclimatic data from Otto-Bliesner et al. (2006) [95] at 30 arc-seconds resolution. There exist nineteen bioclimatic variables are included in the WORLDCLIM current and LIG [95] and LGM [96] datasets. The ArcGIS 10.0 (ESRI: Redlands, CA) was used to extract climatic variable layers to include only Eurasia to improve predictive power of Maxent models [97]. Prior to constructing SDM, the ENMTools (v 1.3) [98] was used to determine which bioclimatic variables were correlated, using $R > 0.90$ as a cutoff, and twelve variables which were found to be correlated with at least one other variable were removed.

We used the Maxent (v 3.4.0) [99] to model current and past *Carassius* distribution, and used the following settings for the Maxent model: hinge features only, regularization multiplier of 1, 10,000 max number of background points, replicate run type of 10 cross-validations, 500 maximum iterations, and 0.00001 convergence threshold. We ran jackknife tests to measure the importance of each bioclimatic variable. Models used 700, 49, 56, 489 and 205 (total 1499) records for testing and six BIOCLIM environmental layers (bio2, 3, 8, 15, 18, 19) to produce models for present and paleodistributions of *C. carassius*, *C. cuvieri*, *C. langsdorfii*, *C. gibelio* and *C. auratus*, respectively.

Declarations

Acknowledgements

We thank reviewers for valuable comments on previous versions of the manuscript. All authors approved the final version of the manuscript.

Authors' contributions

S.J.L conceived and designed the study, Q.H.G drafted the manuscript, analyzed the data, and co-wrote the manuscript. S.W made fish measurements, and carried out the molecular laboratory work, and co-wrote the manuscript, Y.H. and H.Z. carried out tissue dissection, DNA isolation, sample preparation for sequencing, and co-wrote the manuscript. J.L.Y, C.H.Y., X.X.H, X.W.X, and Z.H.W. carried out the molecular laboratory work and collect the mtDNA data from GenBank.

Funding

This work received funding and design inspiration from National Natural Science Foundation of China (Grant No. 31730098, U19A2040, 31802287) and the China Postdoctoral Science Foundation (Grant No. 2018M632969), the High-Level Talent Agglomeration Program of Hunan, China (Grant No. 2019RS1044), the Key Research and Development Program of Hunan Province (Grants No. 2018NK2072) and the Hunan Provincial Natural Science and Technology Major Project (Grant No.2017NK1031), the earmarked fund for China Agriculture Research System (Grant No. CARS-45), 111 Project (D20007), supply enough funding and equipment for the cultivation of crucian carp-like fish in our lab.

Availability of data and materials

Many genetic data were obtained through the Genbank (Additional file 2 and file 3 in Supplementary Information), and the new sequences data for 2nNCRC have been submitted in Genbank.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing of interests.

References

1. Harini BP, Ramachandra NB. Evolutionary experimentation through hybridization under laboratory condition in *Drosophila*: Evidence for Recombinational Speciation. BMC Evol Biol. 2003;3:20.
2. Comeault AA, Matute DR. Genetic divergence and the number of hybridizing species affect the path to homoploid hybrid speciation. Proc Natl Acad Sci USA. 2018;115:9761–9766.

3. Mayr E. *Animal Species and Evolution*. London, UK: Harvard University; 1963.
4. Mallet J. Hybrid speciation. *Nature*. 2007;446:279–283.
5. Lamichhaney S, Han F, Webster MT, Andersson L, Grant BR, Grant PR. Rapid hybrid speciation in Darwin's finches. *Science*. 2018;359:224–227.
6. Liu S, Luo J, Chai J, Ren L, Zhou Y, Huang F, et al. Genomic incompatibilities in the diploid and tetraploid offspring of the goldfish × common carp cross. *Proc Natl Acad Sci USA*. 2016;113:1327–1332.
7. Dowling TE, Secor CL. The role of hybridization and introgression in the diversification of animals. *Annu Rev Ecol Syst*. 1997;28:593–619.
8. Kwan YS, Ko MH, Jeon YS, Kim HJ, Won YJ. Bidirectional mitochondrial introgression between Korean cobitid fish mediated by hybridogenetic hybrids. *Ecol Evol*. 2019;9:1244–1254.
9. Pazmino DA, van Herderden L, Simpfendorfer CA, Junge C, Donnellan SC, Hoyos-Padilla EM, et al. Introgressive hybridisation between two widespread sharks in the east Pacific region. *Mol Phylogenet Evol*. 2019;136:119–127.
10. Meier JI, Marques DA, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat Commun*. 2017;8:14363.
11. Irisarri I, Singh P, Koblmüller S, Torres-Dowdall J, Henning F, Franchini P, et al. Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nat Commun*. 2018;29:3159.
12. Svoldal H, Quah FX, Malinsky M, Ngatunga BP, Miska EA, et al. Ancestral Hybridization Facilitated Species Diversification in the Lake Malawi Cichlid Fish Adaptive Radiation. *Mol Biol Evol*. 2019; 37:1100–1113.
13. Wang S, Tang C, Tao M, Qin Q, Zhang C, Luo K, et al. Establishment and application of distant hybridization technology in fish. *SCI China Life SCI*. 2019;62:22–45.
14. Liu S. *Fish Distant Hybridization*. Beijing: China Social Sciences Press; 2014.
15. Liu S, Duan W, Tao M, Zhang C, Sun Y, Shen J, et al. Establishment of the diploid gynogenetic hybrid clonal line of red crucian carp × common carp. *SCI China Ser C*. 2007;50:186–193.
16. Wang S, Ye X, Wang Y, Chen Y, Lin B, Yi Z, et al. A new type of homodiploid fish derived from the interspecific hybridization of female common carp × male blunt snout bream. *SCI Rep-UK*. 2017;7:4189.
17. Avise JC. Phylogeography: Retrospect and prospect. *J Biogeogr*. 2009;36:3–15.
18. Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, et al. Phylogeography's past, present, and future: 10 years after Avise, 2000. *Mol Phylogenet Evol*. 2010;54:291–301.
19. Ivory SJ, Blome MW, King JW, McGlue MM, Cole JE, Cohen AS. Environmental change explains cichlid adaptive radiation at Lake Malawi over the past 1.2 million years. *Proc Natl Acad Sci USA*. 2016;113:11895–11900.

20. Jensen MP, Fitzsimmons NN, Bourjea J, Hamabata T, Reece J, Dutton PH. The evolutionary history and global phylogeography of the green turtle (*Chelonia mydas*). *J Biogeogr.* 2019;46:860–870.
21. Song K, Gao B, Halvarsson P, Fang Y, Jiang YX, Sun YH, Höglund J. Genomic analysis of demographic history and ecological niche modeling in the endangered Chinese Grouse *Tetrastes sewerzowi*. *BMC Genomics.* 2020;21:581.
22. Dussex N, Alberti F, Heino MT, Olsen RA, van der Valk T, Ryman N, et al. Moose genomes reveal past glacial demography and the origin of modern lineages. *BMC Genomics.* 2020;21:854.
23. Hewitt GM. The genetic legacy of the Quaternary ice ages. *Nature.* 2000; 405:907–913.
24. Gao Y, Wang SY, Luo J, Murphy RW, Du R, Wu SF, et al. Quaternary palaeoenvironmental oscillations drove the evolution of the Eurasian *Carassius auratus* complex (Cypriniformes, Cyprinidae). *J Biogeogr.* 2012;39:2264–2278.
25. Liu XL, Li XY, Jiang FF, Wang ZW, Li Z, Zhang XJ, et al. Numerous mitochondrial DNA haplotypes reveal multiple independent polyploidy origins of hexaploids in *Carassius* species complex. *Ecol Evol.* 2017a;7:10604–10615.
26. Liu HT, Su TT. Pliocene fishes from Yüshe Basin, Shansi. *Vertebrata Palasiatica* 1962;6:1–125. (in Chinese with English abstract)
27. Jeffries DL, Copp GH, Handley LL, Olsén KH, Sayer CD, Hänfling B. Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the crucian carp, *Carassius carassius*, L. *Mol Ecol.* 2016;25:2997–3018.
28. Liu XL, Jiang FF, Wang ZW, Li XY, Li Z, Zhang XJ, et al. Wider geographic distribution and higher diversity of hexaploids than tetraploids in *Carassius* species complex reveal recurrent polyploidy effects on adaptive evolution. *SCI Rep-UK.* 2017b;7:5395.
29. Li J, Wang X, Kong X, Zhao K, He S, Mayden RL. Variation patterns of the mitochondrial 16S rRNA gene with secondary structure constraints and their application to phylogeny of cyprinine fishes (Teleostei: Cypriniformes). *Mol Phylogenet Evol.* 2008;47:472–487.
30. Xu P, Xu J, Liu G, Chen L, Zhou Z, Peng W, et al. The allotetraploid origin and asymmetrical genome evolution of common carp *Cyprinus carpio*. *Nat Commun.* 2019;10:4625.
31. Yuan J, He Z, Yuan X, Jiang X, Sun X, Zou S. Speciation of Polyploid Cyprinidae Fish of Common Carp, Crucian Carp, and Silver Crucian Carp Derived from Duplicated Hox Genes. *J Exp Zool Part B.* 2010;314B:445–456.
32. Huang WL. The geologic distribution of fish fossiles in China. *Chin J Geol.* 1960;5:249–255. (in Chinese)
33. Jiang FF, Wang ZW, Zhou L, Jiang L, Zhang XJ, Apalikova OV, et al. High male incidence and evolutionary implications of triploid form in northeast Asia *Carassius auratus* *Mol Phylogenet Evol* 2013; 66:350–359.
34. Cherfas NB. Gynogenesis in fishes. *Genetic bases of fish selection* (ed. by V.S. Kirichnikov), pp. 255–273. Berlin: Springer-Verlag; 1981.

35. Gui JF. Retrospects and prospects of studies on the mechanism of natural gynogenesis in silver crucian carp (*Carassius auratus gibelio*). Bulletin of the National Natural Science Foundation of China. 1997;1:11–16. (in Chinese with English abstract)
36. Meng QW, Su JX, Miao XZ. Fish taxonomy. Beijing: China Agriculture Press; 1995.
37. Luo J, Zhang YP, Zhu CL, Xiao WH, Huang SY. Genetic diversity in crucian carp (*Carassius auratus*). Biochem Genet. 1999;37:267–279.
38. Copp GH, Černý J, Kováč V. Growth and morphology of an endangered native freshwater fish, crucian carp *Carassius carassius*, in an English ornamental pond. Aquat Conserv. 2008;18:32–43.
39. Podlesnykh AV, Apalikova OV, Brykov VA. Phylogenetic relationships of silver crucian carp in *Carassius auratus* complex based on mtDNA analysis. Russ J Genet. 2012;48:1207–1217.
40. Takada M, Tachihara K, Kon T, Yamamoto GJ. Biogeography and Evolution of the *Carassius auratus* Complex in East Asia, BMC Evol Biol. 2010;10:7.
41. Kimura M. (ed.). The formation of the Ryukyu Arc and migration of biota to the arc. Naha: Okinawa Times; 2002. (in Japanese)
42. Chang MM, Chen YY. Late Mesozoic and Tertiary ichthyofaunas from China and some puzzling patterns of distribution. Vertebrata Palasiatica, 2000;38:161–175.
43. Park YC, Kitade O, Schwarz M, Kim JP, Kimm W. Intraspecific molecular phylogeny, genetic variation and phylogeography of *Reticulitermes speratus* (Isoptera: Rhinotermitidae). Mol Cells. 2006;21:89–103.
44. Dufresnes C, Litvinchuk SN, Borzee A, Jang Y, Li JT, Miura I, et al. Phylogeography reveals an ancient cryptic radiation in East-Asian tree frogs (*Hyla japonica* group) and complex relationships between continental and island lineages. BMC Evol Biol. 2016;16:253.
45. Jeffries DL, Copp GH, Maes GE, Lawson Handley L, Sayer CD, Hänfling B. Genetic evidence challenges the native status of a threatened freshwater fish (*Carassius carassius*) in England. Ecol Evol, 2017;7:2871–2882.
46. Copp GH. Typology of aquatic habitats in the great ouse, a small regulated lowland river. Regul River. 1991;6:125–134.
47. Sun H, Liu XD. Impacts of the uplift of four mountain ranges on the arid climate and dust cycle of inland Asia. Palaeogeogr Palaeocl. 2018;505:167–179.
48. Wang S, Jiao N, Zhao L, Zhang M, Zhou P, Huang X, et al. Evidence for the paternal mitochondrial DNA in the crucian carp-like fish lineage with hybrid origin. Sci China Life Sci. 2020;63:102–115.
49. Wang X, Gan X, Li J, Mayden RL, He S. Cyprinid phylogeny based on Bayesian and maximum likelihood analyses of partitioned data: implications for Cyprinidae systematics. Sci China Life Sci. 2012;55:761–773.
50. Luo J, Chai J, Wen YL, Tao M, Lin GL, Liu XC, et al. From asymmetrical to balanced genomic diversification during rediploidization: Subgenomic evolution in allotetraploid fish. Sci Adv. 2020;6:eaaz7677.

51. Wouters J, Janson S, Lusková V, Olsén KH. Molecular identification of hybrids of the invasive gibel carp *Carassius auratus gibelio* and crucian carp *Carassius carassius* in Swedish waters. *J Fish Biol.* 2012;80:2595–2604.
52. Rylková K, Kalous L, Bohlen J, Lamatsch DK, Petrýl M. Phylogeny and biogeographic history of the cyprinid fish genus *Carassius* (Teleostei: Cyprinidae) with focus on natural and anthropogenic arrivals in Europe. *Aquaculture.* 2013;380:13–20.
53. Manel S, Luikart G, Taberlet P. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol.* 2003;18:189–197.
54. Waters JM, Craw D, Burrige CP, Kennedy M, King TM, Wallis GP. Within-river genetic connectivity patterns reflect contrasting geomorphology. *J Biogeogr.* 2015;42:2452–2460.
55. Cook BD, Kennard MJ, Real K, Pusey BJ, Hughes JM. Landscape genetic analysis of the tropical freshwater fish *Mogurnda mogurnda* (Eleotridae) in a monsoonal river basin: importance of hydrographic factors and population history. *Freshwater Biol.* 2011;56:812–827.
56. Ferreira M, Aleixo A, Ribas CC, Santos MPD. Biogeography of the Neotropical genus *Malacoptila* (Aves: Bucconidae): the influence of the Andean orogeny, Amazonian drainage evolution and palaeoclimate. *J Biogeogr.* 2017;44:748–759.
57. Yap SY. On the distributional patterns of Southeast-East Asian freshwater fish and their history. *J Biogeogr.* 2010;29:1187–1199.
58. Clark MK, Schoenbohm LM, Royden LH, Whipple KX, Burchfiel BC, Zhang X, et al. Surface uplift, tectonics, and erosion of Eastern Tibet from large-scale drainage patterns. *Tectonics.* 2004;23:TC1006.
59. Fan DD, Li CX. Reviews on researches of timing of the Yangtze draining the Tibetan Plateau to the East China sea. *Marine Geology and Quaternary Geology.* 2007;27:121–131. (in Chinese with English abstract).
60. Zheng HB, Wang P, He MY, Luo C, Huang XT, Jia JT. Timing of the Establishment of the East-Flowing Yangtze River and Tectonic-Geomorphic Implications. *Quaternary Sciences.* 2013;33:621–630. (in Chinese with English abstract).
61. Yan F, Zhou WW, Zhao HT, Yuan ZY, Wang, YY, Jiang K, Jin JQ, Murphy RW, Che J, Zhang YP. Geological events play a larger role than Pleistocene climatic fluctuations in driving the genetic structure of *Quasipaa boulengeri* (Anura: Dicroglossidae). *Mol Ecol.* 2013;22:1120-1133.
62. Gu QH, Husemann M, Ding BQ, Luo Z, Xiong BX. Population genetic structure of *Bellamya aeruginosa* (Mollusca: Gastropoda: Viviparidae) in China: weak divergence across large geographic distances. *Ecol Evol.* 2015;5:4906-4919.
63. Gu QH, Husemann M, Wu HH, Dong J, Zhou CJ, Wang XF, et al. Phylogeography of *Bellamya* (Mollusca: Gastropoda: Viviparidae) snails on different continents: contrasting patterns of diversification in China and East Africa. *BMC Evol Biol.* 2019;19:82.
64. Zhisheng A, Kutzbach JE, Prell WL, Porter SC. (2001). Evolution of Asian monsoons and phased uplift of the Himalaya-Tibetan plateau since Late Miocene times. *Nature.* 2001;411: 62–66.

65. Xin SZ, Shen J, Zhang WF, Sun WW, Xiao XY. East Asian winter monsoon evolution since the late Pliocene based on a pollen record from Lake Xingkai, northeast Asia. *Quaternary Res.* 2020;93:40–59.
66. Chen YY. (ed.). *Fauna Sinica, Osteichthyes, Cypriniformes II*. Beijing: Science Press; 1998.
67. An ZS, Zhang PZ, Wang EC, Wang SM, Qiang XK, Li L, et al. Changes of the monsoon-arid environment in China and growth of the Tibetan Plateau since the Miocene. *Quaternary Sciences.* 2006;26:678–693. (in Chinese with English Abstract)
68. Bai DY, Li CA. Status of Quaternary Geology Research of Dongting Basin. *Geological Science and Technology Information.* 2010;29:1–12.
69. Wang MM. A study on Developmental Mechanism and Tectonic Evolution of the Hanzhong Basin. Dissertation, Beijing: Institute of Geology, China Earthquake Administration; 2013.
70. Shen J. Spatiotemporal variations of Chinese lakes and their driving mechanisms since the Last Glacial Maximum: A review and synthesis of lacustrine sediment archives. *Chinese Sci Bull.* 2013;58:17–31.
71. Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, et al. Hybridization and speciation. *J Evolution Biol.* 2013;26:229–246.
72. Koblmüller S, Egger B, Sturmbauer C, Sefc KM. Rapid radiation, ancient incomplete lineage sorting and ancient hybridization in the endemic Lake Tanganyika cichlid tribe Tropheini. *Mol Phylogenet Evol.* 2010;55:318–
73. Danley PD, Husemann M, Ding B, DiPietro LM, Beverly EJ, Peppe DJ. The impact of the geologic history and paleoclimate on the diversification of East African cichlids. *International Journal of Evolutionary Biology.* 2012;2012:574851.
74. Yi CL, Cui ZJ, Xiong HG. Numerical periods of Quaternary glaciations in China. *Quaternary Sciences.* 2005;25:609–619.
75. Cui ZJ, Chen YX, Zhang W, Zhuo SZ, Zhou LP, Zhang M. Research history, glacial chronology and origins of quaternary glaciations in China. *Quaternary Sciences.* 2011;31:749–764.
76. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 2006;33:1870–1874.
77. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol.* 2013;30:772–780.
78. Xia X. DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. *Mol Biol Evol.* 2013;30:1720–1728.
79. Swofford DL. PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sunderland (Massachusetts): Sinauer Associates; 2002.
80. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30:1312–1313.

81. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 2012; 61:539–542.
82. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics (Oxford, England).* 1998;14:817–818.
83. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol.* 2018;67:901-904.
84. Rambaut A. FigTree v1.4.4 2006-2018: Tree Figure Drawing Tool; 2018. Available at: <http://tree.bio.ed.ac.uk/software/figtree/>.
85. Huelsenbeck JP, Rannala B. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science.* 1997; 276:227–232.
86. Liu J, Chang MM. A new Eocene catostomid (Teleostei: Cypriniformes) from northeastern China and early divergence of Catostomidae. *Sci China Ser D.* 2009;52:189–202.
87. Cavender TM. Review of the fossil history of North American freshwater fishes. Pp. 699–724 in C. H. Hocutt and E. O. Wiley, eds. *The zoogeography of North American freshwater fishes.* New York: John Wiley & Sons; 1986.
88. Harrison TM, Copeland P, Kidd W, Yin A. Raising tibet. *Science.* 1992;255:1663-1670.
89. Tang YT, Li CH, Kun Y, Wang H, Feng CG, Tong C, et al. Convergent evolution misled taxonomy in *Schizothoracine* fishes (Cypriniformes: Cyprinidae). *Mol Phylogenet Evol.* 2019;134:323–337.
90. Wang Y, Shen YJ, Feng CG, Zhao K, Song ZB, Zhang YP, et al. Mitogenomic perspectives on the origin of Tibetan loaches and their adaptation to high altitude. *Sci Rep-UK.* 2016;6:29690.
91. Ruber L, Kottelat M, Tan HH, Ng PK, Britz R. Evolution of miniaturization and the phylogenetic position of *Paedocypris*, comprising the world's smallest vertebrate. *BMC Evol Biol.* 2007;7:38.
92. Matzke N. BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R scripts. Berkeley: University of California; 2013.
93. Yu Y, Harris AJ, Blair C, He XJ. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Mol Phylogenet Evol.* 2015;87:46–49.
94. Hasumi H, Emori S. K-1 coupled GCM (MIROC) description. Tech. Report 1. Tokyo: Center for Climate System Research, University of Tokyo; 2004.
95. Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH. Simulating polar amplification of orbital forcing for the last interglacial. *Science.* 2006;311:1751–1753.
96. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol.* 2005;25:1965–1978.
97. Anderson RP, Raza A. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: Preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. *J Biogeogr.* 2010;37:1378–1393.

98. Warren DL, Glor RE, Turelli M. ENMTools: A toolbox for comparative studies of environmental niche models. *Ecography*. 2010;33:607–611.
99. Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecol Model*. 2006;190:231–259.

Figures

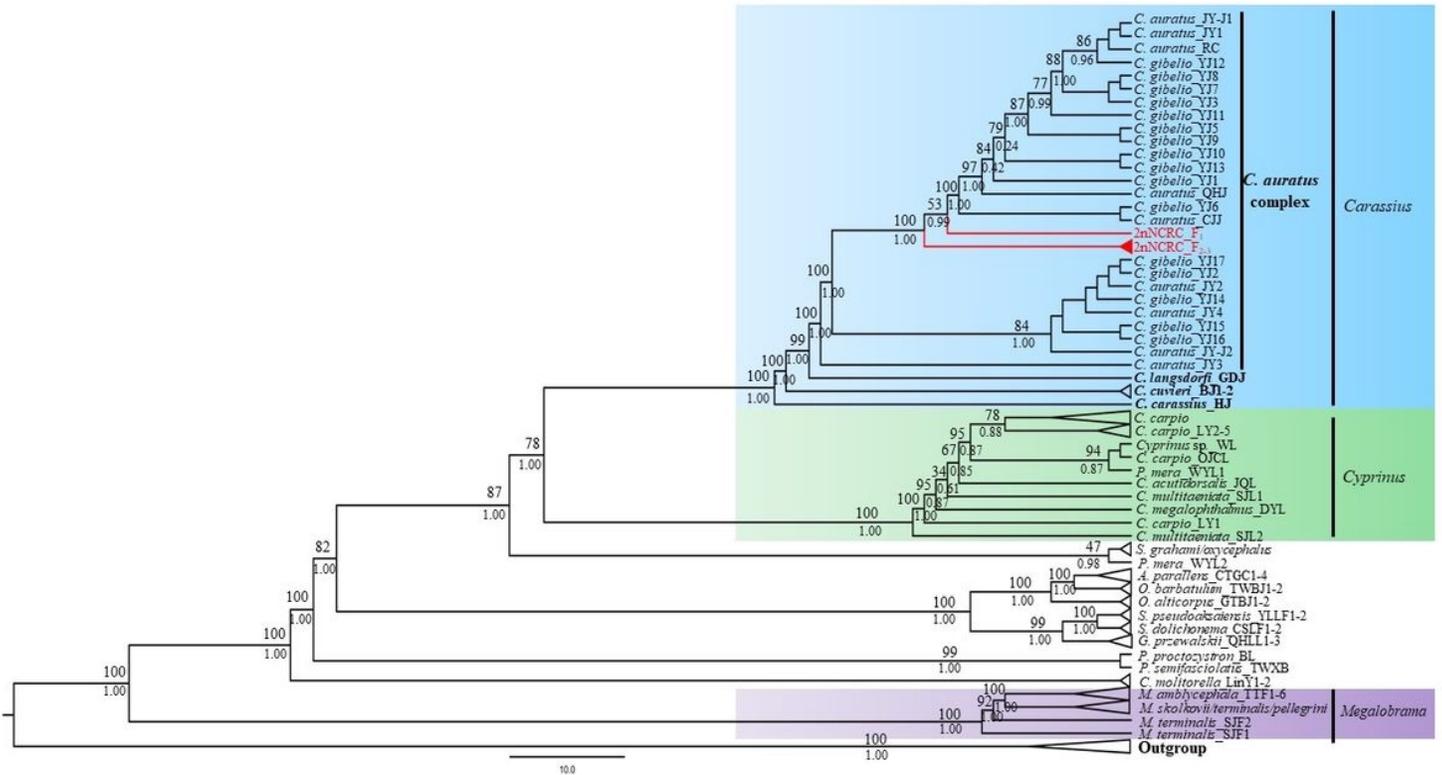


Figure 1

Phylogenetic tree of *Carassius* with 2nNCRC and some representative Cyprinidae fish (including 28 specimens of *Carassius*, 19 of *Cyprinus*, 14 of *Megalobrama* and 24 of other cyprinid fish) using complete mitochondrial genomes based on the concatenated dataset (twelve proteins and two rRNAs). Numbers at nodes are maximum likelihood bootstrap values (up the branch) and Bayesian posterior probabilities (down the branch).

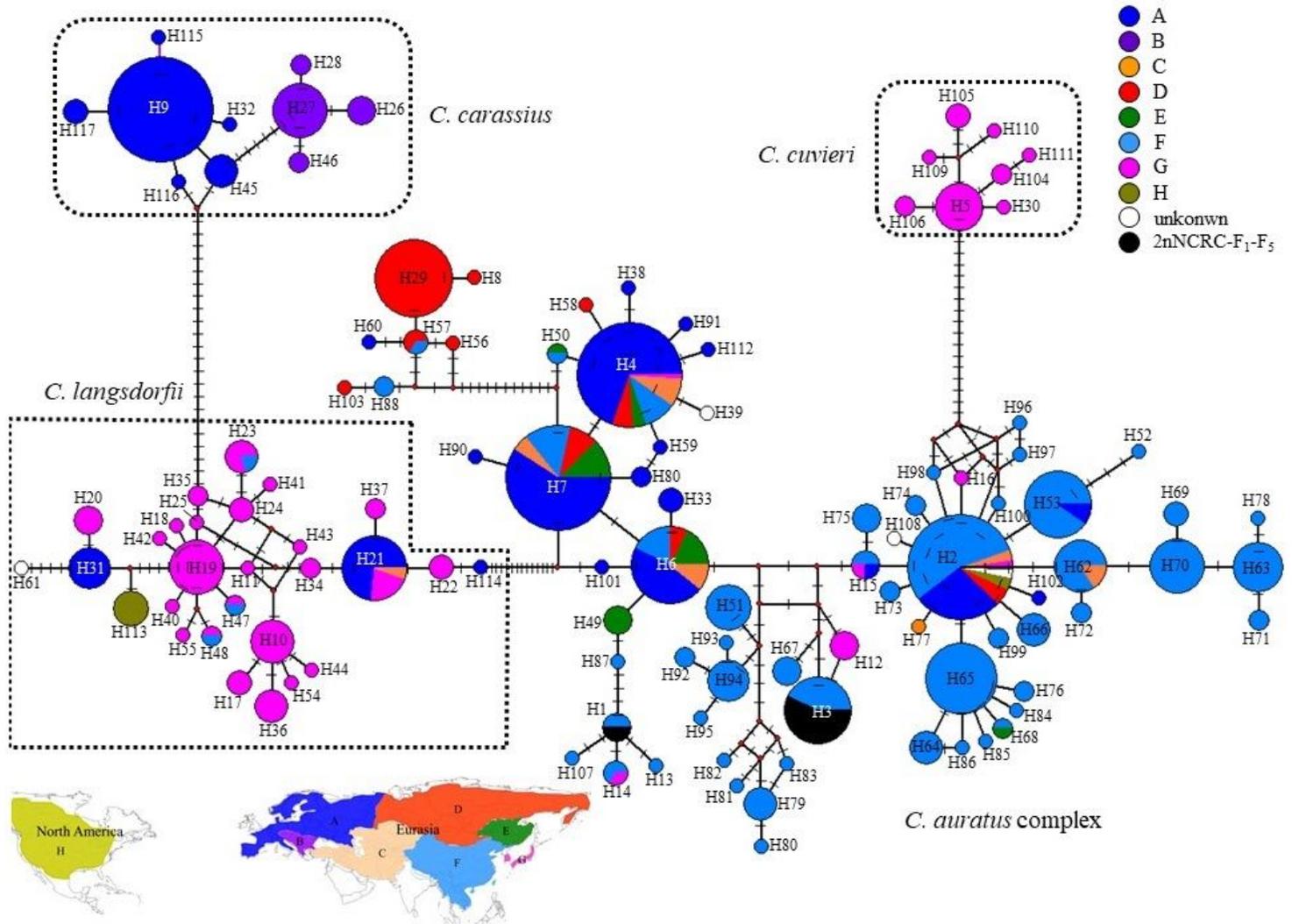


Figure 2

Median-joining haplotype network based on 837 sequences of *cytb* including 15 2nNCRC and 822 specimens of *Carassius* derived from GenBank. The size of the circles represents haplotype frequency. Each connecting line represents a single nucleotide substitution, and each little short line represents mutated position. Where the specimens of *Carassius* in the network were found: (A) the Europe, except for the south of Alps mountains and Danube River basin, (B) the south of Alps mountains and Danube River basin in Europe, (C) Western Asia, (D) Siberia in Russia and Mongolia, (E) the Amur River (including the Amur River in China, Mongolia and Russia), (F) the Yangtze River Basin and its south area in China and Southeast Asia, (G) the East Asia—Japan, (H) the North America. The haplotypes of different geographic samples (see additional file 3 in Supplementary Information) were represented as different colors. The *C. carassius* and *C. cuvieri* formed distinct geographic structure based on the *cytb* gene, while the *C. langsdorfii* and *C. auratus* complex with no apparent geographic pattern. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or

area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

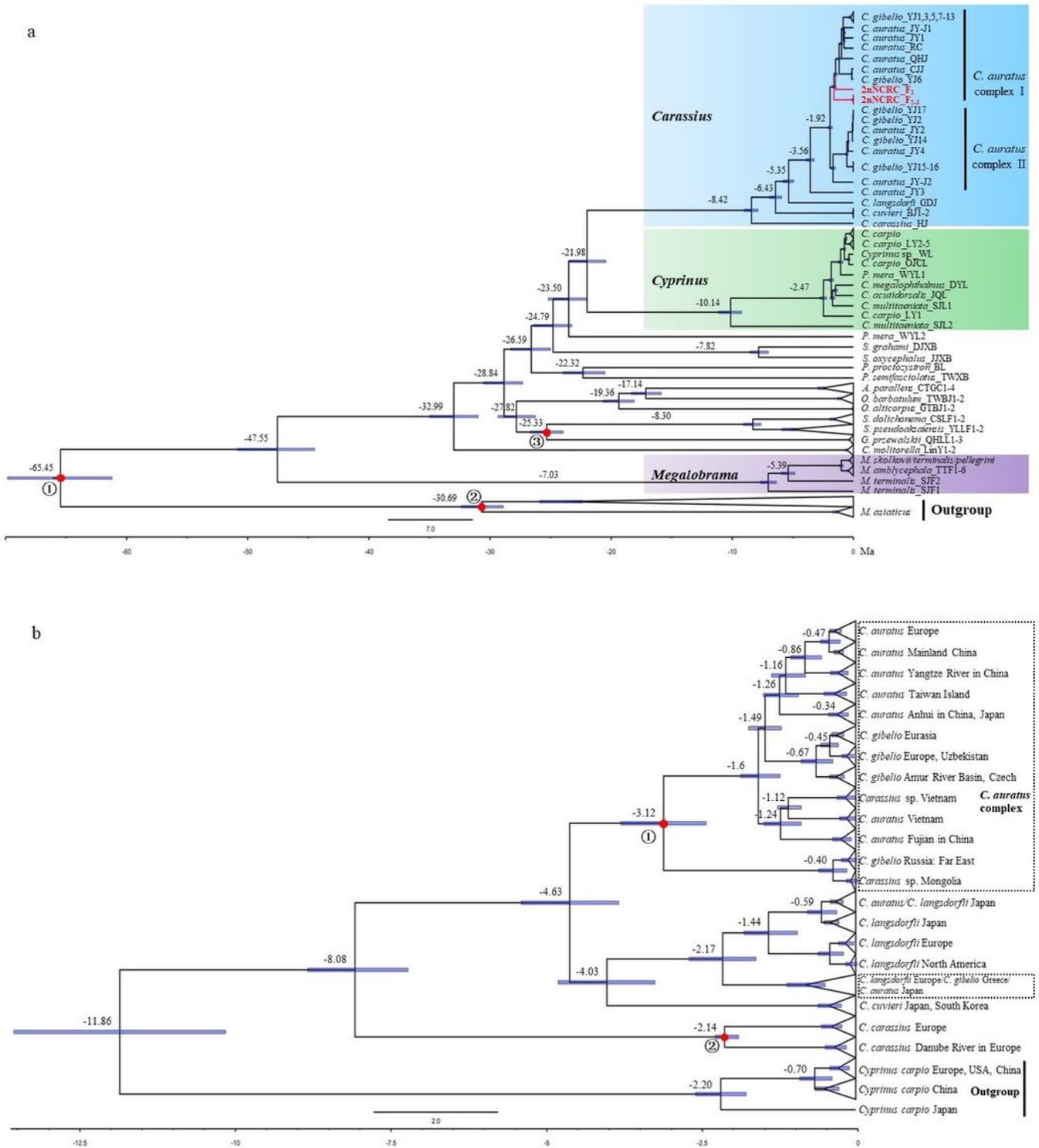


Figure 3

a) Dated phylogeny using the concatenated dataset (twelve proteins and two rRNAs) with three calibration points (red dots indicate time-calibration markers, [1]: Cavender, 1986 [87]; [2]: Liu and Chang, 2009 [86]; [3]: Harrison, 1992 [88]), and b) using cytb gene with two calibration points (red dots indicate

- [additionalfile1.docx](#)
- [Additionalfile2.pdf](#)
- [Addtionalfile3.pdf](#)
- [Addtionalfile4.pdf](#)
- [Addtionalfile5.pdf](#)
- [Addtionalfile6.pdf](#)
- [FigureS1S5SDM.pptx](#)
- [TableS1S4.docx](#)